



Observatorio

Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos

Jorge Serrano-Cobos

Artículos

Tecnologías big data para análisis y recuperación de imágenes web

Sergio Rodríguez-Vaamonde, Ana-Isabel Torre-Bastida y Estibaliz Garrote

El big data transforma la interpretación de los medios sociales

Silvia Martínez-Martínez y Pablo Lara-Navarra

Meloda, métrica para evaluar la reutilización de datos abiertos

Alberto Abella, Marta Ortiz-de-Urbina-Criado y Carmen De-Pablos-Heredero

The generation of large networks from Web of Science data

Loet Leydesdorff, Gohar-Feroz Khan and Lutz Bornmann

Museos sociales. Perfiles museísticos en Twitter y Facebook 2012-2013

Florencia Claes y Luis Deltell

Are libraries sustainable in a world of free, networked, digital information?

Lluís Anglada

Criterios de evaluación y aplicación de fuentes de información web en centros de documentación periodística

Paz Orero y Bernardino J. Cebrián-Enrique

News overload in Spain: The role of demographic characteristics, news interest, and consumer paying behavior

Manuel Goyanes

Aportaciones de un club de lectura escolar a la lectura por placer

Carmen Álvarez-Álvarez y Julián Pascual-Díez

Análisis

Datos enlazados de publicaciones, proyectos y herramientas para investigadores en humanidades digitales: catálogo piloto del centro Clarín IULA-UPF

Silvia Arano-Poggi y Núria Bel

Realidad aumentada en televisión y propuesta de aplicación en los sistemas de gestión cultural

Jorge Caldera-Serrano



BIG DATA + ANALÍTICA WEB

03:14 AM - DATA IMMERSION



El profesional de la información

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

El profesional de la información es una revista de
la editorial EPI SCP

Apartado 32.280 - 08080 Barcelona, España
Tel.: +34 - 609 352 954
<http://www.elprofesionaldelainformacion.com>

Redacción

El profesional de la información
Apartado 32.280
08080 Barcelona, España
Tel.: +34 - 934 250 029
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>

Servicios online

María T. Moreno
mt.moreno@ono.com

Diseño

MASmedios, <http://www.masmedios.com>
Director artístico: Moisés Mañas

Maquetación

Isabel Olea
isabel.iolea@gmail.com

Producción e Impresión

Creación Estelar, S.L.
Pol. Ind. Mascaró
C/ Ponent, nave 6
08756 La Palma de Cervelló (Barcelona, España)
Tel.: +34 - 936 720 099

Distribución online

MetaPress, Birmingham, Alabama, EUA
<http://elprofesionaldelainformacion.metapress.com>

Depósito legal: B. 12.303-1997

Los trabajos publicados en EPI son aprobados según el sistema tradicional "peer review" en doble ciego: son revisados al menos por dos expertos en el tema, del Consejo Asesor de la revista y/o externos.

Para conseguir que los trabajos no pierdan 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
Univ. de Barcelona / Univ. Ramon Llull
<http://sites.google.com/site/sitiodejavierguallar/>

COORDINADOR EDITORIAL

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

Luis Rodríguez-Yunta
CCHS, CSIC, Madrid
luis.ryunta@cchs.csic.es

REDACTORA JEFE

Isabel Olea
isabel.iolea@gmail.com

REDACCIÓN

Natalia Arroyo-Vázquez
Fundación Germán Sánchez Ruipérez
narroyo@fundaciongsr.es

Ricardo Eito-Brun
Grupo GMV
reito@gmv.es

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

Toon Lowette
Grid Electronic Publishing
toon@grid.be

Roser Lozano
CRAI Universitat Rovira i Virgili
roser.lozano@urv.cat

Jorge Serrano-Cobos
MASmedios
jorgeserrano@gmail.com

REVISIÓN DE LENGUA INGLESA

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

COMITÉ CIENTÍFICO

Ernest Abadal
Isidro F. Aguillo
Adela d'Alòs-Moner
Carlos B. Amat
Jesús Bustamante
Carlota Bustelo-Ruesta
Atilio Bustos
Lluís Codina
Emilio Delgado-López-Cózar
Javier Díaz-Noci
Assumpció Estivill
Antonia Ferrer-Sapena
António Fidalgo
Francisco-Javier García-Marco
Johannes Keizer
Jesús Lau
Mari-Carmen Marcos
Pere Masip
Charles McCathieNevile
José-Antonio Moreira-González
Enrique Orduña-Malea
Juan-Antonio Pastor
Fernanda Peset
Joan Roca
Ramón Salaverría
Tomás Saorín
Robert Seal
Ernesto Spinak
Emir-José Suaidén
Daniel Torres-Salinas
Jesús Tramullas

Universitat de Barcelona, Barcelona.
Centro de CC Humanas y Sociales, CSIC, Madrid.
Consultora, Barcelona.
Inst. Agroquím. y Tecn. Alimentos, CSIC, Valencia.
Biblioteca, Cedefop, Salónica, Grecia.
Consultora, Madrid.
Pontificia Univ. Católica de Valparaíso, Chile.
Universitat Pompeu Fabra, Barcelona.
Universidad de Granada, Granada.
Universitat Pompeu Fabra, Barcelona.
Universitat de Barcelona, Barcelona.
Universidad Politécnica de Valencia, Valencia.
Universidade da Beira Interior, Portugal.
Universidad de Zaragoza, Zaragoza.
Food and Agriculture Org. (FAO), Roma, Italia.
Universidad Veracruzana, Veracruz, México.
Universitat Pompeu Fabra, Barcelona.
Blanquerna, Univ. Ramon Llull, Barcelona.
Yandex, Moscow, Rusia.
Universidad Carlos III de Madrid.
Universidad Politécnica de Valencia, Valencia.
Universidad de Murcia, Murcia.
Universidad Politécnica de Valencia, Valencia.
Minnesota State University, Mankato, USA.
Universidad de Navarra, Pamplona.
Universidad de Murcia, Murcia.
Loyola Univ. Chicago, Evanston, Illinois, USA.
Consultor, Montevideo, Uruguay.
IBICT, Brasilia, Brasil.
Universidad de Navarra, Pamplona.
Universidad de Zaragoza, Zaragoza.

BASES DE DATOS

Academic search premier (Ebsco)

<http://www.ebscohost.com/academic/academic-search-premier>

Communication source (Ebsco)

<http://www.ebscohost.com/academic/communication-source>

Dialnet (Universidad de La Rioja)

http://dialnet.unirioja.es/servlet/revista?&clave_revista=469

Economía y negocios (Ebsco)

<http://www.ebscohost.com/academic/economia-y-negocios>

Francis (Inist)

<http://www.inist.fr/?Francis-74>

ISI Social sciences citation index, Social SCI, WoS (Thomson Reuters)

http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci/

Impact Factor 2013 = 0,402

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

<http://bddoc.csic.es:8080/ver//ISOC/revi/0721.html>

ISTA, Information science and technology abstracts (Ebsco)

<http://www.ebscohost.com/public/information-science-technology-abstracts>

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/?PASCAL-73&lang=en>

Scopus (Elsevier) Scimago Journal Rank 2013 = 0,345

<http://www.scopus.com>

CATÁLOGOS Y SERVICIOS BIBLIOTECARIOS

Catálogo de Latindex

<http://www.latindex.unam.mx/buscador/ficRev.html?folio=6772>

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/ec3/Documentacion.html>

ACCESO A LOS TEXTOS COMPLETOS

MetaPress (2000-)

<http://elprofesionaldelainformacion.metapress.com/>

ALPSP Learned journals collection (ALJC) (2011-)

<http://www.swets.com/aljc>

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>

Academic search complete (2000-embargo 1 año)

<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=633>

Business source complete (2000-embargo 1 año)

<http://www.ebscohost.com/academic/business-source-complete>

El profesional de la información (1992-embargo 2 años)

<http://elprofesionaldelainformacion.com/contenidos.html>

SwetsWise (2000-)

<https://www.swetswise.com>

EPI EN REDES SOCIALES

<http://www.facebook.com/elprofesionaldelainformacion>

http://twitter.com/revista_EPI

<http://www.linkedin.com/company/el-profesional-de-la-informacion-epi->

<http://pinterest.com/source/elprofesionaldelainformacion.com>

PLATAFORMA DE PRODUCCIÓN OJS

Recyt, Repositorio español de ciencia y tecnología (Fecyt)

<http://recyt.fecyt.es/index.php/EPI>



Tema central: Big data y analítica digital

OBSERVATORIO

- 561 *Big data* y analítica web. Estudiar las corrientes y pescar en un océano de datos
Jorge Serrano-Cobos

ARTÍCULOS

- 567 *Tecnologías big data* para análisis y recuperación de imágenes web
Sergio Rodríguez-Vaamonde, Ana-Isabel Torre-Bastida y Estibaliz Garrote
- 575 *El big data* transforma la interpretación de los medios sociales
Silvia Martínez-Martínez y Pablo Lara-Navarra
- 582 *Meloda*, métrica para evaluar la reutilización de datos abiertos
Alberto Abella, Marta Ortiz-de-Urbina-Criado y Carmen De-Pablos-Heredero
- 589 *The generation of large networks from Web of Science data*
Loet Leydesdorff, Gohar-Feroz Khan and Lutz Bornmann
- 594 *Museos sociales. Perfiles museísticos en Twitter y Facebook 2012-2013*
Florencia Claes y Luis Deltell
- 603 *Are libraries sustainable in a world of free, networked, digital information?*
Lluís Anglada
- 612 *Criterios de evaluación y aplicación fuentes de información web en centros de documentación periodística*
Paz Orero y Bernardino J. Cebrián-Enrique
- 618 *News overload in Spain: The role of demographic characteristics, news interest, and consumer paying behavior*
Manuel Goyanes
- 625 *Aportaciones de un club de lectura escolar a la lectura por placer*
Carmen Álvarez-Álvarez y Julián Pascual-Díez

ANÁLISIS

- 633 *Datos enlazados de publicaciones, proyectos y herramientas para investigadores en humanidades digitales: catálogo piloto del centro Clarin IULA-UPF*
Silvia Arano-Poggi y Núria Bel
- 643 *Realidad aumentada en televisión y propuesta de aplicación en los sistemas de gestión documental*
Jorge Caldera-Serrano

561 AGENDA

655 EVALUADORES (2013-2014)

660 INFORMACIÓN PARA LOS AUTORES



Creación de Archivos, Museos y 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 o Virtuales

DIGIBIB 8.0

Solución avanzada para la creación de Bibliotecas Digitales y la Gestión Bibliotecaria Multilingüe.

DIGIARCH 2.0

Sistema digital de descripción y gestión archivística. Descripción en ISAD(G) y EAD (2002).

DIGIMUS 2.0

Sistema digital de gestión museológica. Actividad subvencionada por el Ministerio de Educación, Cultura y Deporte.



DIGIHUB. Recolector OAI-PMH

Sistema integrado de recolección de diversos esquemas de metadatos: OAI_DC, ESE v3.4.1, EDM v5.2.5, LIDO 1.0, EAD (2002), EAC-CPF, MARC21...

ADAPTACIÓN A EUROPEANA

Implementación del esquema **ESE 3.4.1** (Europeana Semantic Elements) y **EDM 5.2.5** (Europeana Data Model) Adaptado a la Agenda Digital Europea 2020.

DIGITALIZACIÓN AVANZADA

Con asignación dinámica de metadatos.

- **Recolección en la Web para Entidades e Instituciones de Memoria en OAI-PMH y Dublin Core cualificado con ESE 3.4.1**
- **Consultoría y mapeos a EDM 5.2.5 (Europeana Data Model)**
- **Bibliotecas digitales que permiten la creación, recuperación y recolección de metadatos (MARCXML, Dublin Core y RDF)**
- **Archivos Web que facilitan la creación, recuperación y recolección de metadatos (EAD (2002) y EAC-CPF)**
- **Implementación de la API de búsqueda en Europeana.**
- **Adaptación del repositorio OAI para la transmisión de instancias RDF según ORE**
- **Intercambio de metadatos en METS 1.10 (diferentes Profiles) integrando todos los esquemas de metadatos**
- **Creación de METSRights para el control de los derechos de autor**
- **Reconocimiento Óptico de Caracteres OCR y generación dinámica de ALTO 3.0 (Analyzed Layout and Text Object)**
- **Generación e integración de registros SKOS mediante MARC 21(Up.19)/RDA**
- **Creación de eBooks o libros digitales en formatos: ePub y Mobipocket.**
- **Adaptación de DIGIBIB a Linked Open Data**

ORACLE
PARTNERNETWORK



Incubator Activity



Validación en el Data Providers de la Open Archives Initiative. Genera un Sitemap para Google.



ePUB



Nº ES042816-1



C/ Alenza, 4, 5ª planta. Madrid. Tel.: 914 32 08 88. E-mail: digibis@digibis.com

www.digibis.com

OBSERVATORIO



BIG DATA Y ANALÍTICA WEB. ESTUDIAR LAS CORRIENTES Y PESCAR EN UN OCÉANO DE DATOS



Jorge Serrano-Cobos



Jorge Serrano-Cobos, licenciado en documentación por la *Universidad de Granada*, es profesor asociado de la *Universidad Politécnica de Valencia* donde imparte materias como SEO, analítica web y social media analytics. Ha trabajado en las empresas *MASmedios*, *Google*, *Planeta DeAgostini* y *Serikat*, y ha realizado docenas de proyectos de consultoría y asesoría en marketing académico y científico, estudio de mercados digitales, analítica web, análisis estratégico cualitativo y cuantitativo digital (cibernetría) para e-commerce, e-marketing, arquitectura de la información, diseño de interacción, SEO y search analytics, con vocación de internacionalización.

<http://orcid.org/0000-0002-4394-4883>

MASmedios

Garcilaso, 15. 46003 Valencia, España
jorgeserrano@gmail.com

Resumen

Se realiza un recorrido por las características, posibilidades, disciplinas científicas, técnicas y tecnologías que se recogen dentro del paraguas interdisciplinar del *big data* y la analítica web desde el punto de vista de su aplicación a la praxis. Se realiza una reflexión en torno a los retos, riesgos y problemas que las herramientas y los datos no resuelven por sí solos, así como sobre los contextos de uso de estas técnicas de tratamiento de datos para la toma de decisiones.

Palabras clave

Big data, Analítica web, Analítica predictiva, Analista, *Data mining*, *Text mining*, Análisis, Cibernetría, Marketing, Usuarios, Estadística, Aprendizaje, Informe de situación.

Título: Big data and web analytics. Studying the currents and fishing in an ocean of data

Resumen

A tour is provided of the features, possibilities, scientific, technical and technologies that are collected under the interdisciplinary umbrella of big data and web analytics from the point of view of its application in practice. A reflection is offered about the challenges, risks and problems that the tools and data cannot resolve on their own, as well as the contexts in which these data processing techniques are useful for decision making.

Keywords

Big data, Web analytics, Predictive analytics, Analyst, Analysis, Data mining, Text mining, Cybermetrics, Marketing, Users, Statistics, Machine learning, Situation report.

Serrano-Cobos, Jorge (2014). "*Big data* y analítica web. Estudiar las corrientes y pescar en un océano de datos". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 561-565.

<http://dx.doi.org/10.3145/epi.2014.nov.01>

Introducción

Recuerdo que allá por los 80, la primera vez que vi una computadora con tarjetas perforadas (y no era en un museo) que iba a ser sustituida por un disco (de considerable tamaño, eso sí), me dije "lo que nos queda para *HAL 9000*". En aquella época estaban en boca de todos los amantes de la tecnología las últimas novedades de la inteligencia artificial, y ya nos hacíamos la boca agua con la llamada "quinta generación de computadoras" con capacidad de resolución de problemas complejos imitando la forma de pensar de los seres humanos.

Años más tarde, y antes de otras burbujas, ya tuvimos un "bluff" de esa rama de la inteligencia artificial en este sentido, sustituida por la computación basada en insectos sociales, a su vez sustituida (en cierto modo) por la computación social, esta vez usando personas en lugar de insectos ("*the wisdom of crowds*"). Probablemente no era su momento.

Pero por otro lado, quién le hubiera dicho a Vannevar Bush la que se iba a liar después de la burbuja de internet de 2000 y su *rebranding* a través de la marca "Web 2.0". Viendo cómo han cambiado las cosas, y haciendo un poco de memoria histórica, uno puede tener la tentación de predecir

Artículo recibido el 19-10-2014

hacia dónde irán. Y eso es, hoy día, lo que todo el mundo intenta hacer: lo llaman “*predictive analytics*”.

Variedad, volumen, velocidad, variabilidad

El signo de los tiempos actuales es que nunca en la historia los ciudadanos, instituciones y empresas de todo tamaño hemos tenido acceso a tantos y tan variados datos. La cuestión es qué hacer con ellos y cómo.

Hay una gran variedad de tecnologías, herramientas y disciplinas científicas que intentan lidiar con esa enorme cantidad de datos dispersos que están esperando a ser explotados. Pero para que el consumidor medio pueda explotarlo, o mejor dicho pagar para que se lo hagan, es necesario generar una marca fácilmente reconocible que prometa algo grande sin decir qué: de ahí la marca “*big data*”, que en sí alude a los sistemas que gestionan grandes, enormes, con-

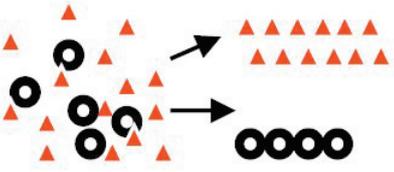
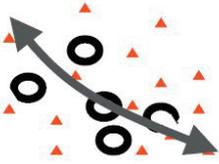
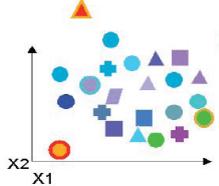
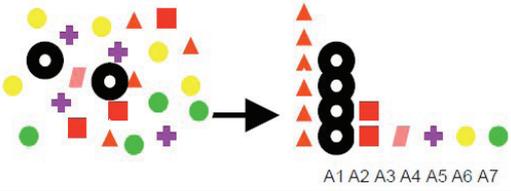
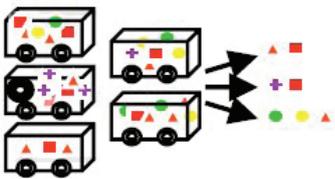
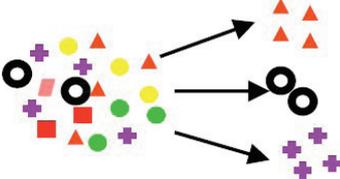
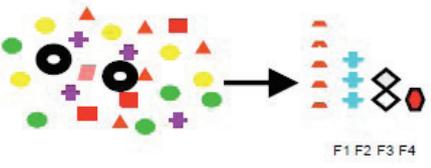
Operaciones y técnicas	Algoritmos	Aplicaciones
<p>Clasificación</p> 	<ul style="list-style-type: none"> -Regresión logística (modelos lineales generalizados, <i>GLM</i>) -Árboles de decisión -Bayes naïf -Máquinas de vectores soporte (<i>support vector machine, SVM</i>) 	<ul style="list-style-type: none"> -Técnica estadística clásica -Popular / reglas / transparencia -App embebida -Amplios / datos estrechos / texto
<p>Regresión</p> 	<ul style="list-style-type: none"> -Regresión múltiple (<i>GLM</i>) -Máquinas de vectores soporte (<i>SVM</i>) 	<ul style="list-style-type: none"> -Técnica estadística clásica -Amplios / datos estrechos / texto
<p>Detección de anomalías</p> 	<ul style="list-style-type: none"> -<i>SVM</i> de una clase 	<p>Falta de ejemplos del campo objetivo</p>
<p>Importancia de atributos</p> 	<ul style="list-style-type: none"> -Longitud de descripción mínima (<i>minimum description length, MDL</i>) 	<ul style="list-style-type: none"> -Reducción de atributos -Identificación de datos útiles -Reducción del ruido de los datos
<p>Reglas de asociación</p> 	<p><i>Apriori</i></p>	<ul style="list-style-type: none"> -Análisis del cesto del mercado -Análisis de enlaces
<p>Clustering (agrupación en racimos)</p> 	<ul style="list-style-type: none"> -Jerárquico K-media -Jerárquico O-cluster (<i>orthogonal partitioning clustering, de Oracle</i>) 	<ul style="list-style-type: none"> -Agrupación de productos -Minería de textos -Análisis de genes y proteínas
<p>Extracción de características</p> 	<p>Factorización no negativa de matrices (<i>non negative matrix factorization</i>)</p>	<ul style="list-style-type: none"> -Análisis de textos -Reducción de características

Figura 1. Esquema de funciones y algoritmos. Adaptado de Berger (2012), p. 32. <http://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataoracleadvancedanalytics11gr2-1930894.pdf>

juntos de datos (*data sets*).

Dentro de esta gestión hay varios problemas de por sí importantes: cómo capturar esos datos y de qué fuentes; dónde almacenarlos, con el costo que conlleva; cómo encontrar la aguja de información en ese pajar de datos y a ser posible, en tiempo real; cómo analizarlo para generar conocimiento práctico con esas agujas de información y, finalmente, cómo representarlo, cómo visualizar ese análisis para que sea comprensible por quienes tienen que tomar las decisiones.

Para esa toma de decisiones, probablemente uno de los cambios experimentados más interesantes viene del hecho de que antes una empresa sólo tenía acceso a sus propios datos financieros, y acaso a algún informe realizado por una consultora sobre un sector completo o hasta cierto punto extrapolable. Ahora puede combinar su información interna con datos del mercado, de lo que se dice de los productos propios y la competencia en internet, de la conducta de los usuarios, de otros estudios sectoriales, de publicaciones científicas, de los datos que abastecen a esas publicaciones científicas, de...

Analítica web

Imaginemos una tienda online. Las hay de todo tipo y tamaño. En principio pensaremos que sólo empresas de internet de la envergadura de *Amazon* o *eBay* harán uso de estas tecnologías (y lo hacen) pero no tiene por qué ser así. Una pequeña tienda online utiliza hoy día ya aplicaciones de analítica web para conocer qué hacen los usuarios que les visitan hasta que compran un producto, cuáles han sido los canales de publicidad y comunicación más eficaces y eficientes, o descubrir fallos de usabilidad o arquitectura de información, que impiden al usuario encontrar y utilizar lo que el responsable del sitio web desearía que encontrara y utilizara.

Sólo entender bien qué es lo que está ocurriendo dentro de su casa ya es un gran adelanto para esa pequeña tienda online, que utiliza programas a veces gratuitos como *Google Analytics* o *Piwik*. Éstos emplean la estadística más básica para analizar los datos de la interacción de los usuarios, acompañándose de gráficos para visualizar o resumir esas interacciones, normalmente en forma de series temporales, gráficos de barras, rankings de contenidos o flujos de navegación.

El problema de la analítica web en ese sentido, no es tanto saber cómo funciona el programa para explotar todas sus posibilidades (que también) sino saber qué buscar, cómo conectar unos datos con otros para encontrar las razones que llevan a los usuarios a clicar o no en un enlace, en un botón de compra, o a realizar las acciones que el dueño del sitio web desearía que se produjesen, lo que podríamos denominar en inglés *better queries* (mejores preguntas) en contraposición a más datos o mejores algoritmos, que pa-

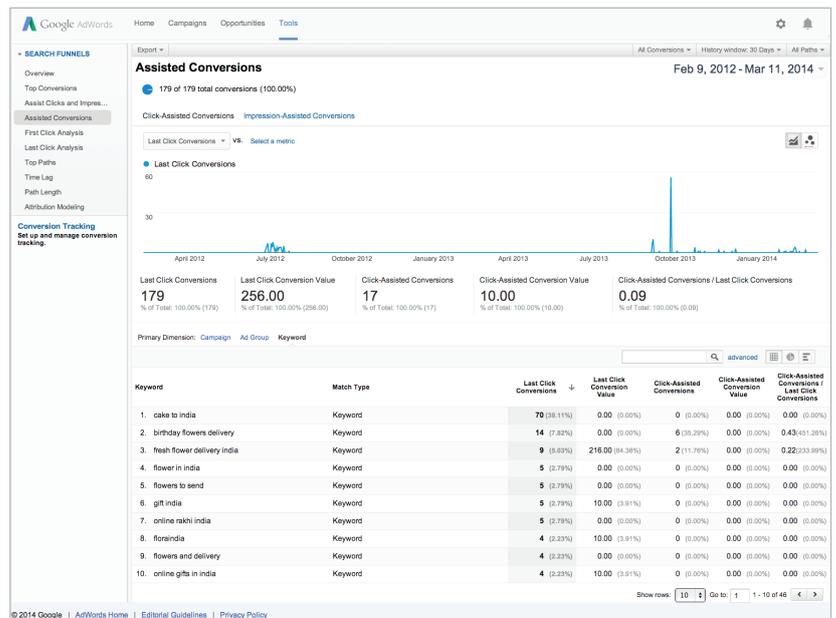


Figura 2. Ejemplo de análisis de conversiones asistidas por más de un canal. Fuente: *Google.com*

rece ser un tema de discusión recurrente en la bibliografía relacionada. Porque el análisis se hace más rico al cruzar distintos indicadores, y de no hacerlo así, podemos fácilmente llegar a conclusiones arriesgadas.

Por ejemplo, sigamos imaginando que esa tienda online quiere mejorar su número de productos vendidos, y para ello realiza diversas acciones de marketing en distintos canales (redes sociales, publicidad en buscadores, publicidad gráfica, email marketing...), que va midiendo. Si contrasta el número de usuarios que llegan al sitio web desde cada canal con las ventas producidas directamente tras llegar de ese canal (lo que en *Google Analytics* se conoce como *last click attribution model*), y llega a la conclusión de que sólo uno de esos canales tiene correlación muy fuerte con un aumento número de ventas inmediatas, corre el riesgo de deducir que debe eliminar su gasto en los otros canales, debido a un típico error, asimilar correlación con causalidad. Y aunque ya sabemos que correlación no implica causalidad, los indicadores nada correlacionados ciertamente no parecen ser causa de ventas directas.

« Uno puede tener la tentación de predecir hacia dónde irán las cosas. Y eso es lo que todo el mundo intenta hacer: lo llaman '*predictive analytics*' »

Y ese es el problema en este caso, analizar sólo las ventas directas (las que se realizan directamente tras clicar en un banner, por ejemplo). Quizá, si la tienda online utilizara un modelo de atribución que estudiara cómo cada canal asiste a los demás a convencerse de que es aquí donde debe realizar su compra, ayudando al usuario que viene al portal, se va a comparar un producto en otros sitios web, consulta foros o descubre opiniones en redes sociales, cambiaría su interés en desechar ciertos canales, por encontrar acaso

indicios de que los distintos canales se ayudaban más de lo que la correlación en ventas directas sugería.

Asimismo, otro problema añadido será adaptar los datos que aporta la aplicación al contexto de uso, a partir de unos objetivos contra los que medirse o unos competidores con los que compararse. Porque el mismo dato puede significar cosas distintas, dependiendo del contexto. Por ejemplo, tomemos el concepto de *engagement*, muy usado en marketing y en gestión de comunidades o redes sociales. En sí *engagement* (para esa tienda online) alude a la implicación o grado de relación entre usuario y marca. Según *Google Analytics*, el *engagement* se mide en función de “cuánto tiempo permanecen los usuarios en su sitio (en segundos) y el número de páginas que visualizan”. Pero un sitio web puede entender que debe medir de otra forma más rica ese compromiso o implicación de los usuarios en su grado de interacción alrededor de un sitio web o de una marca, y por tanto, seleccionar qué fuentes (externas tipo *Twitter*, internas tipo logs de navegación), datos e indicadores va a utilizar para determinar si una marca, un servicio de información o una tienda online consigue que sus usuarios sean más fieles, e interactúen más y mejor con ésta, de forma que rentabilicen esas interacciones, dependiendo (de nuevo) en cada caso de cómo defina “rentabilidad”.

“ Los datos de navegación dentro del sitio web (el reino de la analítica web) se suman ahora a los datos que hay en internet (cibermetría)”

De la analítica web a la cibermetría, *web science*...

Por tanto en un entorno tan competitivo como el actual, ese conocimiento interno probablemente ya no basta. El usuario puede tener implicación para con la marca o sitio web no sólo visitando éste, sino hablando bien (o mal) de él, compartiendo sus contenidos o productos, liderando opiniones... Hace falta pues dotarse de nuevas herramientas para integrar la diversidad de datos que pueden usarse en los distintos análisis, descubrir y gestionar mayor variedad de fuentes de datos, y mejorar la velocidad de procesamiento de los mismos, para entender mejor el escenario en el que ese servicio, proyecto o producto web se mueve.

Los datos de navegación dentro del sitio web (el reino de la analítica web) se suman ahora a los datos que hay en internet (cibermetría) respecto de:

- nichos de mercado online;
- conductas de búsqueda de los usuarios en otros sitios web y en distintos buscadores;
- microsegmentación de esos potenciales clientes;
- la competencia;
- relaciones entre esos competidores y otros sitios web medidas en forma de enlaces;
- relaciones entre las marcas y/o productos y los usuarios que los consumen medidas en forma de comentarios, retweets, respuestas, likes...;

- noticias que hay relativas a las temáticas de esos productos;
- innovación científica que permite intuir hacia dónde va la propuesta de valor única tecnológica de la competencia;
- etcétera, etcétera.

Aquí entramos en el reino de la inteligencia competitiva y la vigilancia tecnológica, en la que necesitaremos la combinación de disciplinas que trabajen con gran cantidad y variedad de datos (*big data*) como la cibermetría, *business intelligence*, *data mining* o *text mining*.

Esta variedad y variabilidad de los datos es un problema en constante evolución. Cuantas más fuentes de datos y más cambiantes, a menudo debido a razones difíciles de detectar, más ángulos de visión tendremos sobre un problema dado, más factores potencialmente causales podremos emplear en los estudios, y más probabilidades habrá de que los análisis sean más ricos, pero más difícil será detectar las fuentes de datos más útiles. La capacidad para lidiar con la complejidad inherente a esta conjugación de datos de fuentes heterogéneas es una de las áreas en las que más está avanzando, por ejemplo gracias no sólo a los repositorios de datos científicos, sino además, a que hoy día se comparten fuentes de datos a través de internet mediante literalmente miles de APIs, que ya incluyen *internet of things*, tomando como fuentes de datos los de productos *wearables* (vestibles) o los producidos por una ciudad en proyectos de *smart cities*.

Otra de las áreas que necesitan avanzar es la de la estandarización, como la que realiza el *Data Mining Group*, creando el lenguaje PMML (*predictive model markup language*) que permite que los modelos que se generen sean interoperables, con independencia del sistema con el que han sido construidos.

Mejores datos versus mejores algoritmos

Cisco estima que en 2012 cada día fueron creados cerca de 2,5 trillones de bytes de datos. Con tantos datos al alcance de la mano, uno en principio pensaría que el principal problema es tecnológico: ¿qué herramientas usar a un precio razonable, dónde custodiar tanta información, qué algoritmos emplear para obtener las conclusiones de mis análisis, y con cuánta velocidad antes de que éstas ya no sean útiles? Estas preguntas han dado lugar a una auténtica explosión de empresas dedicadas a crear herramientas de análisis, sistemas de almacenamiento, técnicas de programación...

Los campos y sectores económicos de aplicación son prácticamente cualesquiera que imaginemos, desde el periodismo de datos (conocido también como periodismo computacional) a la prevención de pandemias, pero permanece la discusión sobre si utilizar mejores algoritmos o más datos. De hecho, incluso se están estudiando algoritmos que usan muy pocos datos (*little data*) aunque lo hacen sobre segmentaciones muy específicas, a las cuales habrán llegado tras analizar gran cantidad de ejemplos o posibles segmentos, por lo que el problema permanece.

Para enfrentarse a las cantidades masivas de datos hay dos grandes escuelas, dos grandes disciplinas, la estadística y la informática con gran diversidad de técnicas:

- *behavioural analysis* o análisis de la conducta;
- *predictive analytics* o análisis predictivo;
- segmentación;
- *sentiment analysis* (mediante *text mining* a la que se aplica procesamiento del lenguaje natural);
- clasificación;
- agrupamiento o *clustering*;
- aprendizaje supervisado y no supervisado;
- regresión;
- árboles de decisión;
- inferencia;
- reconocimiento de patrones;
- representación del conocimiento;
- cálculo de probabilidades;
- reglas de asociación...

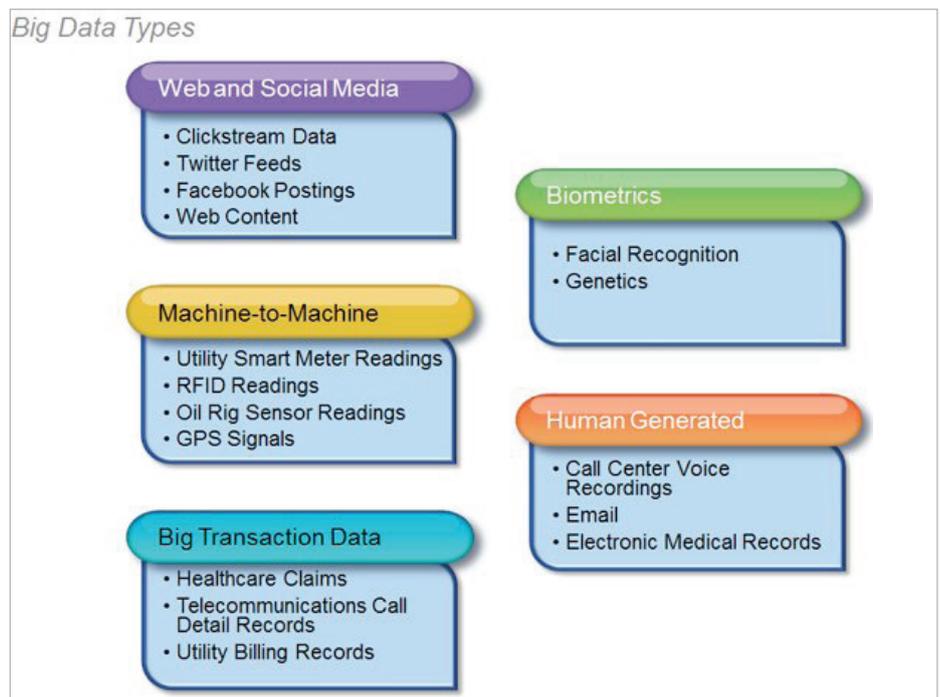


Figura 3. Algunos tipos de *big data*. Fuente: <http://ibm.com>

Actualmente la estadística y la informática se mezclan y tienden a entenderse para contestar distintos aspectos de un mismo problema, aunque no sea nada fácil encontrar a especialistas que abarquen ambas disciplinas enteramente, por lo que las líneas se desdibujan. Probablemente antes debemos entender qué técnicas contestan qué problemáticas, y de ahí aún antes, qué preguntas queremos que sean contestadas.

Conclusión

Hace falta un puente entre lo que necesita un cliente (sea un bibliotecario, un técnico de un ayuntamiento, el CEO de una *start-up* de internet...) y quien conoce cómo obtener respuestas de los datos. Cuando este cliente hable con un especialista en *machine learning* o en estadística, aquél necesita qué, o bien se tengan muy claras las preguntas para las cuales se vislumbra las posibilidades de solución mediante alguna o varias de las técnicas mencionadas, o bien se le forme en la especialidad sectorial o temática, para que desde la suya imagine qué soluciones podría dar sin esperar a la pregunta perfecta. Y es un puente que se podría acortar con más formación... por ambos lados.

Bibliografía

Barranco-Fragoso, Ricardo (2012). *¿Qué es big data?* IBM. <http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data>

Berger, Charlie (2012). *Big data analytics with Oracle advanced analytics in-database option*. <http://www.oracle.com/technetwork/database/options/advanced-analytics/bigdataoracleadvancedanalytics-11gr2-1930894.pdf>

Data Mining Group. *The Data Mining Group releases PMML v 4.2*. <http://www.dmg.org/DMGReleasesPMMLv4.2.pdf>

Garrett, Wu (2013). "Why more data and simple algorithms beat complex analytics models". *DataInformed. Big data and analytics in the enterprise*. <http://data-informed.com/why-more-data-and-simple-algorithms-beat-complex-analytics-models>

Google. *About the attribution models*. <https://support.google.com/analytics/answer/1665189>

Google. *Engagement*. <https://support.google.com/analytics/answer/1144430?hl=en>

MuleSoft. *The top 10 internet of things APIs*. <http://www.mulesoft.com/infographics/api/internet-things>

Winshuttle (2014). *Big data y la historia del almacenamiento de la información*. <http://www.winshuttle.es/big-data-historia-cronologica>

Colección de libros de bolsillo

El profesional de la información (Editorial UOC)



El profesional de la
información



EDITORIAL UOC

Más información:

<http://www.elprofesionaldelainformacion.com/libros.html>

ARTÍCULOS

TECNOLOGÍAS *BIG DATA* PARA ANÁLISIS Y RECUPERACIÓN DE IMÁGENES WEB

Sergio Rodríguez-Vaamonde, Ana-Isabel Torre-Bastida y Estibaliz Garrote



Sergio Rodríguez-Vaamonde es ingeniero de telecomunicaciones por la *Universidad del País Vasco* y diploma de estudios avanzados en ingeniería telemática. Doctorando en anotación automática y recuperación de imágenes en grandes colecciones de datos, cuenta con publicaciones en este ámbito. Trabaja como investigador en el grupo de procesamiento de imagen en el centro tecnológico *Tecnalia*, desarrollando productos de aplicación de las tecnologías de visión artificial a las TICs o el sector industrial.

<http://orcid.org/0000-0003-1982-5128>

sergio.rodriguez@tecnalia.com



Ana-Isabel Torre-Bastida es ingeniera informática por la *Universidad de Deusto* y máster en sistemas informáticos avanzados por la *Universidad del País Vasco*. Es doctoranda, centrando sus investigaciones en web semántica y *linked open data*, e investigadora colaboradora en el centro tecnológico *Tecnalia*, donde lleva a cabo trabajos de aplicación de las tecnologías semánticas y de *big data* a varios sectores.

<http://orcid.org/0000-0003-3005-1100>

isabel.torre@tecnalia.com



Estibaliz Garrote es doctora y licenciada en ciencias físicas e ingeniera en electrónica por la *Universidad del País Vasco*. Como especialista en aplicaciones de visión artificial ha participado en más de 50 proyectos de investigación nacional e internacional sobre control de calidad, procesos de fabricación, biometría, reciclado y accesibilidad. En los últimos años ha trabajado en el desarrollo de modelos neuroinspirados, colaborando en los grupos de Tomaso Poggio (*MIT*), John Mollon (*University of Cambridge*) y Thomas Serre (*Brown University*).

<http://orcid.org/0000-0002-0268-0391>

estibaliz.garrote@tecnalia.com

Tecnalia. División Industria y Transporte
Parque Tecnológico de Bizkaia
Ibaizabal Bidea, Edif. 202. 48170 Zamudio (Bizkaia), España

Resumen

Se aborda el análisis web desde el punto de vista de las imágenes, empleando tecnologías *big data*. Las imágenes cada vez tienen más peso en la web por lo que cualquier análisis que se realice deberá considerar este tipo de información. Los grandes volúmenes de imágenes existentes hacen necesaria la utilización de grandes infraestructuras de computación para realizar este tipo de trabajos, así como tecnologías de visión artificial específicas. Se muestran tecnologías *big data* que pueden ser utilizadas dentro del campo del análisis de imágenes a gran escala. Además, se propone una arquitectura que permite recuperar imágenes de una biblioteca de imágenes de forma eficiente y con un bajo coste computacional. Esta arquitectura puede servir como base para los análisis web e investigaciones que requieran un estudio detallado de las imágenes similares, sin la necesidad de disponer de hardware específico para ello.

Palabras clave

Big data, Imágenes, Procesamiento de imágenes, Visión artificial, Búsqueda, Búsqueda de imágenes, *Map-Reduce*, *LSH*, *Locality-sensitive hashing*.

Title: *Big data* technologies for image retrieval and analysis in web environments

Abstract

This paper addresses web analytics from the point of view of images, using big data technologies. Images are increasingly present on the web, and therefore any web analysis must consider them. The huge volume of available images requires the use of large computation infrastructures in order to generate this kind of analysis, as well as specific computer vision algorithms. Big data technologies are discussed that can be used for large-scale image processing. In addition, a novel distributed algorithm is proposed that can efficiently retrieve images from an online collection at low computational cost. This algorithm can be used in web analysis or future research requiring detailed image study, without requiring any special hardware.

Keywords

Big data, Images, Image processing, Computer vision, Search, Image search, *Map-Reduce*, LSH, Locality-sensitive hashing.

Rodríguez-Vaamonde, Sergio; Torre-Bastida, Ana-Isabel; Garrote, Estibaliz (2014). "Tecnologías big data para análisis y recuperación de imágenes web". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 567-574.

<http://dx.doi.org/10.3145/epi.2014.nov.02>

1. Introducción

Vivimos en un mundo totalmente digitalizado donde el número de usuarios, sensores y dispositivos electrónicos aumenta a cada segundo, en el que se estima que hay más teléfonos móviles conectados a internet que ordenadores (Martin *et al.*, 2013). Cada proceso digital o intercambio de información que éstos producen, genera a su vez una cantidad ingente de datos que se va sumando a la nube ya existente. Este fenómeno se ha denominado *big data* y es un concepto que se usa para referirse a los datos, a los retos y características especiales que éstos engloban y a las nuevas tecnologías desarrolladas para poder tratarlos. La magnitud del fenómeno es tal que los datos generados durante dos días en 2011, por ejemplo, fueron más que los acumulados desde el origen de la civilización hasta principios de 2003 (Lyman; Varian, 2003). Estas magnitudes asustan y por ello la comunidad científica lleva mucho tiempo buscando soluciones al problema de cómo tratarlos.

«Asumir el problema del *big data* es enfrentarse a tres retos fundamentales: almacenamiento, procesamiento y acceso»

El concepto de *big data* engloba grandes cantidades de datos de distintos dominios (que pueden ser complejos, crecientes y variables), junto con las técnicas necesarias para poderlos recolectar, almacenar, gestionar y analizar. Esta definición amplía la inicialmente establecida por Gartner en 2012 (Beyer; Laney, 2012). Para cualquier organización que se enfrenta al problema de tratamiento de sus datos, *big data* representa la frontera que hay que traspasar para hacer efectiva su gestión en tiempo y coste. Cualquiera que asuma este problema se enfrenta a tres retos fundamentales a resolver: almacenamiento, procesamiento y acceso.

Los datos recolectados pueden clasificarse según su naturaleza, formato y estructura en dos grandes grupos: estructurados, que siguen un modelo que proporciona una metainformación que ayuda en el procesamiento; y no estructurados o datos en crudo, que son más arduos de procesar y por lo tanto su análisis conlleva más tiempo y esfuerzo. En este segundo grupo se encuentran las imágenes y los

vídeos, un tipo de información que cada vez está tomando más relevancia en la Web (Rodríguez-Vaamonde; Ruiz-Ibáñez; González-Rodríguez, 2012). Un ejemplo que muestra la tendencia de crecimiento del volumen de imágenes en internet son los 60 millones de fotografías intercambiadas en la red social *Instagram* cada día (*Instagram*, 2014).

2. Big data y la recuperación de imágenes web por su contenido

Desde una perspectiva *big data*, en el ámbito web muchas veces es necesario recuperar una imagen de una biblioteca distribuida. En el análisis de una determinada página web son muchos los escenarios que se pueden encontrar: es posible requerir la obtención de la fuente de las imágenes que aparecen en dicha página, también es posible analizar si sus imágenes han sido publicadas en terceras páginas, o reconocer a una persona en las fotografías de dicha web.

En todos estos casos el problema central es analizar el contenido de las imágenes de interés localizadas en una web para buscar y encontrar una imagen en una biblioteca de imágenes determinadas. Como se puede intuir, cuando se está hablando de grandes cantidades de datos la búsqueda en miles o millones de imágenes será una búsqueda computacionalmente cara y difícil de ejecutar en equipos informáticos básicos.

La pregunta clave en este contexto es si la tecnología evoluciona y crece a la misma velocidad que lo están haciendo estos datos. Para el campo en el que se centra este artículo, el análisis de imágenes, la respuesta es no. Trabajos como los de Guo y Dyer (2005) o White *et al.* (2010), explican como la infraestructura (recursos de almacenamiento y computación necesarios) es difícilmente adquirible para llevar a cabo aplicaciones de análisis de imagen a gran escala, por lo tanto existen pocos investigadores que se aventuren en esta área. Esto a su vez provoca que el número de trabajos dedicados al procesamiento de grandes volúmenes de imágenes o vídeos sea relativamente escaso y monopolizado por grandes grupos de investigación.

El tratamiento de la imagen, por su relevancia y complejidad, debe tener un espacio propio en el mundo *big data*. Si prosiguen las tendencias actuales y la sentencia de que "los datos son el nuevo petróleo del siglo XXI" acuñada por

Andreas Weigend es cierta, el potencial que se pueda extraer de estos datos dependerá irremediamente de las tecnologías y algoritmos desarrollados para ello, y en el caso de las imágenes está claro que se necesita potenciar ambos.

Para ello presentamos en primer lugar un resumen de las principales tecnologías *big data* existentes y su aplicabilidad a los datos en formato imagen. Abarcar el 100% de las tecnologías queda fuera del alcance de este estudio, por lo que nos centraremos en el caso particular de la recuperación rápida de imágenes en base a su contenido y presentaremos una arquitectura que contiene los ingredientes necesarios para hacer búsquedas rápidas de imágenes que permitan un amplio abanico de análisis web sin necesidad de una gran inversión en infraestructura.

En el caso de las imágenes, la tecnología no evoluciona y crece a la misma velocidad que lo están haciendo los datos

3. Tecnologías *big data* aplicables a imágenes

Las necesidades que deben cumplir las tecnologías del *big data* se basan en el procesamiento eficiente de grandes cantidades de datos con un tiempo reducido o tolerable. La complejidad que añade el hecho de que los datos se encuentren en formato imagen es otra variable a considerar.

En general las tecnologías de mayor relevancia para datos no estructurados como imágenes son las bases de datos *NoSQL* (Leavitt, 2010) y los modelos de programación *Map-Reduce* (Bajcsy et al., 2013), ambas relacionadas con el procesamiento de datos en lotes. Por otro lado están los *CEP* (*complex event processing*), los *IMDG* (*in-memory data grids*), o los sistemas de computación distribuida, para el procesamiento de datos en tiempo real. En la tabla 1 se muestra una taxonomía de estas tecnologías en forma de cuadrante ordenado por volumen de datos y tiempos de rendimiento de las tecnologías *big data*.

En los siguientes apartados se detallan las tecnologías más relevantes en los ámbitos anteriores, centrándose en aquellas especialmente útiles para el procesamiento de imágenes a gran escala.

3.1. Bases de datos *NoSQL*

Se pueden definir como una nueva generación de almacenes de datos, que cumplen alguno de estos puntos: ser no relacionales, distribuidos, escalables o de código abierto. Tienen características comunes como:

- esquema libre, no siguen un modelo de datos rígido o uniforme;

Tabla 1. Taxonomía de tecnologías de *big data* aplicables al procesamiento de imágenes

Volumen de datos	Pocos	Muchos
Tiempo real	Análítica stream -Proceso de eventos complejos -Base de datos en memoria	Análítica en tiempo real -Grid de datos en memoria -Plataforma especializada
Lotes	Análítica de operación -OLPT / OLAP -Base de datos relacional	Análítica en lotes -Plataforma <i>Map-Reduce</i> -Base de datos <i>NoSQL</i>

- facilidad de replicación, lo que mejora la tolerancia a fallos y redundancia;
- API de acceso sencilla para el acceso y manipulación de datos;
- capacidad de tratamiento de una enorme cantidad de datos;
- estructura distribuida y escalabilidad horizontal.

Se clasifican en diferentes tipos. Una de las clasificaciones más completas es la de **Stephen Yen** (2009), que se muestra en la tabla 2 junto con diferentes alternativas para cada tipo.

Estos repositorios son la nueva opción de persistencia para aquellos casos en los que las necesidades de distribución y disponibilidad de los datos superen a la necesidad de un esquema complejo y rígido. Es el caso de las imágenes de las páginas web, ya que su generación es distribuida y el acceso a las mismas debe ser altamente disponible.

Las tecnologías de mayor relevancia para imágenes son las bases de datos *NoSQL* y los modelos de programación *Map-Reduce* para el procesamiento de datos en lotes, y *CEP* e *IMDG* para el procesamiento en tiempo real

3.2. Procesamiento en lotes mediante *Map-Reduce*

En segundo lugar se encuentran los modelos de programación para la construcción de aplicaciones distribuidas. El objetivo de estos modelos es ejecutar un procesamiento determinado sobre un conjunto de datos (llamado lote de datos) de forma distribuida en diferentes localizaciones. Los modelos más extendidos de procesamiento en lotes se basan en dos operaciones básicas: *Map* y *Reduce* (Dean; Ghe-

Tabla 2. Clasificación de bases de datos *NoSQL* y productos existentes

Clases de bases de datos <i>NoSQL</i>	Ejemplos
Almacenamiento clave-valor en cache	<i>Memcached, Repcached, Coherence</i>
Almacenamiento clave-valor	<i>KeySpace, Flare, Schema Fre</i>
Almacenamiento clave-valor eventualmente consistente	<i>Dynamo, Voldemort, Dynamite</i>
Almacenamiento clave-valor ordenado	<i>LightCloud, MemCacheDB, Tokyo Tlrant</i>
Servidores de estructuras de datos	<i>Redis</i>
Bases de datos de objetos	<i>SopeDB, DB40, Shoal</i>
Repositorio de documentos	<i>Couch DB, Mongo DB, Scalaris</i>
Repositorio de columnas	<i>Bigtable, Hbase, Cassandra DB, Hipertable</i>

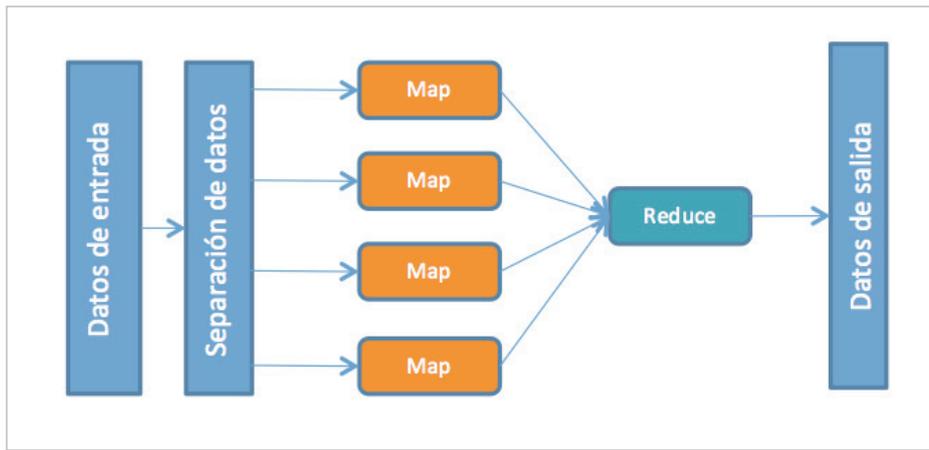


Figura 1. Diagrama de funcionamiento del modelo Map-Reduce

mawat, 2008). En la primera se distribuyen los datos por los nodos del clúster en forma de pares clave-valor y en la segunda se recogen aquellos pares clave-valor que cumplan los criterios del resultado deseado (figura 1).

Uno de los sistemas que ha popularizado esta metodología de trabajo ha sido *Hadoop*, un framework de código abierto para desarrollar y ejecutar aplicaciones distribuidas que procesen una gran cantidad de datos. Su simplicidad y fácil acceso le proporcionan una ventaja en el desarrollo y ejecución de grandes programas distribuidos. La forma de trabajo de *Hadoop* consiste en que los clientes envían sus trabajos a la nube, formada por el clúster de máquinas y es esta la que se encarga de su ejecución mediante el modelo *Map-Reduce* siendo transparente para el usuario la forma en que se haga. Es de un mecanismo muy utilizado en tareas que implican trabajos con grandes volúmenes de datos, como pueden ser el razonamiento de información (Urbani et al., 2009) o la extracción de información de las imágenes (Yan; Huang, 2014).

Para un sistema de visión artificial, lo realmente importante es la representación de la imagen en base a un descriptor visual que sea lo más compacto posible, solventando el problema del almacenamiento

3.3. Tiempo real

En este grupo se enmarcan aquellas soluciones capaces de cumplir con los requisitos de temporalidad del análisis de datos. El resultado del análisis debe de ser inmediato, en tiempo real, para que no pierda su valor y quede obsoleto. Aquí se encuentran dos opciones: los CEP (*complex event processing*) y los IMDG (*in-memory data grids*).

Los CEP son una tecnología de red emergente, que permite adquirir información a partir de sistemas distribuidos de mensajes, bases de datos o aplicaciones, todo ello en tiempo real. Permiten a las organizaciones definir, manejar y predecir eventos y condiciones en una red compleja y heterogénea de actores e informantes. Se centran en el proce-

samiento de eventos como un método para rastrear y analizar grandes volúmenes de información sobre diferentes fenómenos y acciones producidas en un contexto.

En los últimos tiempos ha surgido un nuevo concepto denominado *IMDG*, con el que se denomina a las tecnologías capaces de procesar grandes conjuntos de datos con baja latencia en un contexto de tiempo real. Para ello, estas tecnologías paralelizan el almacenamiento de los datos

gracias a su alojamiento particionado en memoria principal (haciendo que los datos estén cerca de la aplicación que los consume).

En este nuevo contexto se combinan técnicas de cacheo distribuido con potentes herramientas de análisis y gestión, para ofrecer una solución completa que permita gestionar grandes cantidades de datos de naturaleza volátil mediante un clúster de ordenadores.

En la tabla 3 se muestran una serie de productos, así como su fabricante y el lenguaje de programación que aceptan.

Esto ofrece grandes posibilidades a los usuarios, ya que la toma de decisiones será en tiempo real, mejorando la productividad y experiencia de uso.

4. Propuesta de uso de tecnologías de big data aplicadas a imagen

Una vez vistas las tecnologías que se pueden emplear, en este apartado se propone una arquitectura software para construir sistemas más complejos y completos de análisis web en base a imágenes. El objetivo de esta propuesta es doble:

- aliviar la carga computacional y ser capaces de realizar búsquedas rápidas sobre grandes colecciones de imágenes;
- demostrar que el uso de tecnologías genéricas *big data* no requiere disponer de grandes infraestructuras de computación para el procesamiento y recuperación de grandes cantidades de imágenes, eliminando la barrera de entrada de numerosos investigadores a este tipo de estudios.

Tabla 3. Ejemplos de sistemas *in-memory data grids* (IMDG)

Producto	Lenguaje	Empresa
Oracle Coherence	Java	Oracle
Ehcache BigMemory	Java	Terracota
GemFire	Java	VMware
JBoss Infinispan	Java	RedHat
Ncache	.net	AlaChisoft
WebSphere eXtreme	Java	IBM

A continuación se definirá el flujo de trabajo mínimo para el análisis del contenido de una imagen, se describirá la arquitectura propuesta para la búsqueda de imágenes a escala web y se mostrará cómo la propuesta puede ser considerada como base para siguientes trabajos en este ámbito ya que mejora en gran medida los sistemas básicos actuales.

4.1. Análisis de imagen, almacenamiento y recuperación

Una imagen se describe como un conjunto de píxeles en formato matricial que contiene información sobre la escena que representa. Este conjunto de píxeles en crudo no permite a los sistemas de computación conocer el contenido, sino que es necesario generar descriptores visuales de la imagen.

Un descriptor visual se puede definir como la representación matemática del contenido de la imagen. Así, el descriptor visual más básico es un histograma de color, donde en un conjunto limitado de valores representa estadísticamente el número de píxeles contabilizados de cada color en una imagen. Además de este, existen múltiples descriptores que tratan de representar una información de mayor nivel semántico como los bordes presentes (Lowe, 2004) o las formas básicas que aparecen en la imagen (Ferrari; Jurie; Schmid, 2010).

Para un sistema de visión artificial, lo realmente importante es la representación de la imagen en base a un descriptor visual que sea lo más compacto y representativo posible. En este punto entra en juego el concepto de almacenamiento. Si bien una imagen puede ocupar unos cientos o miles de kilobytes, un descriptor apenas ocupa unas decenas. El problema relacionado con *big data* es que en grandes bibliotecas de imágenes (como puede ser la propia web), existirán millones de estos descriptores que representan las imágenes y por ello el espacio de almacenamiento requerido será inmenso. En este punto es donde las tecnologías *big data* son de gran utilidad.

Una vez se tiene toda la información almacenada, la clave es el acceso a la misma. En el caso que se está analizando el objetivo es recuperar las imágenes por su contenido y, específicamente, se desea recuperar las que sean lo más similares visualmente a una imagen dada (figura 2).

Para poder realizar esta búsqueda, el procedimiento directo es comparar los descriptores visuales con una métrica que mida el grado de similitud por su contenido visual. Estas métricas no son más que distancias computadas para cada uno de los dos vectores x e y a comparar y algunas de las más utilizadas se presentan en la tabla 3.

El problema de esta aproximación es claro: si se tienen mi-

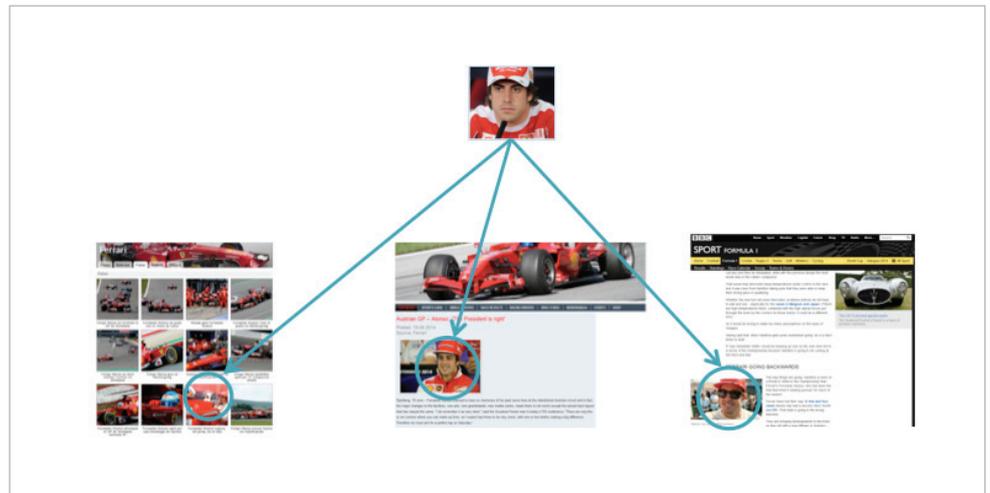


Figura 2. Ejemplo de uso en el que se quiere buscar las páginas web con una fotografía de Fernando Alonso para analizar su aparición en internet

liones de imágenes, comparar y calcular una distancia para cada vector de forma directa es una tarea que llevará una cantidad excesiva de tiempo.

Debido a estos dos principales problemas derivados de los datos y comunes a cualquier sistema *big data*, es necesario utilizar tecnologías adaptadas al uso de grandes cantidades de datos.

4.2. Arquitectura software de recuperación rápida de imágenes similares

En el apartado anterior se ha mostrado que la aproximación directa de la recuperación de imágenes similares no es algo factible cuando se trata de bibliotecas de un gran volumen como pueden ser las relacionadas con el análisis web. Para poder hacer este tipo de análisis y recuperación, se propone la utilización de tecnologías *big data* que permitan un acceso eficiente a toda la información disponible en imágenes. La propuesta se puede resumir en la figura 3.

El primer problema es el almacenamiento eficiente de las imágenes de entrada. Estas imágenes pueden entrar al sistema de diferentes formas, en función del análisis al que se esté dedicando esta arquitectura. En el caso de análisis de imágenes web, la entrada sería por un robot automático que obtuviese las imágenes de las páginas web. En el caso

Tabla 4. Distancias utilizadas para medida de similitud entre descriptores visuales de imágenes

Distancia	Fórmula
Euclídea	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Coseno	$d(x, y) = 1 - \cos(\theta) = 1 - \frac{x \cdot y}{\sqrt{(x \cdot x)(y \cdot y)}}$
Chi Cuadrado	$d(x, y) = \frac{1}{2} \sum_{i=1}^n \frac{ x_i - y_i ^2}{x_i + y_i}$

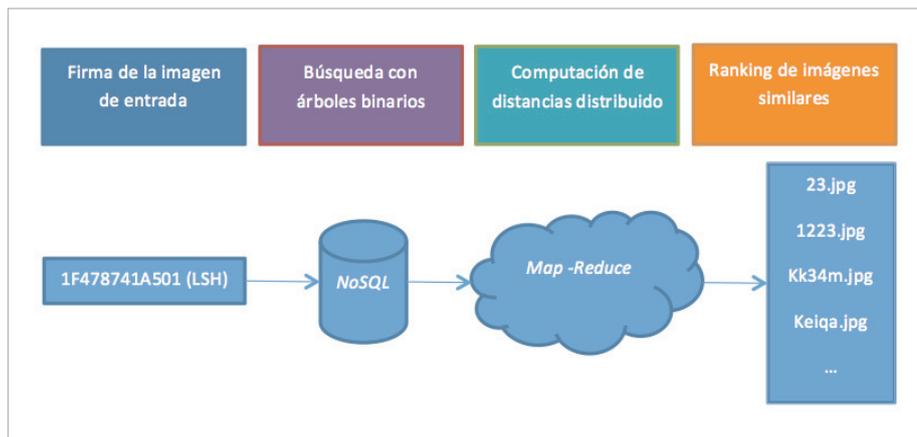


Figura 3. Fases de la arquitectura propuesta para la recuperación de imágenes similares a gran escala

de búsqueda de una imagen concreta, se dispondría de una base de datos de imágenes y habría que introducir manualmente la nueva imagen de consulta.

En cualquier caso el problema principal consiste en almacenar las imágenes y, además de la propia imagen, es crucial almacenar sus descriptores visuales. Ya que esta no es una información estática y es dependiente del análisis o búsqueda a realizar, es necesario disponer de un almacén de datos lo suficientemente flexible. Además se debe poder guardar una gran colección de datos, como son el origen de las imágenes (web, fecha de acceso, etc.), anotaciones manuales o comentarios. Para ello, cualquier base de datos *NoSQL* de las presentadas en la tabla 1 es una opción válida ya que en general permiten disponer de un esquema flexible y son capaces de tener réplicas o nodos distribuidos. Esta última característica la hace ideal para el análisis de imágenes web en cualquier punto geográfico, ya que permite replicas distribuidas en todo el mundo.

Para un sistema de búsqueda por similitud no es factible computar para cada imagen de la colección la distancia a la imagen de consulta; es fundamental utilizar algún sistema de búsqueda aproximada de las más similares

Sobre este almacén de datos *NoSQL* es necesario construir el sistema de recuperación de figuras. No es factible computar para cada imagen de la colección la distancia a la imagen de consulta. Por ello es fundamental utilizar algún sistema de búsqueda aproximada de las más cercanas.

El algoritmo más utilizado en el campo del análisis de imágenes (Kulis; Grauman, 2009) es *locality-sensitive hashing* (LSH) (Slaney; Casey, 2008), por lo que éste será el algoritmo base para la recuperación de figuras. Este algoritmo permite generar una firma numérica (o *hash*) para cada descriptor o conjunto de descriptores de imagen, de tal forma que aquellos vectores que tengan una distancia euclídea muy baja, y por tanto sean vectores muy similares, posean la misma firma numérica. Este tipo de algoritmos es

muy útil para encontrar entradas similares dentro de grandes colecciones de datos, por ejemplo buscando páginas web similares (Slaney; Casey, 2008), por lo que es lógica su aplicación a los descriptores visuales de imágenes.

Una vez se tienen las firmas para todas las imágenes de la colección, la búsqueda de las similares es sencilla: dada una imagen de consulta, se generará su *hash*. Con cada firma se buscará en toda la base de datos las que posean la misma firma y todas ellas serán las imágenes más similares.

Para hacer esta comparación, se puede pensar en que existe el mismo problema de búsqueda que antes, pero nada más lejos de la realidad. Las bases de datos *NoSQL* actuales para *big data*, disponen de técnicas de indexación y búsqueda rápida de un número único, como puede ser el algoritmo de búsqueda en árbol binario de la base de datos *NoSQL MongoDB* o el uso de cualquier tecnología *IMDG* de las propuestas. Por ello la búsqueda ya no se circunscribe a calcular una distancia entre vectores sino a usar una arquitectura de índices para encontrar un número concreto.

Tras este paso ya se dispone de un conjunto de imágenes de la biblioteca similares a la de la entrada. En función de la analítica web que se esté ejecutando, quizá conocer este número es suficiente. En muchos casos la recuperación de figuras tiene como objetivo aquella que más se parece a la de entrada. En este caso, es obligatorio calcular la distancia concreta, pero ya que se ha obtenido un conjunto de imágenes parecidas, se puede calcular la distancia sobre ese conjunto de unas decenas de imágenes similares en unos pocos segundos, en vez de sobre el total de la biblioteca.

Para este último paso, también se va a aprovechar el almacén de datos *NoSQL* distribuido propuesto en el inicio. Ya que las figuras pueden estar almacenadas en localizaciones diferentes y que cada cálculo de la distancia entre la imagen de consulta y la similar es independiente, es posible usar el paradigma *Map-Reduce* expuesto con anterioridad. Este modelo permitirá en la función *Map* el cálculo de la distancia entre cada imagen similar y la de consulta, ejecutándose en cada nodo de la red distribuida de almacenamiento. Por otro lado, el método *Reduce* se encargará de ordenar todas las distancias y podrá generar el ranking final de imágenes similares útiles para la analítica.

5. Resultados y conclusiones

En este artículo se ha mostrado cómo es necesario aplicar tecnologías de *big data* cuando se trata del almacenamiento y acceso de grandes volúmenes de imágenes, como es el caso del análisis de imágenes en la web. Se han expuesto un conjunto de tecnologías que potencialmente pueden ser aplicables a este campo y también se ha realizado una propuesta de arquitectura software que puede ser utiliza-

da en cualquier análisis web que requiera conocer imágenes presentes en las páginas web.

Para comprobar si esta arquitectura es la más adecuada, se ha realizado un experimento comparativo de diferentes métodos de recuperación de imágenes similares del estado del arte. Para ello, se ha utilizado la base de datos *ImageNet* (Deng et al., 2009), que posee más de 14 millones de imágenes. De ellas se han generado diferentes subconjuntos aleatorios, en los que se han agrupado las imágenes de cada una de las 7 pruebas realizadas. En cada prueba el tamaño del conjunto de imágenes que conformaba la base de datos ha sido incremental: empezando con 100, cada prueba ha añadido 200.000 adicionales hasta 1,2 millones. Con estas bases de datos se han ejecutado 10 búsquedas de imágenes similares y se ha medido el tiempo (en segundos) que se ha tardado en computar la distancia a todos los documentos, almacenando el valor del tiempo medio en esas 10 imágenes. Todas las pruebas se han realizado en un único PC, con 16 núcleos de computación a 3GHz, con 32GB de memoria RAM.

Esta operación de búsqueda se ha realizado para cuatro algoritmos, tres propuestos en el estado del arte, y el cuarto es la propuesta realizada:

- el primer algoritmo es el utilizado por la mayor parte de los trabajos, y trata de cargar en memoria RAM todas las imágenes y computar la distancia de forma exhaustiva (Saratxaga et al., 2014);
- el segundo trabajo es computar la distancia exhaustiva utilizando el algoritmo *Map-Reduce* (Bajcsy et al., 2013) donde cada nodo es un núcleo de procesamiento del PC;
- el tercero es utilizar la aproximación LSH de forma directa con los *hashes* cargados en memoria RAM (Slaney; Casey, 2008);
- la propuesta que realizamos es utilizar LSH para representar las imágenes, un *Binary Tree* para hacer una búsqueda aproximada y posteriormente *Map-Reduce* para obtener de forma precisa un subconjunto de distancias.

En todos ellos se ha comprobado cómo el resultado de las 10 imágenes más similares de la base de datos es el mismo, por lo que sólo varía el tiempo de ejecución, no la precisión de la búsqueda.

En la figura 4 se muestra cómo el uso de la aproximación propuesta es muy útil cuando aumenta el número de imágenes.

En el eje horizontal, se ven diferentes números de imágenes que se han introducido en la bases de datos de validación. En el eje vertical se ve el tiempo en segundos que se ha tar-

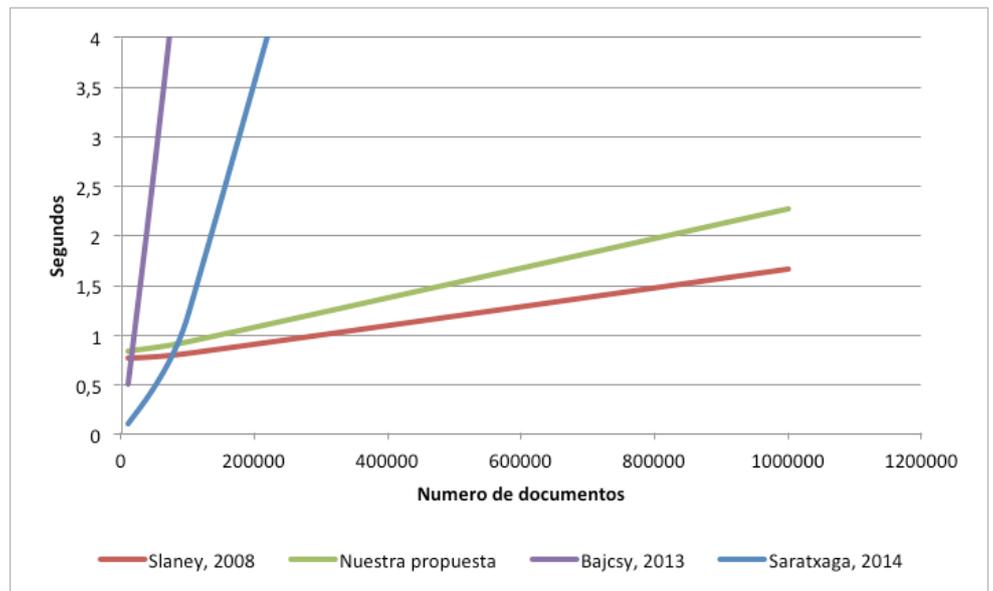


Figura 4. Comparativa de tiempos medios de búsqueda de una imagen usando cuatro aproximaciones diferentes

dado de media en buscar una única imagen similar dentro de la biblioteca de validación.

El resultado muestra que el tiempo de búsqueda en las técnicas de búsqueda exacta aumenta de forma exponencial con el número de figuras almacenado, mientras que en las búsquedas aproximadas usando LSH el aumento es lineal. Como es de esperar, el acceso a memoria es mucho más rápido que el uso de árboles de indexación distribuidos, por ello la aproximación de LSH en memoria (Slaney; Casey, 2008) es la que menos tiempo tarda. Aun así, esto no es eficiente para los análisis web distribuidos, ya que la multitud de nodos hace que no sea viable mantener todo en memoria. Por ello, la aproximación distribuida propuesta es la más útil para este tipo de analíticas.

La conclusión de este artículo está claramente alineada con el impulso de la utilización de tecnologías *big data* en el campo del análisis de las imágenes para recuperación a escala web. La motivación principal es que cada vez existen más datos en forma de imágenes y videos, y su investigación y análisis se monopoliza en los grandes grupos de investigación con grandes infraestructuras dedicadas al manejo de estos datos. En este artículo se muestra cómo las tecnologías *big data* más populares se pueden aplicar a la búsqueda de documentos en formato de imagen. Por tanto el uso de herramientas estándar, principalmente de código abierto, y optimizaciones de las mismas generan que sea posible utilizar sistemas de computación clásicos, y no necesariamente grandes clústeres de computación, para la analítica de estos datos.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno Vasco dentro del proyecto *Smartur* bajo el programa de investigación fundamental *Etortek*. También ha sido parcialmente financiado por la Comisión Europea bajo el proyecto *Biopool*. Los autores quieren agradecer el trabajo a todos los socios de este proyecto.

<http://www.biopoolproject.eu>

6. Bibliografía

- Bajcsy, Peter; Vandecreme, Antoine; Amelot, Julien; Nguyen, Phuong; Chalfoun, Joe; Brady, Mary** (2013). "Terabyte-sized image computations on Hadoop cluster platforms". En: *IEEE Intl conf on big data*, pp. 729-737.
<http://dx.doi.org/10.1109/BigData.2013.6691645>
- Beyer, Mark A.; Laney, Douglas** (2012). *The importance of big data: A definition*. Gartner.
<http://www.gartner.com/id=2057415>
- Dean, Jeffrey; Ghemawat, Sanjay** (2008). "MapReduce: simplified data processing on large clusters". *Communications of the ACM*, v. 51, n. 1, pp. 107-113.
<http://static.googleusercontent.com/media/research.google.com/es//archive/mapreduce-osdi04.pdf>
- Deng, Jia; Dong, Wei; Socher, Richard; Li, Li-Jia; Li, Kai; Fei-Fei, Li** (2009). "ImageNet: A large-scale hierarchical image database". *IEEE Computer vision and pattern recognition (CVPR)*.
<http://dx.doi.org/10.1109/CVPR.2009.5206848>
- Ferrari, Vittorio; Jurie, Frederic; Schmid, Cordelia** (2010). "From images to shape models for object detection". *Intl journal of computer vision*, v. 87, n. 3, pp. 284-303.
https://lear.inrialpes.fr/pubs/2010/FJS10/vitto_final_ijcv.pdf
<http://dx.doi.org/10.1007/s11263-009-0270-9>
- Guo, Guodong; Dyer, Charles R.** (2005). "Learning from examples in the small sample case: face expression recognition". *IEEE Transactions on systems, man, and cybernetics, Part B: Cybernetics*, v. 35, n. 3, pp. 477-488.
<ftp://ftp.cs.wisc.edu/computer-vision/repository/PDF/guo.2005.smc.pdf>
<http://dx.doi.org/10.1109/TSMCB.2005.846658>
- Instagram (2014). *Instagram press stats*.
<http://instagram.com/press>
- Kulis, Brian; Grauman, Kristen** (2009). "Kernelized locality-sensitive hashing for scalable image search". En: *IEEE 12th Intl conf on computer vision*, pp. 2130-2137.
http://www.cs.utexas.edu/~grauman/papers/iccv2009_klsh.pdf
<http://dx.doi.org/10.1109/ICCV.2009.5459466>
- Leavitt, Neal** (2010). "Will NoSQL databases live up to their promise?". *IEEE Computer*, v. 43, n. 2, pp. 12-14.
<http://leavcom.com/pdf/NoSQL.pdf>
- Lowe, David G.** (2004). "Distinctive image features from scale-invariant keypoints". *Intl journal of computer vision*, v. 60, n. 2, pp. 91-110.
<http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>
<http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- Lyman, Peter; Varian, Hal R.** (2003). *How much information?* Berkeley, CA: University of California at Berkeley, School of Information Management and Systems.
<http://www2.sims.berkeley.edu/research/projects/how-much-info/how-much-info.pdf>
- Martín, David; López-de-Ipiña, Diego; Alzua-Sorzabal, Aurenkene; Lamsfus, Carlos; Torres-Manzanera, Emilio** (2013). "A methodology and a web platform for the collaborative development of context-aware systems". *Sensors*, v. 13, n. 5, pp. 6032-6053.
<http://www.mdpi.com/1424-8220/13/5/6032>
<http://dx.doi.org/10.3390/s130506032>
- Rodríguez-Vaamonde, Sergio; Ruiz-Ibáñez, Pilar; González-Rodríguez, Marta** (2011). "Uso combinado de tecnologías semánticas y análisis visual para la anotación automática de imágenes y su recuperación". *El profesional de la información*, v. 21, n. 1, enero-febrero, pp. 27-33.
http://www.computervisionbytecnalia.com/wp-content/uploads/2012/12/2012_JointUseSemantics.pdf
<http://dx.doi.org/10.3145/epi.2012.ene.04>
- Saratxaga, Cristina L.; Picón, Artzai; Rodríguez-Vaamonde, Sergio; López-Carrera, Ángel; Echazarra, Jone; Bereciartua, Arantza; Garrote, Estibaliz** (2014). "Plataforma de búsqueda de imágenes histológicas por similitud visual". En: *XVII Congreso nacional de informática de la salud*.
<http://www.computervisionbytecnalia.com/wp-content/uploads/2014/04/PlataformaBusquedadelmagenesHistologicas.pdf>
- Slaney, Malcolm; Casey, Michael** (2008). "Locality-sensitive hashing for finding nearest neighbors". *IEEE Signal processing magazine*, v. 25, n. 2, pp. 128-131.
<http://web.iitd.ac.in/~sumeet/Slaney2008-LSHTutorial.pdf>
- Urbani, Jacopo; Kotoulas, Spyros; Oren, Eyal; Van Harmelen, Fran** (2009). "Scalable distributed reasoning using MapReduce". En: *ISWC '09 Procs of the 8th Intl semantic web conf*, pp. 634-649.
<http://www.few.vu.nl/~jui200/papers/ISWC09-Urbani.pdf>
http://dx.doi.org/10.1007/978-3-642-04930-9_40
- White, Brandyn; Yeh, Tom; Lin, Jimmy; Davis, Larry** (2010). "Web-scale computer vision using mapreduce for multimedia data mining". En: *Procs of the 10th Intl workshop on multimedia data mining*, pp. 1-10.
<http://www.umiacs.umd.edu/~lsd/papers/brandyn-kdd-cloud.pdf>
<http://dx.doi.org/10.1145/1814245.1814254>
- Yan, Yuzhong; Huang, Lei** (2014). "Large-scale image processing research cloud". En: *Cloud computing 2014, The 5th Intl conf on cloud computing, grids, and virtualization*, pp. 88-93.
http://www.thinkmind.org/index.php?view=article&articleid=cloud_computing_2014_4_20_20069
- Yen, Stephen** (2009). "NoSQL is a horseless carriage". NoSQL Oakland.



EL *BIG DATA* TRANSFORMA LA INTERPRETACIÓN DE LOS MEDIOS SOCIALES



Silvia Martínez-Martínez y Pablo Lara-Navarra



Silvia Martínez-Martínez, doctora europea en comunicación por la *Universidad CEU Cardenal Herrera*, es profesora y coordinadora de posgrado en los *Estudios de Ciencias de la Información y de la Comunicación* de la *Universitat Oberta de Catalunya*.

<http://orcid.org/0000-0002-6666-7954>

smartinezmartinez1@uoc.edu



Pablo Lara-Navarra es licenciado en documentación por la *Universidad de Granada*, doctor por la *Universitat Pompeu Fabra*, y master en sociedad de la información y conocimiento por la *Universitat Oberta de Catalunya (UOC)*. Ha sido director de *Innovación* de la *UOC*. Es profesor de los *Estudios de Ciencias de la Información y de la Comunicación* de la *UOC* y asesor de empresas tecnológicas.

<http://orcid.org/0000-0003-0595-3161>

plara@uoc.edu

Universitat Oberta de Catalunya

Rambla de Poblenou, 156. 08018 Barcelona, España

Resumen

En la sociedad de la información, donde el volumen de datos crece de forma exponencial, la eclosión del *big data* ha impactado en ámbitos diversos. La popularidad del término ha desdibujado las fronteras de un concepto que no sólo incide en la dimensión sino también en el valor de los datos recopilados y procesados. Los *social media*, caracterizados por su rápida expansión y por la variedad de interacciones y de contenidos que en ellos circulan, se han incorporado al estudio del *big data* al convertirse en fuente de datos útiles para investigadores, entidades y empresas. Se observan las singularidades propias del análisis de los medios sociales dentro del *big data* y, tras escrutar recursos de medición existentes, se reflexiona sobre la oportunidad y necesidad de una solución que permita examinar tendencias y el impacto de perfiles que configuran amplias redes de usuarios.

Palabras clave

Big data, Medios sociales, Comunicación, Información, Conocimiento, Perfiles de influencia, Pirámide informacional, Herramientas de medición, Monitorización.

Title: Big data transforms the interpretation of the social media

Abstract

In the information society where the volume of data grows exponentially, the blossoming of the big data has had an impact on several fields. The popularity of the expression has blurred the outlines of the concept that not only insists on the size but also on the value of the collected and processed data. Social media, which are characterized by a rapid expansion and by the diversity of interactions and contents that circulate in them, have been incorporated to the big data studies to become resource of useful data for researchers, organizations and companies. This paper observes the typical singularities of the social media analysis inside the big data and, after scrutinizing existing measuring tools, reflects on the opportunity and the necessity of a solution to examine trends and the impact of the profiles that get to set extensive networks.

Keywords

Big data, Social media, Communication, Information, Knowledge, Influencers, Informational pyramid, Measuring tools, Monitoring.

Martínez-Martínez, Silvia; Lara-Navarra, Pablo (2015). "El *big data* transforma la interpretación de los medios sociales". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 575-581.

<http://dx.doi.org/10.3145/epi.2014.nov.03>

Artículo recibido el 30-07-2014

Aceptación definitiva: 07-10-2014

1. Introducción: definición, alcance e impacto del big data

El volumen de datos generados crece exponencialmente. Sólo en los dos últimos años la cifra mundial ha aumentado un 92% (West, 2013). La sociedad asiste a la eclosión de la era del *big data* (Villars; Olofson; Eastwood, 2011) con la generación de colecciones de datos estructurados y no estructurados en tiempo real o diferido (Bowden, 2014). Este fenómeno no se caracteriza sólo por la magnitud o el tamaño del conjunto de datos disponible sino que posee otros atributos. Russom (2011) identifica tres elementos que sintetiza bajo la expresión de las tres V's: volumen, velocidad (por la frecuencia o inmediatez con la que se generan datos actuales o atemporales) y variedad, que hace referencia a la diversidad de tipos de datos que se manejan. Algunos autores añadirían una cuarta V para incluir el valor de los datos, pues de ellos se debe poder extraer información valiosa (Zaslavsky; Perera; Georgakopoulos, 2012).

Los datos por sí mismos pueden no aportar nada o incluso generar ruido, especialmente si se encuentran en volúmenes elevados, por lo que su procesado se hace indispensable. Éste consiste en su extracción y análisis para "adquirir o descubrir conocimiento" (De-la-Rosa-Troyano; Martínez-Gasca, 2007, p. 7). Ésta es una idea que remite a la pirámide informacional (figura 1) que se puede presentar, según señala García-Marco (2011, p. 14), como un modelo válido, una metáfora fértil aplicable a campos transdisciplinares que aborden "los sistemas de información en un sentido amplio".

Esta metáfora visual lleva a una reflexión más amplia que pone de relieve que el impacto del *big data* se mide en términos cuantitativos y también cualitativos. Mayer-Schönberger y Cukier (2013, pp. 22-26) subrayan los cambios que introduce en la forma de comprender la sociedad ya que permite el acercamiento a lo granular, ganar en percepción macro y descubrir tendencias, pautas y correlaciones que permiten hacer predicciones en sectores tan variados como los negocios o la salud.

La explotación de los datos puede suponer un aumento de la innovación, la eficiencia y la productividad (Tene; Polonetsky, 2013, p. 29). La tecnología ha avanzado en los últimos

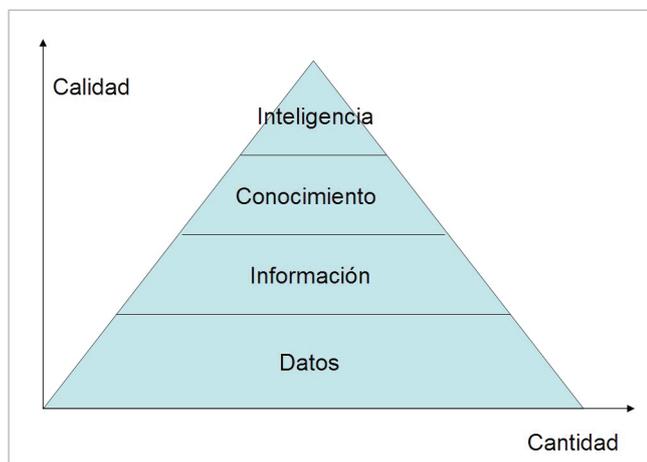


Figura 1. Pirámide informacional. Fuente: Ponjuán (1998)

años en capacidad y abaratamiento de costes permitiendo y mejorando las posibilidades de almacenamiento, recuperación, análisis y visualización pero todavía se enfrenta a algunos retos en su aplicación a distintas disciplinas y ámbitos. Uno de ellos es el análisis de las redes sociales (Magnusson, 2012, p. 53).

El impacto del *big data* no sólo se mide en términos cuantitativos sino también cualitativos

2. Los medios sociales como objeto del big data

Con la web 2.0 y el mayor protagonismo que adquiere el usuario, aumenta el número de sujetos activos en el intercambio de opiniones e informaciones. En este contexto los medios sociales han tenido un espectacular incremento en los últimos años y han marcado una revolución en los hábitos y posibilidades de comunicación, intercambio y publicación de contenidos. Algunas cifras permiten medir la magnitud de tal crecimiento: Según ComScore, de 2007 a 2011 la audiencia mundial en sitios sociales registra un aumento del 174% y en junio de 2012 sumaban 1.258 millones de usuarios (Foldes, 2013). Los datos por plataformas también resultan espectaculares por su dimensión. Por ejemplo, en marzo de 2014 Facebook (2014a) contaba con 802 millones de usuarios activos al día, y la comunidad de profesionales que se conectan a través de LinkedIn (2014) sumaba en abril de ese año, más de 300 millones de miembros. Atendiendo al intercambio de contenidos destacan los 500 millones de tweets al día que registra Twitter (s. f.), las 100 horas de vídeo que cada minuto se suben a YouTube (s. f.) y los 25 mil millones de 'pins' que se contabilizan en Pinterest (s. f.).

Las consecuencias de este incremento de usuarios y de flujo comunicativo inciden de forma directa en la forma de relacionarnos y acceder a contenidos: "Por primera vez podemos seguir los pensamientos, opiniones, ideas y sentimientos de cientos de millones de personas. Podemos ver las imágenes y los vídeos que crean y comentar, monitorizar las conversaciones que mantienen, leer sus blogs y tweets, navegar por sus mapas, escuchar sus listas de música, y seguir sus movimientos en el espacio físico" (Manovich, 2011).

El individuo como ser social experimenta la necesidad de contar y narrar sus vivencias y ahora esa necesidad se traslada a internet gracias a las posibilidades que brindan los medios sociales. La comunicación se intensifica ante grandes acontecimientos de distinta naturaleza, como se ha podido registrar en el uso que se hizo de Twitter tras el terremoto de Japón en 2011 (Watters, 2011) o en el Mundial de fútbol 2014, pues sólo en 18 días de torneo ya se habían alcanzado los mil millones de mensajes en Facebook (2014b) relacionados con la competición. Además de los comentarios publicados, se genera otra información muy valiosa gracias al uso de nuevos dispositivos que facilitan datos de localización geográfica y otra serie de metadatos en el propio contenido generado (Smith et al., 2013). Se trata del fenómeno que Manovich llama *big social data* (Burgess; Bruns, 2012)

y que pone de relieve los puntos en común que las plataformas de medios sociales tienen con el *big data* en tanto que éstas se presentan como nodo del intercambio de contenido y persiguen no sólo obtener un mayor flujo comunicativo sino también ser capaces de aprovechar el volumen de datos generados (Barrera, 2013).

Así, los atributos del *big data* (volumen, velocidad, variedad, valor) se presentan como un eje y foco de atención para el estudio, tratamiento y procesado de los datos de medios sociales (figura 2).

El valor del análisis del *social media data* se observa no sólo en la inversión que se está asignando a su investigación sino también en su utilidad para, entre otras cuestiones, comprender el pensamiento y actuación de las personas, mejorar los procesos que llevan a la toma de decisiones y dirigir de forma eficaz productos y servicios (Hobbs, 2014).

3. Características y singularidades del *social media data*

La información y los mensajes que se intercambian en los *social media* circulan en una dimensión semipública de la comunicación pues, aunque algunos perfiles y plataformas permiten al usuario regular la privacidad de los contenidos que comparte, estos quedan expuestos, cuando menos, al resto de participantes o miembros de su red de contactos (Oboler; Welsh; Cruz, 2012). Asimismo algunos *social media* facilitan interfaces de programación de aplicaciones (APIs) que permiten acceso gratuito a los datos procedentes de los contenidos públicos e incluso las compañías transfieren datos a terceros de cuentas privadas si los usuarios han dado su consentimiento o si éstos se amparan bajo la cobertura del anonimato (Hobbs, 2014). Ello ha abierto el debate sobre el respeto al derecho a la intimidad y explica la presencia de voces que reivindican mayor protección en este tipo de plataformas (Tene; Polonetsky, 2012, p. 65). La discusión se complica si entra en juego el derecho al olvido, “el derecho de las personas físicas a hacer que se borre la información sobre ellas después de un período de tiempo determinado” (De-Terwangne, 2012, p. 54), y las propias limitaciones que incorpora la normativa, como por ejemplo la incluida en la *Directiva europea 95/46/CE (Unión Europea, 1995)* que admite, tal y como se recoge en su artículo 6, “el tratamiento posterior de datos con fines históricos, estadísticos o científicos” aunque inicialmente no fueran recogidos con ese propósito.

Dado el valor de los datos que se pueden recopilar sobre los usuarios de *social media* y su interés para los propios usuarios, investigadores, empresas y gobiernos, “el dominio de las redes sociales en internet alberga detrás innumerables cuestiones relacionadas con el poder” (Segovia, 2013). Ello justifica no sólo los movimientos de expansión vertical de algunas compañías, que multiplican así su presencia en plataformas sociales accediendo a más parcelas de vida social de los usuarios, sino también que entre las APIs se encuentren restricciones en el volumen de datos transferidos. Esta evolución actualiza a su vez la reflexión sobre la neutralidad de la Red (Cullell-March, 2012) y la gobernanza de internet (Pérez; Olmos, 2009).

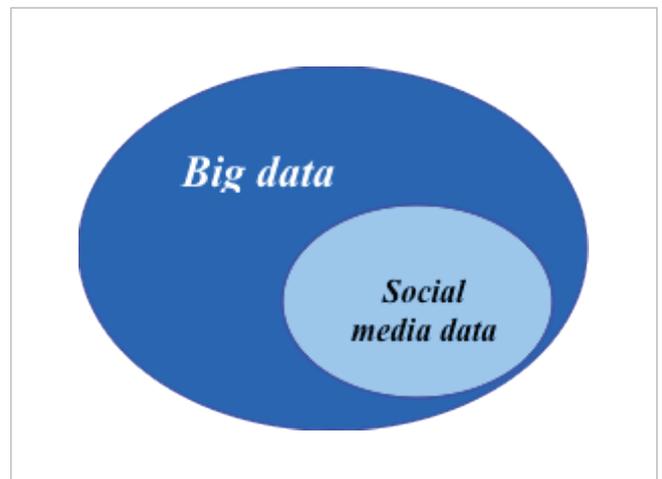


Figura 2. Representación de los datos procedentes de medios sociales dentro del fenómeno *big data*

La datificación en *social media* se relaciona entre otros con los mensajes y contenidos publicados y/o compartidos, los estados de ánimo o gustos expresados (a través de los botones “me gusta” o “favorito”), la descripción de los perfiles, los sentimientos, experiencias profesionales y aspectos aparentemente imperceptibles recogidos en forma de metadatos (Mayer-Schönberger; Cukier, 2013, p. 166). Pero su objeto también son las relaciones o las propias interacciones producidas. En la comunidad o red, no obstante, no todos los nodos o perfiles tienen el mismo valor, especialmente en un contexto en el que, siguiendo a Goldhaber (1997), se asiste a una economía de la atención. Esta se presenta como un recurso fundamental ante el volumen de intercambio de contenidos. La intermediación en un entorno de confianza se convierte en valor agregado y dota a los sujetos que la ejercen de mejor posición o crédito dentro de una comunidad o red (Velázquez; Rey-Marín, 2007; Boyd, 2009).

De este modo, a pesar del potencial que internet ofrece para el desarrollo de una comunicación horizontal, en los *social media* el concepto de “influencia” se redimensiona. En el entorno digital se actualizan, adquiriendo características propias, teorías clásicas sobre los efectos de la comunicación que señalaban la importancia de los líderes de opinión (Katz; Lazarsfeld, 1955) o la relevancia de los canales interpersonales para la difusión que el modelo en “J” de Greenberg¹ reconocía. Además, ahora nuevas personas distintas a las que ejercen influencia en la comunicación cara a cara, pueden alcanzar un estatus destacado en la Red (González-García, 2010).

Es un reto identificar a los denominados *influencers*, así como medir o evaluar el impacto que, por su posición estratégica, pueden ejercer en distintos sectores, incluyendo el del negocio (Roy, 2014). Reflexiones como la publicada por Adi Avnit (2009) con el título *The million followers fallacy* o estudios como los realizados en el *Max Planck Institute for Software Systems* en el marco del *Twitter Project* (Cha et al., 2010) desmienten algunos falsos mitos sobre el papel que ejercen de forma aislada en la influencia aspectos como la popularidad, el número de seguidores o fans, y ponen de relieve la necesidad de combinar varios criterios para medirla.



Figura 3. <http://kred.com>

4. Medición, monitorización y análisis del *social media data*

El análisis de los datos generados en los medios sociales se presenta como un nuevo campo de estudio en el que se hace necesaria la aplicación de nuevos métodos y tecnologías. Han surgido algunos productos que sin embargo presentan sesgos o bien ofrecen soluciones parciales, por lo que los analistas necesitan utilizar varios y combinarlos según el tipo y características de cada plataforma.

Existen programas de gestión de *social media* como *Tweetdeck*, que trabaja sobre *Twitter*, o *Hootsuite*, que opera con distintos medios sociales. También hay sistemas para monitorización y rastreo de temas, marcas, palabras clave o perfiles, pudiendo en algún caso recopilar información sobre los sentimientos que se asocian a determinados contenidos. Se encuentran aquí productos como *Radian6*, *Socialmention* y el sistema *Truthly* que, aunque en su origen se diseñó para identificar campañas de desinformación en *Twitter* (McKelvey et al., 2012), permite agrupar *tweets* en función de mensajes o contenidos relacionados. Asimismo hay sistemas de medición y analítica como las estadísticas de *Facebook Insights* o *Google Analytics*, que incluye una sección de medios sociales que se centra en las referencias, actividades e interacciones. También se debe citar *Socialbro*, que ofrece servicios de analítica para *Twitter* y que incorpora *Kred*, uno de los recursos que, como *Klout*, miden la influencia de perfiles en medios sociales a partir de algoritmos y la puntúan (figuras 3 y 4).

Por un lado, según Scarfi (2012), no existe un sistema que permita dar sentido a los datos de los medios sociales a través de herramientas analíticas; por otro, los sistemas de análisis están limitados a aquellos capaces de dar cabida a grandes cantidades de datos y que sacrifican, según el caso, la complejidad del cálculo o la actualización en tiempo real. Como consecuencia, existen muchos interrogantes sobre la gestión y generación de conocimiento a partir del *social media data* (Sandoval-Almazán; Gómez-Díaz, 2012) y se necesitan soluciones para mejorar las técnicas no supervisadas de detección e identificación de líderes de opinión.

En este sentido, para identificar nodos hiperconectados de valor, se propone² la utilización de metodologías de cálculo y tecnologías para la detección y el análisis de tendencias, junto al análisis de emociones y la apertura a recursos externos.

La detección de tendencias requiere potentes herramientas capaces de gestionar casi de forma automática datos generados en los *social media* y, aunque se estudian soluciones que permitan el análisis comparativo de tendencias en distintos períodos (Shih; Liu; Hsu, 2010; Kim; Suh; Park, 2008), desde la perspectiva de la investigación se trata de adaptar y aprovechar metodologías, algoritmos y funciones existentes para el análisis.

Además, se requiere el uso combinado de grafos para detectar relaciones y perfiles similares y contenidos afines. Esto permite el cálculo de relaciones de la red social a partir de métricas que incluyen entre otros el análisis de *authorities* y de *hubs*³ que establecen la importancia de los nodos y su relevancia respecto al resto de nodos.

“ A pesar del potencial que internet ofrece para la comunicación horizontal, en los *social media* el concepto de ‘influencia’ se redimensiona ”

A ello se deben sumar otras métricas para registrar determinados aspectos en la red como: *closeness* (cercanía), *betweenness* (intermediación), *degree centrality* (centralidad), *eigenvalue*, *page rank* y *community* (comunidad). Las técnicas de análisis de redes sociales (*SNA* o *ARS*) que permiten el estudio estructural (Sanz-Menéndez, 2003) o reticular, contribuyen a generar información sobre algunas de estas cuestiones si bien el componente estático de las descripciones resultantes representa una limitación.

Para gestionar el cálculo en tiempo real, el *big data* se presenta como la solución idónea al permitir trabajar con bases de datos caracterizadas por los bajos tiempos de latencia de respuesta, la posibilidad de ser escalable y asegurar la consistencia y recuperación ante fallos del sistema.

5. Conclusiones

Los *social media*, caracterizados por su rápida expansión (West, 2013) y por la variedad de interacciones y de contenidos que en ellos circulan, se han incorporado al estudio del *big data* al convertirse en fuente de datos útiles para investigadores, entidades y empresas. A pesar de compartir los atributos del fenómeno *big data*, que se sintetizan en las tres V's identificadas por Russom (2011) a las que se debe sumar el valor de la información, el *big social data* presenta una particular complejidad derivada de la propia naturaleza de las interacciones y del distinto valor que alcanzan los

nodos o del interés que despierta el peso que pueden ejercer los denominados *influencers*. Con todo existen muchos interrogantes sobre la gestión y generación de conocimiento a partir del *social media data* y las herramientas o soluciones de analítica y medición existentes son parciales y no están exentas de sesgos.

La oportunidad se presenta en la generación de aplicaciones que no sólo permitan recabar los datos generados en esos entornos sino también analizarlos de manera que los resultados puedan ofrecer nuevas soluciones para optimizar los recursos y convertirse en una herramienta más para la toma de decisiones.

La combinación de algoritmos que permitan el análisis de tendencias, métricas que faciliten entre otros el estudio de *hubs* y *authorities*, y la aplicación de bases de datos cuyas propiedades encajen en el paradigma *big data*, permitirán identificar nodos hiperconectados y diseñar estrategias para la difusión selectiva de contenidos, valorar modelos económicos de la Red, gestionar perfiles que consigan configurar amplias redes de usuarios y mejorar el retorno de la inversión en la comunicación en *social media*.

El análisis de los datos generados en los medios sociales se presenta como un nuevo campo de estudio en el que se hace necesaria la aplicación de nuevos métodos y tecnologías

Notas

1. El modelo en J de Greenberg explica las posibilidades de difusión del mensaje y es considerada una de las teorías clásicas de comunicación que estudian los efectos a largo plazo. Este modelo, cuya representación corresponde a una curva en forma de J (de ahí su nombre), valora ante acontecimientos con distinto grado de relevancia o de implicación, el volumen de personas que conocen el hecho y las fuentes por las que se informaron (más allá de los medios de comunicación).
2. Propuesta recogida en el proyecto *Social engagement. Solución linked big data para el establecimiento de modelos económicos en la Red*. Referencia RTC-2014-2178-7 de la Convocatoria Retos-colaboración 2014 del Ministerio de Economía y Competitividad de España.
3. El análisis de *authorities* y de *hubs* permite determinar el valor de los nodos en función de los enlaces de entrada que reciben o de salida que contienen respectivamente.

6. Bibliografía

Avnit, Adi (2009). "The million followers fallacy". *Internet draft, Pravda Media*.

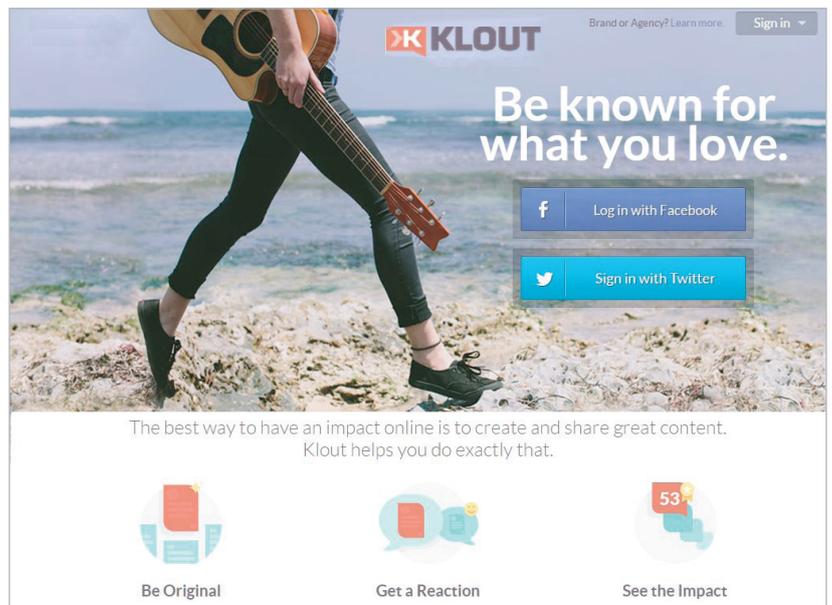


Figura 4. <https://klout.com>

Barrera, Alba (2013). "Así circula el contenido social". *Baquía.com*, 19 agosto.

<http://goo.gl/MCfQg4>

Bowden, Jason (2014). "Reasons to explore big data with social media analytics". *SocialMediaToday.com*, 15 enero.

<http://www.socialmediatoday.com/content/reasons-explore-big-data-social-media-analytics-videos>

Boyd, Danah (2009). "Streams of content, limited attention: the flow of information through social media". *Web2.0 Expo*. NY, November 17.

<http://www.danah.org/papers/talks/Web2Expo.html>

Burgess, Jean; Bruns, Axel (2012). "Twitter archives and the challenges of big social data for media and communication research". *M/C Journal*, v. 15, n. 5.

<http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/561>

Cha, Meeyoung; Haddadi, Hamed; Benevenuto, Fabricio; Gummadi, Krishna P. (2010). "Measuring user influence in Twitter: The million follower fallacy". *Intl AAAI Conf on weblogs and social media (Icwsm)*.

<http://twitter.mpi-sws.org>

Cullell-March, Cristina (2012). "El futuro de la Web ante la neutralidad de la Red: estado de la cuestión en la Unión Europea". *El profesional de la información*, v. 21, n.1, enero-febrero, pp.77-82.

<http://repositori.uji.es/xmlui/handle/10234/70240?locale-attribute=en>

<http://dx.doi.org/10.3145/epi.2012.ene.10>

De-la-Rosa-Troyano, Francisco-Fernando; Martínez-Gasca, Rafael (2007). "Sistemas de inteligencia web basados en redes sociales". *Redes. Revista hispana para el análisis de redes sociales*, v. 12, n. 9.

<http://revistes.uab.cat/redes/article/view/105>

De-Terwangne, Cécile (2012). "Privacidad en internet y el derecho a ser olvidado/derecho al olvido". *Revista*

d'internet, dret i política, n. 13, febrero. Monográfico VII Congreso intl internet, derecho y política. Neutralidad de la red y otros retos para el futuro de internet.

<http://journals.uoc.edu/index.php/idp/issue/view/n13>

Facebook (2014a). "Company info". *Newsroom.fb.com*.

<http://newsroom.fb.com/company-info>

Facebook (2014b). "World cup 2014: Facebook tops a billion interactions". *Newsroom.fb.com*, 30 junio.

<http://newsroom.fb.com/news/2014/06/world-cup-2014-facebook-tops-a-billion-interactions>

Foldes, Abi (2013). "Tendances clés du marché des réseaux sociaux en France et dans le monde en 2012". *ComScore*, 18 octubre.

<http://goo.gl/B02Ufc>

García-Marco, Francisco-Javier (2011). "La pirámide de la información revisitada: enriqueciendo el modelo desde la ciencia cognitiva". *El profesional de la información*, v. 20, n. 1, enero-febrero, pp. 11-24.

<http://pciucr.files.wordpress.com/2011/03/61263234.pdf>

<http://dx.doi.org/10.3145/epi.2011.ene.02>

Goldhaber, Michael (1997). "Attention economy". *Wired magazine*.

González-García, María (2010). "Los líderes de opinión o *influentials* en las redes sociales de internet. Hacia el *three-step flow* que defina los procesos de influencia personal en la Red". *Comunicación y desarrollo en la era digital. Congreso internacional AE-IC*, febrero, Málaga.

<http://www.ae-ic.org/malaga2010/upload/ok/41.pdf>

Hobbs, Abbi (2014). "Social media and big data". *POSTnote*, n. 460, March. Houses of Parliament. Parliamentary Office of Science & Technology.

<http://www.parliament.uk/briefing-papers/POST-PN-460/social-media-and-big-data>

Katz, Elihu; Lazarsfeld, Paul F. (1955). *Personal influence. The part played by people in the flow of mass communications*. NY: The Free Press. ISBN: 987 1412805070

Kim, Young-Gil; Suh, Jong-Hwan; Park, Sang-Chan (2008). "Visualization of patent analysis for emerging technology". *Expert systems with applications*, v. 34, n. 3, pp. 1804-1812.

<http://dx.doi.org/10.1016/j.eswa.2007.01.033>

LinkedIn (2014). "LinkedIn reaches 300 million members worldwide". *Press.linkedin.com*, 18 abril.

<http://press.linkedin.com/News-Releases/333/LinkedIn-reaches-300-million-members-worldwide>

Magnusson, Jonathan (2012). *Social network analysis utilizing big data technology*. Uptec. Uppsala Universitet.

<http://www.diva-portal.org/smash/get/diva2:509757/FULLTEXT01.pdf>

Manovich, Lev (2012). "Trending: the promises and the challenges of big social data". En: Gold, Matthew (Ed.). *Debates in the digital humanities*, Minneapolis: U. Minnesota Press, pp. 460-475. ISBN: 978 0 8166 7795 5

Mayer-Schönberger, Viktor; Cukier, Kenneth (2013). *Big data. La revolución de los datos masivos*. Madrid: Turner

Publications. ISBN: 978 8415832102

McKelvey, Karissa; Rudnick, Alex; Conover, Michael D.; Menczer, Filippo (2012). "Visualizing communication on social media. Making big data accesible". *Procs. CSCW Workshop on collective intelligence as community discourse and action*.

Oboler, Andre; Welsh, Kristopeher; Cruz, Lito (2012). "The danger of big data: social media as computational social science". *First Monday*, v. 17, n. 7.

<http://journals.uic.edu/ojs/index.php/fm/article/view/3993/3269>

<http://dx.doi.org/10.5210/fm.v17i7.3993>

Pérez, Jorge; Olmos, Ana (2009). "Introducción, gobernanza de internet". *Telos. Cuadernos de comunicación e innovación*, n. 80, junio-septiembre.

<http://telos.fundaciontelefonica.com>

Pinterest (s. f.). *Pinterest.com*

<https://es.pinterest.com>

Ponjuán, Gloria (1998). *Gestión de la información en las organizaciones: principios, conceptos y aplicaciones*. Universidad Chile. ISBN: 9567782008

Roy, Arnaud (2014). *Informe Augure sobre el estatus del marketing de influencers*. Augure.

<http://www.augure.com/es/blog/influencer-marketing-estatus-20140220>

Russom, Phillip (2011). *Big data analytics*. TDWI Best practices report.

<http://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx?tc=page0>

Sandoval-Almazán, Rodrigo; Gómez-Díaz, María-Rocío (2012). "Un análisis exploratorio del uso de las redes sociales en internet como herramienta para la gestión del conocimiento". *Revista eletrônica de sistemas de informação*, v. 11, n. 1.

<http://goo.gl/KsdR7M>

Sanz-Menéndez, Luis (2003). "Análisis de redes sociales: o como representar las estructuras sociales subyacentes". *Apuntes de ciencia y tecnología*, junio, n. 7, pp. 21-29.

<http://www.aacte.eu/wp/apuntes>

Scarfí, Marita (2012). "Social media and the *big data* explosion". *Forbes.com*, 28 June.

<http://www.forbes.com/sites/onmarketing/2012/06/28/social-media-and-the-big-data-explosion>

Segovia, Adrián (2013). "Las redes sociales se agotan en el PC". *Estrategia digital*, *Elpais.com*, 5 diciembre.

<http://blogs.elpais.com/estrategia-digital/2013/12/las-redes-sociales-se-agotan-en-los-ordenadores.html>

Shih, Meng-Jung; Liu, Duen-Ren; Hsu, Ming-Li (2010). "Discovering competitive intelligence by mining changes in patent trends". *Expert systems with applications*, v. 37, n. 4, pp. 2882-2890.

<http://dx.doi.org/10.1016/j.eswa.2009.09.001>

Smith, Matthew; Szongott, Christian; Henne, Benjamin; Von-Voigt, Gabriele (2013). "Big data privacy issues in pu-

blic social media". *Digital ecosystems technologies*, 2012. 6th IEEE Intl Conf.
<http://dx.doi.org/10.1109/DEST.2012.6227909>

Tene, Omer; Polonetsky, Jules (2013). "Big data for all: privacy and user control in the age of analytics". *Northwestern journal of technology and intellectual property*, v. 11, n. 5.
<http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1>

Twitter (s. f.). "Uso de Twitter". *About.twitter.com*
<https://about.twitter.com/es/company>

Unión Europea (1995). "Directiva 95/46/CE del Parlamento Europeo y del Consejo, 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos". *Diario oficial*, n. L 282 de 23/11/1995, pp. 0031-0050.
http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_es.pdf

Velázquez, Alejandro; Rey-Marín, Lluís (2007). "El valor agregado de las redes sociales: propuesta metodológica

para el análisis del capital social". *Redes*, v. 13, n. 5.
<http://revistes.uab.cat/redes/article/view/111>

Villars, Richard; Olofson, Carl W.; Eastwood, Matthew (2011). *White paper. Big data: what it is and why you should care*. IDC.
http://sites.amd.com/sa/Documents/IDC_AMD_Big_Data_Whitepaper.pdf

Watters, Audrey (2011). "Visualization of the week: Twitter's global pulse". *O'Reilly Radar*, 8 July.
<http://radar.oreilly.com/2011/07/visualization-twitter-pulse-japan-earthquake.html>

West, Owen (2013). "Big data insights". *Webit Congress 2013*. ComScore.
<https://www.comscore.com/es/Insights/Presentations-and-Whitepapers/2013/Big-Data-Insights>

YouTube (s. f.). "Estadísticas". *Youtube.com*
<https://www.youtube.com/yt/press/es/statistics.html>

Zaslavsky, Arkady; Perera, Charith; Georgakopoulos, Dimitrios (2012). "Sensing as a service and big data". En: *Procs of the Intl conf on advances in cloud computing*, India, July. ISBN: 978 8173717789

Anuario ThinkEPI 2014



384 páginas de análisis de tendencias en información, documentación y comunicación

Formulario de compra:

<http://www.elprofesionaldelainformacion.com/suscripciones.php>

Información y pedidos:

Isabel Olea

epi.iolea@gmail.com

+34 608 491 521

Ahora disponible en:

<http://recyt.fecyt.es/index.php/ThinkEPI>

MELODA, MÉTRICA PARA EVALUAR LA REUTILIZACIÓN DE DATOS ABIERTOS

Alberto Abella, Marta Ortiz-de-Urbina-Criado y Carmen De-Pablos-Heredero



Alberto Abella es ingeniero superior de telecomunicaciones, profesor asociado en la *Universidad Rey Juan Carlos* y estudiante de la *Escuela Internacional de Doctorado* en el Programa de Ciencias Sociales y Jurídicas en la *Universidad Rey Juan Carlos*. Es experto en los campos de *open data* y *open government* aplicados al sector público y privado. Es presidente de la *Open Knowledge Foundation* en España.

<http://orcid.org/0000-0003-0243-4606>

alberto.abella@meloda.org



Marta Ortiz-de-Urbina-Criado es doctora en dirección y organización de empresas y profesora titular en la *Universidad Rey Juan Carlos*. Ha publicado artículos en revistas de impacto como *International business review*, *International journal of technology management*, *Journal of knowledge management* y *Revista española de financiación y contabilidad*, entre otras.

<http://orcid.org/0000-0001-7527-6798>

marta.ortizdeurbina@urjc.es



Carmen De-Pablos-Heredero es doctora en ciencias económicas y empresariales, profesora titular de universidad y directora del máster universitario en organización de empresas en la *Universidad Rey Juan Carlos*. Ha publicado artículos en revistas de impacto como *CEDE*, *Universia*, *Interciencia*, *TIBE*, *Journal of entrepreneurship management*, *International journal of marketing research*, *Revista de ciencias sociales*, *Revista de economía mundial*, *Dyna*, *Pensée*, *Medical economics*, *REIS*, etc.

<http://orcid.org/0000-0003-0457-3730>

carmen.depablos@urjc.es

Universidad Rey Juan Carlos
Facultad de Ciencias Jurídicas y Sociales
Paseo de los Artilleros, s/n. 28032 Madrid, España

Resumen

Las políticas actuales para el aumento de la transparencia, la implantación del gobierno abierto o de las ciudades inteligentes tienen en la publicación de información uno de sus pilares fundamentales. Sin embargo, la disponibilidad de mecanismos de evaluación del uso y valor de esta información es limitada. Se presenta una nueva métrica, *Meloda*, que permite calificar la información y evaluar su grado de reutilización. Se describen las cuatro dimensiones de la métrica: estándares técnicos, acceso, legal, y modelo de datos, y se explica su proceso de evaluación.

Palabras clave

Datos abiertos, Reutilización de la información, Métrica, Gobierno abierto, *Meloda*.

Title: *Meloda*, a metric to assess open data reuse

Abstract

Current policies demanding an increase in transparency, open government and smart cities share open data publication as one of their basic pillars. However, there is a limited availability of mechanisms to assess the use and value of this information. The goal of this paper is to present a new metric, *Meloda*, that qualifies the information and helps to assess its reusability. *Meloda's* four dimensions are described: technical standards, access, legal considerations, and data model. Finally *Meloda's* assessment process is explained.

Keywords

Open data, Information reuse, Metric, Open government, *Meloda*.

Artículo recibido el 22-07-2014

Aceptación definitiva: 08-10-2014

Abella, Alberto; Ortiz-de-Urbina-Criado, Marta; De-Pablos-Heredero, Carmen (2014). "Meloda, métrica para evaluar la reutilización de datos abiertos". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 582-588.

<http://dx.doi.org/10.3145/epi.2014.nov.04>

1. Introducción

La alianza para el gobierno abierto, nacida en 2011 y que incluye a 60 países (España desde abril de 2012), es una prueba del interés que el *open government* o gobierno abierto está suscitando en todo el mundo (Ramírez-Alujas; Dassen, 2014). Se trata de "una forma de comunicación abierta, permanente y bidireccional entre la administración y sus administrados, ciudadanos, empresas y sociedad civil" (Hofmann; Ramírez-Alujas; Bojórquez-Pereznieto, 2012). Los tres pilares que lo definen son participación, colaboración y rendición de cuentas (publicación de información).

Las instituciones europeas están dando impulso a las políticas de reutilización de información como dinamizadoras de la economía digital y como base de la transparencia democrática.

Dentro del movimiento *open data*, que promueve la liberación de datos (Peset; Ferrer-Sapena; Subirats-Coll, 2011), se aboga porque las administraciones públicas pongan a disposición de la sociedad la información pública que tienen en su poder, a lo que se denomina *open government data*, para que cualquier persona u organización pueda, a partir de la misma, crear nuevas informaciones y servicios (Marcos-Martín; Soriano-Maldonado, 2011; Weiskopf; Weng, 2013). La apertura de la información del sector público permite establecer mecanismos de transparencia y favorece la colaboración ciudadana; aunque algunos autores apuntan a que la política de puesta a disposición de los datos no implica necesariamente que haya transparencia, ya que para lograrla es necesario que los datos estén tratados y presentados de tal forma que puedan reutilizarse para crear estudios, servicios de valor añadido o generar conocimiento para la sociedad (Porter; Millar, 1985; Shapiro; Varian, 1998; Ferrer-Sapena; Peset; Alexandre-Benavent, 2011).

Reutilizando esta información se pueden ofrecer nuevos productos y servicios digitales, lo cual dinamiza la actividad económica y empresarial. Pueden crearse nuevas empresas que con pocos recursos propios lleven a cabo modelos de negocio basados en la reelaboración de dicha información, tratándola y añadiéndole valor (Marcos-Martín; Soriano-Maldonado, 2011).

España está entrando en una dinámica activa en la liberación de datos públicos con la apertura de numerosos portales. El localizador de datos abiertos a nivel mundial de la web de la Fundación CTIC ofrece un mapa de la situación en el que se incluyen 267 portales con conjuntos de datos públicos, de los cuales 32 son españoles. Sin embargo, hemos realizado una revisión de los portales locales, regionales y nacionales y hemos observado que esta cifra es superior a 50.
<http://datos.fundacionctic.org/sandbox/catalog/faceted>

Hoy en día las ciudades inteligentes¹ ofrecen una gran cantidad de información pública a través de sus portales. Esta información está recogida en catálogos que facilitan el acceso a las fuentes de información y que pueden ser utilizados

para la creación de nuevos servicios por parte del sector privado, siendo, además, fuentes valiosas para los negocios de *big data* (LaValle et al., 2011; Kitchin, 2014).

La actualización de la *Directiva 2003/98/CE relativa a la reutilización de la información del sector público en la Unión Europea*, de junio de 2013 (Comisión Europea, 2013), confirma la importancia otorgada a este tema desde las más altas instancias políticas europeas. La puesta a disposición de información pública en España está regulada principalmente por la *Ley 37/2007 de 16 de noviembre sobre reutilización de información* (España, 2007), que tiene como objetivo "armonizar la explotación comercial de la información del sector público y la publicación de todos los documentos de libre disposición que obran en su poder, al considerar que son un instrumento esencial para el desarrollo del derecho al conocimiento, que constituye un principio básico de la democracia". Esta ley se complementa con el *Esquema nacional de interoperabilidad (RD 4/2010, de 8 de enero)* (España, 2010), así como por la reciente *Ley de transparencia, acceso a la información y buen gobierno (Ley 19/2013, de 9 de diciembre)* (España, 2013). Si bien otras normas, tanto nacionales como autonómicas y locales, condicionan aspectos o ámbitos de esta publicación.

Las políticas para el aumento de la transparencia, la implantación del gobierno abierto o de las ciudades inteligentes tienen en el *open data* uno de sus pilares fundamentales

La liberación de los datos supone un cambio organizativo, estructural y en la forma de trabajar de las administraciones, que tienen que establecer unas políticas de uso y privacidad sobre los que vayan a liberar, a cambio de ahorrar ciertos costes en la elaboración de los informes por parte de terceros (Ferrer-Sapena; Peset; Alexandre-Benavent, 2011; LaValle et al., 2011). Sin embargo en muchos casos el sector público no tiene establecidas, o no es el objeto de su actividad, las rutinas organizativas que le permitan desarrollar la capacidad para extraer todo el valor del *open data*. Y es aquí donde las empresas pueden intervenir creando servicios de valor añadido que hagan la información más comprensible y asequible (Ferrer-Sapena; Peset; Alexandre-Benavent, 2011; Lee; Kwak 2012).

Son todavía muy pocos los gobiernos o entidades públicas que puedan asegurar, o al menos estimar cuantitativamente los beneficios de sus políticas de apertura de datos públicos. En la mayoría de los casos sus estrategias de reutilización de datos se basan más en principios como el apoyo a las libertades ciudadanas, el cumplimiento de legislaciones propias, o declaraciones públicas de transparencia (Huijboom; Van-den-Broek, 2011). En España los informes elaborados

por la administración pública a través de la entidad pública empresarial *Red.es* (*Red.es*, 2011; 2012) analizan el impacto del sector económico que reutiliza la información aunque no establecen una relación causal con sus políticas de publicación de datos.

Casi todas las iniciativas analizadas presentan una carencia de métricas tanto cuantitativas como cualitativas que puedan estimar adecuadamente su impacto. **Fournier-Tombs** (2011) explica que es técnicamente complicado el análisis del impacto general de los sitios de publicación de datos. Hace una revisión de los trabajos que han estudiado la influencia de los portales sobre la comunidad de usuarios y observa que los métodos basados en la medición del uso (encuestas sobre los reutilizadores y usuarios) son de utilidad limitada porque no se pueden obtener muestras estadísticamente significativas.

Por tanto se presenta como una necesidad estratégica disponer de métricas que evalúen la reutilización. Estas métricas pueden facilitar el uso eficiente de recursos de las administraciones públicas y servir a las empresas para conocer qué informaciones pueden tener un mayor potencial económico para la creación de negocios innovadores. El objetivo de este trabajo es explicar el desarrollo y utilidad de una nueva métrica que permite evaluar el grado de reutilización de esta información.

2. Reutilización de datos

La definición de reutilización según la RAE es “utilizar algo, bien con la función que desempeñaba anteriormente o con otros fines”. La importancia de la reutilización de la información ya ha sido estudiada en otros muchos campos, desde el ámbito financiero (**Chan; Greenbaum; Thakor**, 1986), la medicina clínica (**Kim**, 2005) o la información científica (**Faniel; Jacobsen**, 2010).

La *Ley 37/2007* (España, 2007) define la reutilización de la información del sector público como “el uso de documentos que obran en poder de las administraciones y organismos del sector público, por personas físicas o jurídicas, con fines comerciales o no comerciales, siempre que dicho uso no constituya una actividad administrativa pública”. Se establecen 4 condiciones generales de reutilización:

- el contenido de la información no debe ser alterado;
- no se debe desnaturalizar el sentido de la información;
- se debe citar la fuente;
- se debe mencionar la fecha de la última actualización.

En este trabajo definiremos información reutilizable como aquella que cumple 4 condiciones mínimas que facilitan su uso:

- 1) no existencia de barreras técnicas a la reutilización;
- 2) posibilidad de acceso automatizado a la información (**Eaves**, 2010);
- 3) existencia de un esquema legal que permita su uso (**Eaves**, 2010);
- 4) acceso al conocimiento de la estructura de la información publicada.

Algunos autores han propuesto clasificaciones y formas de evaluar la calidad de la información pero no su reutilización.

Tabla 1. Principios del *open government data*
<http://opengovdata.org>

<i>Complete</i>	Acceso a toda la información
<i>Primary</i>	No se accede a los datos de forma agregada
<i>Timely</i>	Datos disponibles lo antes posible
<i>Accessible</i>	Datos accesibles para todos los propósitos y tipos de usuarios
<i>Machine processable</i>	Procesamiento de datos de forma automática
<i>Non-discriminatory</i>	Datos disponibles para todos los usuarios sin necesidad de registro
<i>Non-proprietary</i>	Datos almacenados en un estándar abierto
<i>License-free</i>	Datos no sujetos a restricciones en su uso

Por ejemplo, la clasificación en cinco estrellas de **Berners-Lee**² está orientada a evaluar si los datos están preparados para ser conectados (*linked data*) con otros conjuntos de datos. La métrica de **Pipino, Lee y Wang** (2002), que cuenta con 16 dimensiones de análisis, evalúa la calidad de la información, pero no sus posibilidades de reutilización. **Ren y Glissmann** (2012) analizan la arquitectura de datos de la organización para generar la publicación de datos abiertos con la mayor calidad de la información posible. Para ello presentan un proceso con 5 etapas para identificar la información relevante, comprobar su estado y diseñar los mecanismos de publicación. Sin embargo en su propuesta no se incluye la evaluación de la reutilización de esa información.

“ La métrica *Meloda* permite calificar la información y evaluar su grado de reutilización ”

3. Meloda

Nació en 2011 como reacción a la falta de homogeneidad en los conjuntos de datos que se publicaban en los portales *open data*, para analizar el grado de reutilización de la información. Para la elaboración de esta métrica se han considerado las tres leyes del *open government data* (**Eaves**, 2010):

- ser accesible por internet;
- ser legible por máquinas;
- que el marco legal permita usos distintos de los originales, así como los principios del *open government data* (tabla 1) (**Lessig**, 2007).

En las primeras versiones estables se consideraron tres dimensiones:

- Estándares técnicos. Trata de reflejar la importancia que tiene que la información sea almacenada en un estándar que no sea privativo (*non-proprietary*).
- Acceso a toda la información legible de forma automatizada, con el mayor detalle posible y para el mayor número de usos (*complete, machine procesable, primary y accessible*).

Tabla 2. *Meloda*: dimensiones y niveles

Estándares técnicos	Acceso	Legal	Modelo de datos
1. Estándar privativo Ej.: .xls, .shp, .doc	1. Sin acceso Ej.: mail no automático o acceso en persona	1. <i>Copyright</i> Ej.: <i>copyright</i>	1. Sin modelo publicado Ej.: tabla de datos sin descripción de los campos
2. Estándar abierto Ej.: .csv, .ods, wms	2. Acceso vía web con registro Ej.: formulario manual	2. Uso privado Ej.: <i>copyright</i> permitiendo uso personal	2. Modelo con campos de datos Ej.: tabla de datos con descripción de los campos
3. Estándar abierto con metadatos Ej.: rdf, rss, json	3. Acceso directo vía web Ej.: url único	3. Uso no comercial Ej.: CC BY-NC 4.0	3. Modelo con especificaciones de campos Ej.: vocabularios disponibles
	4. Acceso vía web con parámetros Ej.: url con parámetros	4. Uso comercial Ej.: CC BY-SA 4.0	4. Modelo externo normalizado Ej.: vocabularios disponibles aceptados por organización de normalización
	5. Acceso completo (API) Ej.: punto de acceso <i>Sparql</i>	5. Uso no limitado con autoría Ej.: CC BY 4.0	5. Modelo externo y generalizado Ej.: vocabularios disponibles aceptados por organización de normalización y reconocidos

- Legal. Importancia del marco legal en el uso de la información, y sin barreras legales para su uso, independientemente del respeto a la privacidad y a la seguridad (*license-free*).

Las primeras aplicaciones de la métrica se realizaron sobre 200 conjuntos de datos provenientes de portales de *open data* regionales, locales y nacionales de España. Se observó que era necesario incluir una cuarta dimensión que tuviera en cuenta el modelo de datos a publicar, que reflejara la importancia que tiene la estructura de datos para poder procesar la información (*machine processable*). La última versión, *Meloda 3.10*, incluye las cuatro dimensiones. <http://bit.ly/meloda301>

3.1. Dimensiones de *Meloda*

Las cuatro dimensiones que analiza son:

1) Estándares técnicos o estructura técnica en que se ofrece el dato. Califica las informaciones en tres niveles (tabla 2):

- 1: la información se ofrece en estándares privativos que incluyen restricciones de propiedad intelectual para su uso;
- 2: se alcanza cuando los datos están almacenados en formatos que no tienen restricciones para su utilización;
- 3: se alcanza cuando además del segundo nivel se incluye para cada dato un esquema de metadatos.

2) Acceso a la información o mecanismo por el cual se hace posible la descarga o conexión con la información. Presenta cinco niveles (tabla 2):

- 1: conjunto de datos que, o bien no tienen acceso web, o bien requieren una petición manual;
- 2: acceso mediante interacción manual del usuario o necesidad de registro;
- 3: url que sólo permite descargar el conjunto completo de datos;
- 4: al url se pueden añadir parámetros con los que no es necesario descargar o acceder a todo el conjunto de datos;
- 5: incluye aquellos conjuntos de datos que ofrecen la posibilidad de realizar consultas sobre los datos y cruzarlos con fuentes externas.

3) Marco legal: licencia que se asigna al conjunto de datos. Tiene cinco niveles (tabla 2):

- 1: el uso de la información se restringe completamente sin permiso explícito del autor;
- 2: se permite el uso de la información con fines particulares;
- 3: se permite el uso con fines no lucrativos;
- 4: se permite el uso comercial;
- 5: se permite un uso no limitado y la creación de derivados, siempre y cuando se reconozca la autoría.

La última versión, *Meloda 3.10*, incluye cuatro dimensiones: estándares técnicos, acceso a la información, marco legal y modelos de datos

4) Modelos de datos: modelo utilizado para publicar la información y la cantidad de veces que es utilizado en otras fuentes de datos. Tiene cinco niveles (tabla 2).

- 1: no se tiene un modelo publicado;
- 2: se identifican los campos de los datos;
- 3: se detallan y se publican las especificaciones de los campos que componen el modelo, aunque sea propio;
- 4: se usa un modelo externo normalizado, aunque esté poco extendido;
- 5: se usa un modelo generalizado.

La tabla 3 muestra la relación entre los niveles de las dimensiones de *Meloda* y la definición de *open data* (niveles en verde). Se puede observar que, aunque *Meloda* está centra-

Tabla 3. Relación entre dimensiones *Meloda* y *open data*

Legal	Acceso	Estándares técnicos	Modelo datos
5	5	3	5
4	4		4
3	3	2	3
2	2		2
1	1	1	1

da en información *open data*, incluye también informaciones que no cumplen con la definición y la métrica puede ser de utilidad también en otros ámbitos.

3.2. Proceso de evaluación de *Meloda*

A continuación se describe el proceso de evaluación (figura 1). La tabla 4 describe los pesos que se asignan a cada uno de los niveles.

La puntuación final se calcula como la raíz cuarta del producto de las puntuaciones obtenidas en cada dimensión, ya que esta fórmula permite obtener una distribución más uniforme de los resultados. Además se puede analizar el efecto que tiene un cambio en la puntuación de una o más dimensiones sobre el grado de reutilización³.

La tabla 5 presenta los rangos de reutilización, entre 0 y 100, que se proponen tras realizar una primera evaluación de los resultados con datos reales.

$$Meloda = 100 \times \sqrt[4]{Estándar-técnico \times Acceso \times Legal \times Modelo-datos}$$

A continuación se incluye un ejemplo de una evaluación realizada para un conjunto de datos con *Meloda*. La institución A publica un conjunto de datos como *linked data* pero con un modelo de datos publicado propio, con licencia de uso comercial, y con un mecanismo por el cual se puede utilizar un recurso *Sparql* para interrogar al conjunto de datos. Los niveles que tiene en cada dimensión son los siguientes:

Estándares técnicos	Nivel 3 (RDF con metadatos)	100%
Acceso	Nivel 5 (Interrogable vía <i>Sparql</i>)	100%
Legal	Nivel 4 (comercial)	90%
Modelo de datos	Nivel 2 (modelo de datos <i>ad hoc</i>)	30%

Tabla 4. Ponderación de la métrica *Meloda* (%)

#	Legal	Acceso	#	Estándares técnicos	#	Modelo datos
5	100	100	3	100	5	100
4	90	90			4	90
3	25	25	2	60	3	50
2	10	10			2	10
1	0	0	1	20	1	0

Tabla 5. Rangos de reutilización

Rangos	Calificación
75-100	Reutilización avanzada
50-75	Reutilización avanzada con alguna característica mejorable
25-50	Reutilización básica
0-25	Inadecuado para reutilización

$$Meloda = 100 \times \sqrt[4]{1 \times 1 \times 0,9 \times 0,3} = 72,08$$

Se observa que el conjunto de datos está en el rango de reutilización apta pero con alguna característica mejorable (tabla 5).

4. Conclusiones

En este trabajo se ha descrito una nueva métrica que caracteriza la reutilización a través de cuatro dimensiones y propone una forma cuantitativa para evaluarla. La última versión (*Meloda* 3.10) se encuentra en proceso de mejora y actualización. Para el desarrollo de nuevas versiones se está considerando la inclusión de la dimensión temporal y geográfica y el ajuste de las ponderaciones de los distintos niveles.

Esta métrica tiene interés para todas las organizaciones que publican datos, públicas y privadas, y para todas las entidades, con o sin ánimo de lucro, que elaboran productos y servicios basados en ellas. Un tema interesante para futuros trabajos es el análisis del uso de la información que hacen las organizaciones (públicas, privadas, con ánimo de lucro o no), para identificar cuáles son las que crean más valor gracias a la reutilización.

Los estudios previos sobre

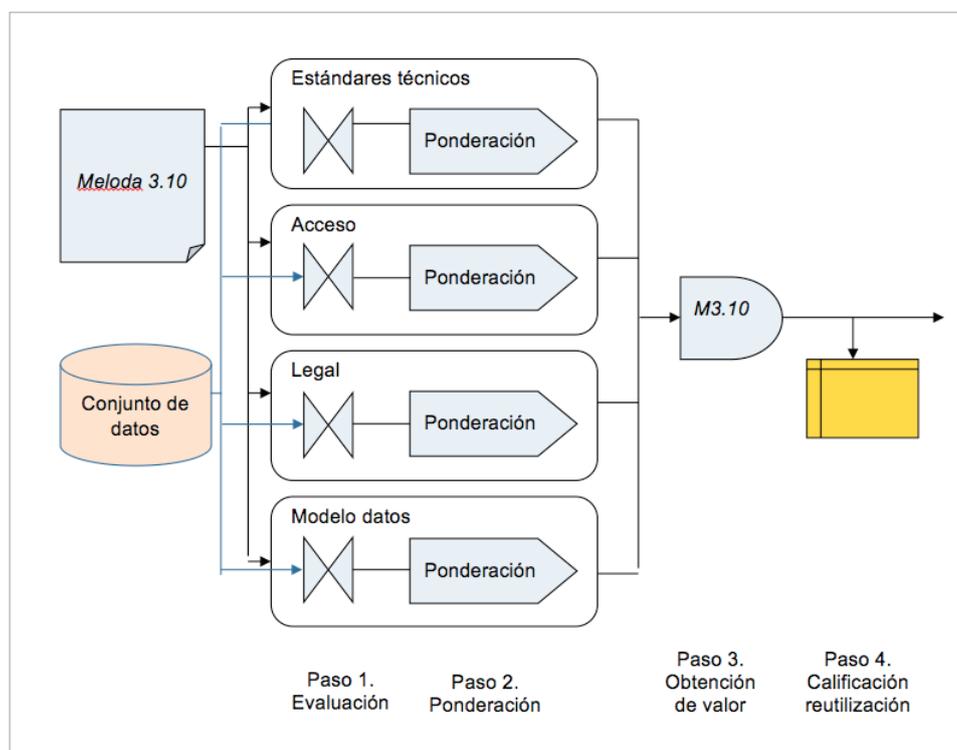


Figura 1. Proceso de evaluación de *Meloda*

gobierno abierto suelen analizar las políticas gubernamentales y la creación de valor social (transparencia y colaboración ciudadana). Por otra parte, la literatura sobre reutilización de la información centra su atención en el uso de la información y en el análisis del valor económico que se puede crear mediante el uso eficiente de los recursos públicos –eficiencia- y/o de la creación de nuevos productos y servicios –innovación- (Jetzek; Avital; Bjørn-Andersen, 2013; 2014). Sin embargo, hay pocos trabajos que demuestren empíricamente el impacto económico de la reutilización de la información. Por ello, la investigación futura analizará, utilizando esta métrica, el valor económico que tiene la reutilización de la información.

Notas

1. Las ciudades inteligentes son un ecosistema público-privado que con uso intensivo de la tecnología proporciona servicios a los ciudadanos y a sus organizaciones.
2. Las cinco estrellas son: licencia abierta, datos estructurados, estándares abiertos, disponer de un identificador de recursos uniforme (URI) y fuentes conectadas.
<http://5stardata.info>
3. En las primeras versiones de la métrica se comprobó que la función suma de las dimensiones daba resultados incoherentes, y que la distribución uniforme de la puntuación de los niveles no permitía identificar bien los rangos de reutilización.

Agradecimientos

Este trabajo ha sido financiado por el proyecto ECO2012-36775 del *Ministerio de Economía y Competitividad* (España) y ha recibido el apoyo del proyecto *Apps4EU* financiado por *The Information and Communication Technologies Policy Support Programme (ICI PSP)*.

Bibliografía

- Chan, Yuk-Shee; Greenbaum, Stuart I.; Thakor, Anjan V. (1986). "Information reusability competition and bank asset quality". *Journal of banking and finance*, v. 10, n. 2, junio, pp. 243-253.
<http://goo.gl/JX7OHi>
[http://dx.doi.org/10.1016/0378-4266\(86\)90008-7](http://dx.doi.org/10.1016/0378-4266(86)90008-7)
- Comisión Europea (2013). "Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público". *Diario oficial de la Unión Europea*, 27 de junio.
https://www.boe.es/diario_boe/txt.php?id=DOUE-L-2013-81251
- Eaves, David (2010). "The three laws of open government data". *Conference for Parliamentarians: balancing openness and the public interest in protecting information*.
<http://bit.ly/eaves2010>
- España (2007). "Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público". *Boletín oficial del estado*, n. 276, 17 de noviembre, pp. 47160-47165.
<http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>

España (2010). "Real decreto 4/2010, de 8 de enero, por el que se regula el Esquema nacional de interoperabilidad en el ámbito de la administración electrónica". *Boletín oficial del estado*, n. 25, 29 de enero, pp. 8139-8156.
<http://www.boe.es/boe/dias/2010/01/29/pdfs/BOE-A-2010-1331.pdf>

España (2013). "Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información y buen gobierno". *Boletín oficial del estado*, n. 295, 10 de diciembre, pp. 97922-97952.
<http://www.boe.es/boe/dias/2013/12/10/pdfs/BOE-A-2013-12887.pdf>

Faniel, Ixchel M.; Jacobsen, Trond E. (2010). "Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data". *Computer supported cooperative work*, v. 19, n. 3-4, August, pp. 355-375.
<http://dx.doi.org/10.1007/s10606-010-9117-8>

Fournier-Tombs, Eleonore (2011). "Evaluating the impact of open data websites". *Working papers series-SSRN*.
<http://dx.doi.org/10.2139/ssrn.1926201>

Ferrer-Sapena, Antonia; Peset, Fernanda; Alexandre-Benavent, Rafael (2011). "Acceso a los datos públicos y su reutilización: open data y open government". *El profesional de la información*, v. 20, n. 3, mayo-junio, pp. 260-269.
<http://eprints.rclis.org/20997>
<http://dx.doi.org/10.3145/epi.2011.may.03>

Hofmann, Andrés; Ramírez-Alujas, Álvaro; Bojórquez-Pereznieto, José-Antonio (coords.) (2012). *La promesa del gobierno abierto*, Itiap.
<http://www.lapromesadelgobiernoabierto.info>

Huijboom, Noor; Van-den-Broek, Tijs (2011). "Open data: an international comparison of strategies". *European journal of e-practice*, n. 12, March-April, pp. 1-13.
<http://www.epractice.eu/en/document/5290090>

Jetzek, Thorhildur; Avital, Michel; Bjørn-Andersen, Niels (2013). "Generating value from open government data". En: *34th Intl conf on information systems*, Milan, pp. 1-20.
<http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1181&context=icis2013>

Jetzek, Thorhildur; Avital, Michel; Bjørn-Andersen, Niels (2014). "Data-driven innovation through open government data". *Journal of theoretical and applied electronic commerce research*, v. 9, n. 2, May, pp. 100-120.
<http://dx.doi.org/10.4067/S0718-18762014000200008>

Kim, Hyeoneui (2005). *Exploring data reusability: standardized representation of domain contents and feasibility testing*. Doctoral dissertation, University of Minnesota, USA. ISBN: 0 542 31045 7

Kitchin, Rob (2014). "The real-time city? Big data and smart urbanism". *GeoJournal*, v. 79, Febr., pp. 1-14.
<http://link.springer.com/article/10.1007%2Fs10708-013-9516-8>

LaValle, Steve; Lesser, Eric; Shockley, Rebecca; Hopkins, Michael S.; Kruschwitz, Nina (2011). "Big data, analytics and the path from insights to value". *MIT Sloan management review*, v. 52, n. 2, enero, pp. 21-32.

<http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value>

Lee, Gwanhoo; Kwak, Young-Hoon (2012). "An open government maturity model for social media-based public engagement". *Government information quarterly*, v. 29, Oct., pp. 492-503.
<http://dx.doi.org/10.1016/j.giq.2012.06.001>

Lessig, Larry (2007). "Eight principles of open government data". *Open government working group*.
<http://www.opengovdata.org/home/8principles>

Marcos-Martín, Carlos; Soriano-Maldonado, Salvador-Luis (2011). "Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta". *El profesional de la información*, v. 20, n. 3, mayo-junio, pp. 291-297.
<http://www.elprofesionaldelainformacion.com/contenidos/2011/mayo/07.pdf>
<http://dx.doi.org/10.3145/epi.2011.may.07>

Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma (2011). "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información*, v. 20, n. 2, marzo-abril, pp. 165-173.
<http://www.elprofesionaldelainformacion.com/contenidos/2011/marzo/06.pdf>
<http://dx.doi.org/10.3145/epi.2011.mar.06>

Pipino, Leo L.; Lee, Yang W.; Wang, Richard Y. (2002). "Data quality assessment". *Communications of the ACM*, April, v. 45, n. 4, pp. 211-218.
<http://web.mit.edu/t dq m / w w w / t dq m p u b / PipinoLeeWangCACMApr02.pdf>

Porter, Michael E.; Millar, Victor E. (1985). "How information gives you competitive advantage". *Harvard business re-*

view, v. 63, n. 4, pp. 149-160.

<http://www.im.ethz.ch/education/HS11/MIS/docu/Porter85.pdf>

Ramírez-Alujas, Álvaro; Dassen, Nicolás (2014). *Vientos de cambio: El avance de las políticas de gobierno abierto en América Latina y el Caribe*. Banco Interamericano de Desarrollo. Nota técnica #idb-tn-629.
<http://bit.ly/ramirez2014>

Red.es (2011). *Characterization study of the infomediary sector 2011*.
http://datos.gob.es/datos/sites/default/files/files/aporta_infomediary%20sector_2011.pdf

Red.es (2012). *Characterization study of the infomediary sector 2012*.
http://www.ontsi.red.es/ontsi/sites/default/files/121001_red_007_final_report_2012_edition_vf_en_1.pdf

Ren, Guang-Jie; Glissmann, Susanne (2012). "Identifying information assets for open data. The role of business architecture and information quality". En: *Procs. of the IEEE 14th Intl conf on commerce and enterprise computing*, Sept., pp. 94-100.
<http://doi.ieeecomputersociety.org/10.1109/CEC.2012.23>

Shapiro, Carl; Varian, Hal R. (1998). *Information rules*. Harvard Business Press, Boston. ISBN: 0 87584 863 X

Weiskopf, Nicole G.; Weng, Chunhua (2013). "Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research". *Journal of the American Medical Informatics Association*, January, v. 20, n. 1, pp. 144-151.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3555312>
<http://dx.doi.org/10.1136/amiajnl-2011-000681>

Nueva colección EPI Scholar

Libros académicos y científicos de Información y Documentación



Primer título:

Cibermetría. Midiendo el espacio red de Enrique Orduña-Malea e Isidro F. Aguillo

A pesar del crecimiento de la disciplina de la Cibermetría desde mediados de la década de los noventa, son escasos los libros académicos o manuales dedicados en exclusiva a la misma desde un contexto de las ciencias de la información y documentación. Este libro pretende cubrir este claro hueco en la literatura tanto nacional como internacional.



THE GENERATION OF LARGE NETWORKS FROM WEB OF SCIENCE DATA



Loet Leydesdorff, Gohar-Feroz Khan and Lutz Bornmann



Loet Leydesdorff (Ph.D. Sociology, M.A. Philosophy, and M.Sc. Biochemistry) is professor at the *Amsterdam School of Communications Research (ASCoR)* of the *University of Amsterdam*. He is Honorary Professor of the *Science and Technology Policy Research Unit (SPRU)* of the *University of Sussex*, Visiting Professor of the *Institute of Scientific and Technical Information of China (ISTIC)* in Beijing, Guest Professor at *Zhejiang University in Hangzhou*, and Visiting Professor at the *School of Management, Birkbeck, University of London*. He has published extensively in systems theory, social network analysis, scientometrics, and the sociology of innovation. With Henry Etzkowitz, he initiated a series of workshops, conferences, and special issues about the *Triple Helix of University-Industry-Government Relations*. He received the *Derek de Solla Price Award* for Scientometrics and Informetrics in 2003 and held "The City of Lausanne" Honor Chair at the *School of Economics, Université de Lausanne*, in 2005. In 2007, he was Vice-President of the 8th *International Conference on Computing Anticipatory Systems (Casys'07, Liège)*. In 2014, *Thomson Reuters* listed him as a highly-cited author.

<http://www.leydesdorff.net/list.htm>

<http://www.leydesdorff.net/th2/index.htm>

<http://highlycited.com>

<http://orcid.org/0000-0002-7835-3098>

University of Amsterdam, Amsterdam School of Communication Research (ASCoR)
PO Box 15793, 1001 NG Amsterdam, The Netherlands
loet@leydesdorff.net



Gohar-Feroz Khan is an Assistant Professor at the *Korea University of Technology and Education*. Since his PhD from *Kaist* in 2011, Khan has published over 30 articles in refereed journals, conference proceedings, and book chapters. His research has been published in *Online information review*, *Social science computer review*, *Government information quarterly*, *Journal of the American Society for Information Science and Technology*, *Scientometrics*, *Information development*, *Internet research*, *Asian journal of communication*, and *Asia Pacific journal of information system*, among others. His research interests include, 1) social information systems & social media, 2) e-government, government 2.0, and 3) networked science. Prior to his doctoral studies, Dr. Khan held a senior management position with the *Afghan Ministry of Communications and Information Technology*. He is also an Associate Editor of *Journal of contemporary Eastern Asia* and founding Director of *Center for Social Technologies*.

<http://gfkhan.wordpress.com/publications>

<http://orcid.org/0000-0003-2784-0918>

Korea University of Technology & Education (KoreaTECH)
1600 Chungjol-ro Byungcheon-myun
Cheonan city, 330-708, South Korea
gohar.feroz@kut.ac.kr



Lutz Bornmann works as a sociologist of science at the *Division for Science and Innovation Studies* in the administrative headquarters of the *Max Planck Society* in Munich (Germany). Since the late 1990s, he has been working on issues in the promotion of young academics and scientists in the sciences and on quality assurance in higher education. His current research interests include research evaluation, peer review and bibliometrics, and altmetrics. *Thomson Reuters* lists him among the most-highly cited researchers worldwide over the last ten years.

<http://highlycited.com>

<http://orcid.org/0000-0003-0810-7091>

Division for Science and Innovation Studies.
Administrative Headquarters of the Max Planck Society
Hofgartenstr., 8. 80539 Munich, Germany
bornmann@gv.mpg.de

Abstract

During the 1990s, one of us developed a series of freeware routines (<http://www.leydesdorff.net/indicators>) that enable the user to organize downloads from the *Web of Science* (Thomson Reuters) into a relational database, and then to export matrices for further analysis in various formats (for example, for co-author analysis). The basic format of the matrices displays each document as a case in a row that can be attributed different variables in the columns. One limitation to this approach was hitherto that relational databases typically have an upper limit for the number of variables, such as 256 or 1024. In this brief communication we report on a way to circumvent this limitation by using *txt2Pajek.exe*, available as freeware from <http://www.pfeffer.at/txt2pajek>

Keywords

Web of Science, Bibliometric network, *Pajek*, *txt2Pajek*.

Título: Generación de grandes redes a partir de datos de la *Web of Science*

Resumen

Durante la década de 1990, uno de nosotros desarrolló una serie de rutinas de software gratuito (<http://www.leydesdorff.net/indicators>) que permiten organizar las descargas desde la *Web of Science* (Thomson Reuters) en una base de datos relacional, y luego exportar matrices para su posterior análisis en varios formatos (por ejemplo, para el análisis de co-autores). El formato básico de las matrices muestra cada documento en una fila a la que se le pueden atribuir diferentes variables en las columnas. Una limitación que entonces tenía este enfoque era que las bases de datos relacionales suelen tener un límite superior en el número de variables, por ejemplo, 256 o 1.024. En esta breve comunicación se presenta una forma de eludir esta limitación utilizando *txt2Pajek.exe*, disponible como freeware en el url <http://www.pfeffer.at/txt2pajek>

Keywords

Web of Science, Redes bibliométricas, *Pajek*, *txt2Pajek*.

Leydesdorff, Loet; Feroz-Khan, Gohar; Bornmann, Lutz (2014). "The generation of large networks from *Web of Science* data". *El profesional de la información*, v. 23, n. 6, November-December, pp. 589-593.

<http://dx.doi.org/10.3145/epi.2014.nov.05>

Introduction

In recent decades, one of us has developed a series of software routines that enables the user to organize downloads from the *Web of Science* (Thomson Reuters) into a relational database, and then to export matrices for further analysis in various formats; for example, for co-author analysis, co-citation analysis, bibliographic coupling, etc. (Cobo *et al.*, 2011). The basic format of each matrix shows each document as a case in a row that can be attributed with different variables in the columns. Variables can be author names, institutional addresses, cited references, etc. One can also combine types of variables such as authors, title words, and institutional addresses (Leydesdorff, 2014; Vlieger; Leydesdorff, 2011). Multiplication of the asymmetrical word/document matrix with its transposed leads to a co-word matrix; and this operation can be done *mutatis mutandis* for other (sets of) variables attributable to documents. <http://www.leydesdorff.net/indicators>

One limitation to this approach was hitherto that relational databases typically have an upper limit for the number of variables, such as 256 or 1024, whereas the number of cases (documents) is limited only by considerations of disk space¹. In this brief communication, we report on a way to circumvent this limitation easily by using *txt2Pajek.exe* (Pfeffer; Mrvar; Bagatelj, 2013). *Txt2Pajek* enables the user to generate a 2-mode (asymmetrical) matrix of cases (documents) and variables in the *Pajek* format for an unlimited

number of variables from a text file. Within *Pajek* (De-Nooy; Mrvar; Batagelj, 2011) the newly generated 2-mode file can be further transformed into a 1-mode network file that can also be used in other software programs for network analysis and visualization such as *Gephi*, *UCInet*, or *VOSViewer*. <http://www.pfeffer.at/txt2pajek>

Data

One of us (GFK) encountered the systems limitation of 1024 variables when generating a co-author network at the level of institutional addresses using *instcoll.exe* for analysis and visualization. Using the eight journals listed in the so-called *Senior Scholars' Basket* of the *Association for Information Systems* (AIS) that were used for the ranking, 3,587 documents were downloaded for the period 1995-2014 (table 1). The set contains 7,397 institutional addresses, of which 4,617 are unique (Khan, in preparation). The author wished to pursue a network analysis using these names of institutions as nodes and had already organized the data download in a relational database using *isi.exe*². <http://www.leydesdorff.net/software/instcoll/index.htm> <http://www.leydesdorff.net/software/isi>

Analysis

The institutional names are organized by *isi.exe* in a separate table (named *cs.dbf*) that contains the document numbers for relational database management and the address information. Using *Excel* or a similar program, one can open

Table 1: The data of 3,587 documents in the eight journals (1995-2014) in the basket used by the *Association for Information Systems (AIS)* for ranking.

Journals	N
<i>European journal of information systems</i>	613
<i>Information systems journal</i>	341
<i>Information systems research</i>	549
<i>Journal of information technology</i>	447
<i>Journal of management information systems</i>	534
<i>Journal of strategic information systems</i>	331
<i>Journal of the Association for Information Systems</i>	239
<i>MIS quarterly</i>	533
Total	3,587

this table and save it as a comma-separated-variables (.csv) file or as tab-delimited. A program entitled *dbf2csv.exe* has additionally been made available at this url for a direct transformation:

<http://www.leydesdorff.net/software/dbf2csv/dbf2csv.exe>

The comma-separated files can be read as text files into *txt2Pajek.exe* and are transformed in 2-mode *Pajek* files.

One can further refine the address information by using functions of *Excel*. For example, the first address in the file was "Unist, Sch Technol Management, Ulsan, South Korea" in cell B2. Using the function "=left(B2, find(";",B2)-1)", one obtains the institutional name "Unist" in another cell (e.g., C2). Since institutional names are now considerably standardized in *WoS*, one can drag the function along the column in *Excel* and thus obtain a field with only institutional names. There are 1,364 unique institutional names in the set based on 3,564 (of the 3,578) documents. Similarly, one can extract country names on the right side of the string using more composed functions or by writing a routine³.

The .csv file should be re-named with the extension ".txt" and one should take care that the content is either lower or upper case (or capitalized case) because the default cases were changed in *WoS* during the 1990s. The transformation

by *txt2Pajek* is straightforward and provides a file with the same name, but with the extension ".net" in the *Pajek* format. This file can be read into *Pajek* or another network-analysis program that is able to read this format. The *Pajek* format is nowadays increasingly the standard currency for exchanges among network analysis and visualization programs.

Network analysis in *Pajek*

When the 2-mode network generated by *txt2pajek.exe* is read into *Pajek* (v.3), it can be transformed into a 1-mode network (in this case of institutes) under *Network > 2-Mode Network > 2-Mode to 1-Mode > Columns*. The option "multiple lines" should be set ON. Thereafter the multiple lines have to be summed under *Network > Create New Network > Transform > Remove > Multiple Lines > Sum Values*. The lines of the network (edges) can now be visualized with different widths. Similarly, one can size the nodes using "weighted degrees" for the number of occurrences under *Network > Create Vector > Centrality > Weighted Degree > All*. Using the Draw-menu now visualizes the network (*Draw > Network + First Partition + First Vector*; **Bruun**, 2009).

As would be expected, institutional collaboration networks contain lots of isolates, dyads, triads, etc. These small networks are not necessarily connected among themselves. The network under study thus contains 153 components with a largest component of 1,171 (85.9% of the 1,364) nodes. Figure 1 shows this largest component as a heat map after exporting to *VOSViewer* (**Van-Eck**; **Waltman**, 2010).

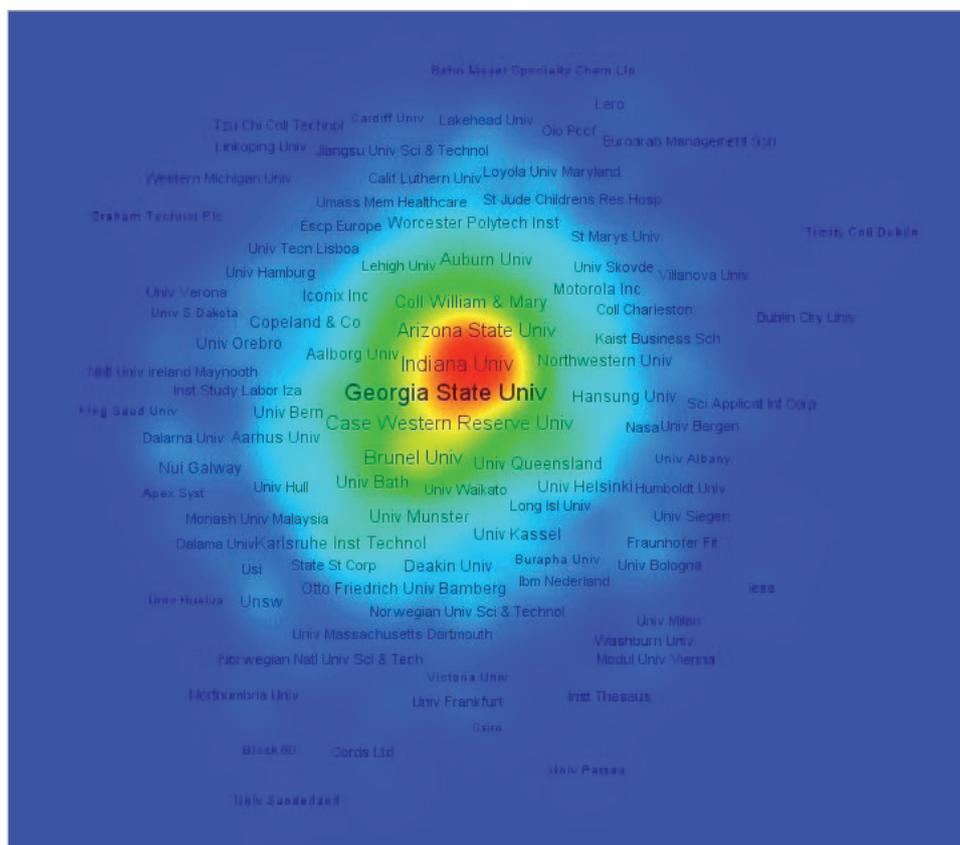


Figure 1. Heat map of the largest component (N = 1,171) of the network of institutional collaborations in the AIS-basket of 8 journals in "information systems".

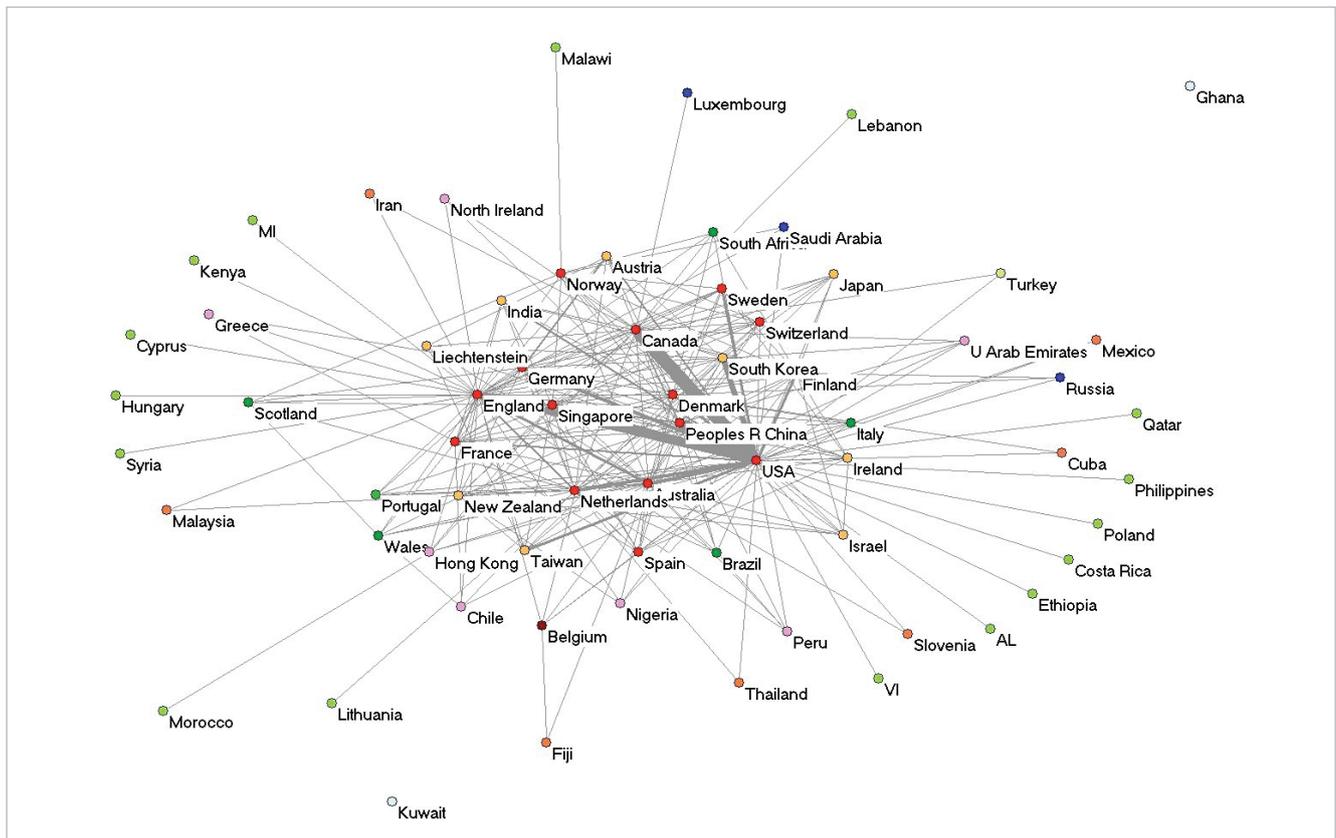


Figure 2. Network of international collaborations in the AIS basket; 67 countries; 3,563 documents. Kamada & Kawai (1989) used for the mapping.

The second largest component contains only nine institutes.

Figure 2 shows the network of 67 countries named in the institutional addresses of the 3,563 (99.6%) documents that provide such information. Note that in WoS, “England” is counted separately from the other countries of the UK.

Conclusions and summary

Using this pathway, one can visualize both smaller and very large networks, for example, of authors in large consortia (such as at CERN; Milojević, 2010). The routines *isi.exe* and *txt2pajek.exe* have no systems limitations except disk sizes. Bringing the files into network analysis and visualization programs, one can study degree distributions, clustering coefficients, modularity, etc., and visualize subsets accordingly. An alternative route for achieving this is provided by *Wos2Pajek*, but in this case the data is not organized relationally into databases. We have demonstrated the possibilities for analysis and visualization of collaborations in the specialty of “information systems” both at the institutional and international levels.

<http://pajek.imfm.si/doku.php?id=wos2pajek>

Notes

1. In a 32-bit operating environment, file sizes are limited to 2 GB, but this limitation is removed in the environment of a 64-bit operating system.
2. One can use *scopus.exe* at <http://www.leydesdorff.net/scopus> for transforming data from Scopus into this format.
3. The table *cs.dbf* already contains country names as a separate (third) field.

References

Bruun, Jesper (2009). *Physics and didactics: maps of text on scientific literacy*. <http://absalon.itslearning.com/jbruun/blog>

Cobo, Manuel-Jesús; López-Herrera, Antonio G.; Herrera-Viedma, Enrique; Herrera, Francisco (2011). “Science mapping software tools: Review, analysis, and cooperative study among tools”. *Journal of the American Society for Information Science and Technology*, v. 62, n. 7, pp. 1382-1402. <http://dx.doi.org/10.1002/asi.21525>

De-Nooy, Wouter; Mrvar, Andrej; Batagelj, Vladimir (2011). *Exploratory social network analysis with Pajek* (2nd edition). New York, NY: Cambridge University Press. ISBN: 978 0521174800

Khan, Gohar-Feroz (in preparation). *Roots and fruits of information system domain: a network perspective*.

Kamada, Tomihisa; Kawai, Satoru (1989). “An algorithm for drawing general undirected graphs”. *Information processing letters*, v. 31, n. 1, pp. 7-15. [http://dx.doi.org/10.1016/0020-0190\(89\)90102-6](http://dx.doi.org/10.1016/0020-0190(89)90102-6)

Leydesdorff, Loet (2014). “Science visualization and discursive knowledge”. In: Cronin, Blaise; Sugimoto, Cassidy (Eds.). *Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact*, pp. 167-185. Cambridge MA: MIT Press. ISBN: 978 0262525510

Milojević, Stasa (2010). “Modes of collaboration in modern science: beyond power laws and preferential attachment”.

Journal of the American Society for Information Science and Technology, v. 67, n. 7, pp. 1410-1423.

<http://arxiv.org/pdf/1004.5176.pdf>

<http://dx.doi.org/10.1002/asi.21331>

Pfeffer, Jürgen; Mrvar, Andrej; Batagelj, Vladimir (2013). *txt2pajek: Creating Pajek files from text files technical report*, CMU-ISR-13-110. Carnegie Mellon University: School of Computer Science, Institute for Software Research.

<http://www.pfeffer.at/txt2pajek/txt2pajek.pdf>

Van-Eck, Nees-Jan; Waltman, Ludo (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, v. 84, n. 2, pp. 523-538.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2883932>

<http://dx.doi.org/10.1007/s11192-009-0146-3>

Vlieger, Esther; Leydesdorff, Loet (2011). "Content analysis and the measurement of meaning: the visualization of frames in collections of messages". *The public journal of semiotics*, v. 3, n. 1, pp. 28.

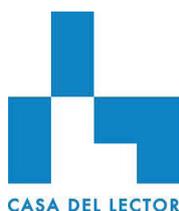
<http://www.leydesdorff.net/semiotics/semiotics.pdf>

1ª CONFERENCIA INTERNACIONAL SOBRE INDUSTRIA Y MERCADO DE LA INFORMACIÓN (Confimi)

Madrid, 5-6 de febrero de 2015

Organizada por:

- El profesional de la información;
- Biblioteca de la Universidad Complutense de Madrid; y
- Casa del Lector, de la Fundación Germán Sánchez Ruipérez.



El profesional de la
información

Dirigida a:

- responsables de adquisiciones y profesionales de la información de universidades, consorcios, bibliotecas virtuales, redes cooperativas, centros de investigación, empresas y administraciones;
- investigadores y profesores de biblioteconomía, documentación y ciencias de la información; y
- empresas proveedoras de contenido, tanto productoras como distribuidoras.

Objetivos:

- analizar problemáticas del mercado de la información (costes, valor, beneficios, análisis beneficio/coste, productividad, política de adquisiciones, suscripciones, evolución y tendencias...); y
- evaluar la oferta de nuevos productos y servicios que ofrecen los proveedores.

Más información
<http://confimi.info>





MUSEOS SOCIALES. PERFILES MUSEÍSTICOS EN TWITTER Y FACEBOOK 2012-2013



Florencia Claes y Luis Deltell



Florencia Claes trabaja en *Globomedia* y es investigadora, colaboradora honorífica y doctoranda en el *Departamento de Comunicación Audiovisual y Publicidad I* de la *Universidad Complutense de Madrid (UCM)*. Licenciada en comunicación audiovisual con premio extraordinario en 2010, su campo de estudio son las redes sociales y los líderes de opinión en espacios virtuales de España e Iberoamérica. Ha participado en congresos internacionales e investiga la relación de *Twitter* y el mercado audiovisual español e iberoamericano.
<http://orcid.org/0000-0003-4291-0507>

florenciaclaes@ucm.es



Luis Deltell es profesor contratado doctor en la *Universidad Complutense de Madrid (UCM)*. Su campo de investigación es la historia de la imagen y el análisis de las nuevas tecnologías de las ciencias de la información. Pertenece al grupo complutense de investigación sobre cine español. Ha sido *visiting scholar* en *Stanford University* y profesor invitado en *La Sapienza Università di Roma* y en *El Cairo University*. Trabaja en la medición de audiencias por medio de *Twitter*.
<http://orcid.org/0000-0002-5230-1409>

ldeltell@ccinf.ucm.es

*Universidad Complutense de Madrid
Facultad de Ciencias de la Información
Av. Complutense, s/n. 28040 Madrid, España*

Resumen

Se presentan las conclusiones de una investigación sobre la evolución del museo social o museo 2.0 y la comunicación entre visitantes virtuales y dichas organizaciones. Se han estudiado los perfiles en medios sociales de los principales museos y centros de arte de San Francisco, Londres y Madrid. El seguimiento y monitorización de estos perfiles se realizó durante un año desde diciembre de 2012 a diciembre de 2013. En esta investigación no sólo es importante el número de seguidores o de perfiles usados en la comunicación online, sino también las estrategias comunicativas y curatoriales que los centros llevan adelante. Se analiza el posible desarrollo de espacios de diálogo, interacción, participación y creación que conducen hacia el museo virtual y social pleno.

Palabras clave

Twitter, Facebook, Museo, Museo social, Comunicación, Audiencia, Cooperación, Participación, Internet, Red social, Medios sociales, Web 2.0, Estrategias, Influencia digital, Klout, Visitantes, Online.

Title: Social museums: social media profiles in *Twitter* and *Facebook* 2012-2013

Abstract

The results of a study on the development of a museum-2.0 or social museums and the communication between these organizations and their virtual visitors are presented. We based it on an analysis of the social media profiles of major museums and art centers in San Francisco, London and Madrid. Tracking and monitoring of these profiles was performed from December 2012 to December 2013. In this study, not only is it important how many followers or how many profiles are used in online media communication, but also the communication and curatorial strategies that these institutions apply. We explore the possible development of dialogue spaces, interaction, participation and creation that move toward the idea of a complete virtual and social museum.

Keywords

Twitter, Facebook, Museum, Social museum, Communication, Audience, Cooperation, Participation, Internet, Social networks, Social media, Web 2.0, Strategies, Digital influence, Klout, Visitors, Online.

Artículo recibido el 20-03-2014
Aceptación definitiva: 28-10-2014

Claes, Florencia; Deltell, Luis (2014). "Museos sociales. Perfiles museísticos en *Twitter* y *Facebook* 2012-2013". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 594-602.

<http://dx.doi.org/10.3145/epi.2014.nov.06>

1. Introducción

Pocas instituciones han experimentado una evolución tan grande como los museos en las últimas décadas. Si bien desde principios del siglo XX su función era ya cuestionada por futuristas y dadaístas, no es hasta finales del siglo cuando los centros de arte se transforman en una rémora para los creadores y los visitantes más inquietos. Ejemplo de esta actitud la representa **Isidoro Valcárcel-Medina** que se negó a exponer en el *Museo Nacional Centro de Arte Reina Sofía* (*Mncars*) y declaró que "los museos son mausoleos, siempre se ha dicho, pero hoy, los museos son cementerios, simplemente" (**Valcárcel-Medina**, 2003).

La museología crítica comenzó hace 30 años a "replantear la función y el espacio del museo y lucha por convertirlos en espacios de debate, reflexión y negociación" (**Martínez-Sanz**, 2012). Estos lugares tuvieron que cuestionarse el marketing y la "comunicación hacia y con los visitantes" (**Kotler; Kotler; Kotler**, 2008). Los museos, las filmotecas y los lugares de exhibición derivaron en espacios de arte y experimentación donde se pretende generar el debate y se aceptan, al menos en sus programas, otras formas y opiniones, incluso a aquellas que cuestionan a la propia institución. Sirva de ejemplo, la aportación de **Manuel Borja-Villel**, director del *Mncars*, que logró la elaboración de una exposición de **Valcárcel-Medina** en dicho centro en 2009.

La creación de un museo social globalizado no se materializa. El número de seguidores está relacionado directamente con el territorio próximo al espacio físico de la institución

La ruptura de la comunicación dominante de los museos y la apertura al diálogo y a la negociación con los visitantes se convierte en un tema clave en los inicios del siglo XXI. Es precisamente en esos años cuando surgen los medios sociales (*Facebook* en 2004 y *Twitter* en 2006). La web 2.0 supone un reto para los centros museísticos: una oportunidad y una condena al mismo tiempo. Como indica **Del-Río-Castro**, la implantación de las redes sociales "encuentra reticencias, al ser espacios poco serios o que afectan a la percepción profesional del museo" (**Del-Río-Castro**, 2011) y son criticadas desde los sectores más conservadores de las colecciones.

Sin embargo, los representantes de la museología crítica creen que internet y las redes sociales, además de "alzar y expandir el mensaje del museo, son un excelente canal de escucha que permite la bidireccionalidad y en consecuencia un *feedback*" (**Martínez-Sanz**, 2012). Las redes sociales se plantean como un vehículo para consolidar la promoción del museo e iniciar una nueva etapa de "participación equi-

tativa" entre centro y visitante (**Viñarás-Abad; Cabezuelo-Lorenzo**, 2012).

Ahora bien, algunos investigadores consideran que este diálogo no se produce en la actualidad. Muy especialmente el uso de internet muestra cómo no se materializa una "participación equitativa". Así, desde la implantación de las primeras webs de los museos se observa cierta incoherencia en las estrategias de comunicación y participación (**Sabin**, 1997). Incluso en fechas recientes, algunos analistas detallan cómo los espacios museísticos "están en una fase embrionaria de adopción de un nuevo modelo de gestión de sus públicos" aunque ya poseen una "tendencia a ampliar su gestión" (**Oliveira; Capriotti**, 2013). En similar línea se encuentran las nuevas teorías del marketing y la comunicación museística, que observan un cambio de planteamiento más participativo, pero que no acaba de fraguarse en la realidad diaria (**Gürel; Kavak**, 2010).

El último aspecto interesante es la desubicación virtual del museo. Internet posibilita la existencia no sólo de un visitante físico, sino también de un visitante virtual. "El espacio físico de la institución deja de ser la única vía para acceder al centro y empiezan a generarse nuevos caminos para conocer las colecciones museísticas. La evolución de este último modelo y la intervención directa del público como elemento activo en la construcción de la idea de museo han dado lugar a que actualmente no sólo exista un museo abierto, sino que empiece a hablarse ya de un museo social" (**Gómez-Vílchez**, 2012).

Este "museo social", o "museo 2.0" (**Simon**, 2010), incluye el uso de las redes sociales y las teorías de la museología crítica de finales del siglo XX y pretende expandir la experiencia de los visitantes más allá del espacio físico (**Sookhanaphibarn; Chatuporn**, 2013). El "museo social o 2.0" encaja plenamente en el universo de una "autocomunicación de masas" formulado por **Manuel Castells** (2009) donde los individuos ejercen una resistencia (y en consecuencia un poder) que los equilibra con el poder de las instituciones, en este caso los museos. Como indica **Nina Simon**, el museo 2.0 deberá ser una plataforma de creación, debate y negociación entre centro, comisarios, artistas y visitantes.

2. Objetivos e hipótesis

Nuestra investigación utiliza y se fundamenta en una rica bibliografía científica existente en torno a los museos y las redes sociales. Estos trabajos precedentes configuran un corpus joven pero valioso. El experimento concluyó en enero de 2014 y pretende monitorizar la actividad de los "museos sociales" para valorar su grado de participación y de diálogo en internet (*feedback*). Se recurrió a la monitorización durante un año (diciembre de 2012 a diciembre de 2013) de los perfiles en medios sociales de los principales museos de arte y filmotecas (o salas de reposición) de las ciudades de San Francisco, Londres y Madrid.

La selección se debe a los siguientes criterios:

a. Se han escogido tres países en los que los usuarios de internet superan el 70% de su población (*Banco Mundial*).

<http://www.worldbank.org>

b. Londres y Madrid son las capitales de sus respectivos países con un elevado número de turistas, mientras que San Francisco,

a pesar de ser una ciudad de segunda magnitud político-administrativa, alberga en su área metropolitana las principales universidades, centros de investigación, sedes de los medios sociales, como *Facebook* y *Twitter* y en las cercanías la matriz de *Google*.

c. Se encuentran entre las cincuenta urbes más visitadas del mundo (**Hedrick-Wong; Choog, 2013**).

d. Las tres presentan museos internacionales de primera categoría en arte contemporáneo, arte tradicional, europeo y no europeo. Los principales centros de dichas ciudades reciben más de un millón de visitas físicas anuales (*INE, Office for National Statistics y California Department of Finance*).

<http://www.ine.es>

<http://www.statistics.gov.uk/hub/index.html>

<http://www.dof.ca.gov>

Así hemos escogido las siguientes organizaciones:

1) San Francisco

- *Asian Art Museum*
- *De Young Museum*
- *San Francisco Museum of Modern Art (Sfmoma)*.

2) Londres

- *The British Museum*
- *The National Gallery*
- *Tate*.

3) Madrid

- *Museo Nacional del Prado*
- *Museo Nacional Centro de Arte Reina Sofía (Mncars)*
- *Museo Thyssen Bornemisza*.

Además, se han monitorizado los perfiles de otros espacios como medición de control. Aunque éstos no cumplían todos los parámetros de la muestra nos resultan interesantes para explicar algunos procesos:

- *Legion of Honor*
- *Saatchi Gallery*
- *Design Museum*
- *Victoria and Albert Museum*
- *La Casa Encendida*
- *Caixa Forum Madrid*
- *Museo Nacional del Romanticismo*.

Especial interés aportan los datos de las instituciones culturales menores como las filmotecas y cines de reposición, *rep theater* o *repertory theater*:

- *Castro Theatre*

Tabla 1. Datos básicos de las ciudades

Año 2012	San Francisco	Londres	Madrid
Población	825.111	9.787.426	3.207.247
Densidad hab/km²	6.818	5.285	5.295
Población metropolitana	4.203.898	15.010.295	6.369.162
Turistas totales que pasan la noche	2.300.000	16.900.000	9.700.000
Visitantes totales a los 3 museos estudiados	2.408.170	16.318.688	6.532.860
Usuarios de internet en el país (por cada 100 habitantes)	81	87	72

- *British Film Institute*
- *Filmoteca Española (Cine Doré)*.

Estos usuarios muestran la tendencia virtual entre las organizaciones con menos presupuesto y actividad.

Objetivos del experimento:

a. Monitorización del número total de amigos y seguidores de los perfiles de los museos en las redes sociales y en *Twitter* durante doce meses.

b. Medir y registrar el impacto de los museos en la Red, flujo de actividad y factor *Klout* durante el período del experimento.

c. Monitorizar la actividad y participación de los perfiles de los museos. Cuantificar la interacción entre los visitantes, seguidores y organizaciones.

d. Utilizar los datos obtenidos y contrastarlos con las estrategias comunicativas de los centros.

e. Comparar estos datos con las actividades y las propuestas curatoriales de los museos.

Nuestra hipótesis de trabajo es que los medios y las redes sociales han supuesto un profundo cambio en el planteamiento de los espacios museísticos y que éstos han comenzado una transformación de local a global. Es decir, de la dedicación exclusiva a un público físico-presencial se deriva a una dedicación compartida con el público virtual.

“ Todos los grandes museos han creado un gran sistema de medios sociales ”

3. Metodología

Este estudio se basa en una monitorización de datos cuantitativa y en un posterior análisis cualitativo de los resultados. El corpus lo forman los perfiles de los museos y centros de arte más importantes de las mencionadas ciudades. En la tabla 2 se muestran los datos del experimento.

Para la confección de nuestras tablas hemos utilizado las siguientes herramientas online:

- *Museum Analytics* es el referente internacional básico para la medición de la actividad de los perfiles museísticos en *Facebook* y *Twitter*. Los resultados de esta web, que aún se encuentra en su versión beta, son cuantitativos. <http://www.museum-analytics.org>

- *Twitalyzer*
<http://www.twitalyzer.com>
- *Hootsuite*
<http://www.hootsuite.com>
- *Klout*
<http://www.klout.com>
- los buscadores avanzados de los 8 medios sociales investigados y muy especialmente el sistema universitario de monitorización de *Twitter T-Hoarder*.
<http://www.t-warder.com>
- *SocialWin*
<http://www.socialwinapp.com>
- *LikeAlyzer*
<http://www.likealyzer.com>

Estas cifras resultan valiosas y sirven para contrastar nuestra propia captura de información. Sin embargo, como indica **Elena Villaespesa** “el mero hecho de contar números no tiene valor en sí mismo, a no ser que los relacionemos con los objetivos del museo” (Villaespesa, 2013).

Se ha monitorizado el comportamiento de los perfiles de los usuarios de los museos escogidos pero, a la vez, se ha llevado a cabo un seguimiento de las exposiciones, de los eventos que dichos centros han realizado a lo largo de un año y de las propuestas interactivas dentro del ámbito físico. Es decir, pretendemos investigar todo el “ecosistema digital y social del museo que se compone básicamente de la web, canales en redes sociales y la incorporación de las tecnologías en el museo” (Villaespesa, 2013), y los datos obtenidos son comparados con las estrategias curatoriales comunicativas de estas instituciones. No se trata sólo de obtener o “contar números” sino de entender la “plataforma completa” que se crea en el ámbito físico y virtual entre museo y espectador-autor (Simon, 2013).

Uno de los aspectos fundamentales para mostrar el liderazgo y el impacto de los perfiles de los museos en *Twitter* es seguir el factor *Klout*, como recomienda Soler-Herreros (2013), pero también el número de seguidores y el alcance de sus tuits por medio de los retuits y menciones (Deltell; Claes; Osteso, 2013). Por ello estudiamos centros monitorizando los seguidores y la evolución del factor *Klout*, que es un baremo de la influencia de un usuario en varias redes sociales (Serrano-Puche, 2012).

La utilización de *Twitter*, y el volumen del flujo de tuits para predecir el comportamiento social es algo extendido y suele recurrirse a él para mediciones sociales diversas, predicción de taquillas cinematográficas (Ishii et al., 2012), elecciones políticas (Tumasjan et al., 2010) –investigación fuertemente rebatida en la actualidad- o audiencias televisivas (Congosto et al., 2013). No es de extrañar que también se haya propuesto en el entorno museístico. Este trabajo atiende también a la posible relación entre el flujo de tuits y el número total de visitantes físicos.

<http://www.museumnerd.org>

Tabla 2. Datos del experimento

Espacios museísticos de referencia	Asian Art Museum, De Young Museum, San Francisco Museum of Modern Art. The British Museum, The National Gallery, Tate. Museo Nacional del Prado, Museo Nacional Centro de Arte Reina Sofía, Museo Thyssen-Bornemisza.	
Espacios museísticos de control	Legion of Honor, Saatchi Gallery, Design Museum, Victoria and Albert Museum, La Casa Encendida, Caixa Forum Madrid, Museo Nacional Romántico	
Filmotecas	Castro Theatre, British Film Institute, Cine Doré	
Total de seguidores en medios sociales	Google+	3.556.689
	Facebook	2.190.406
	Instagram	124.326
	Pinterest	46.522
	YouTube	41.121
	Flickr (nº fotos)	24.189
	Storify	1.008
	Tumblr	no cuantificable
Fechas de control	Diciembre 2012	Diciembre 2013
Total seguidores Twitter	1.797.171	2.583.550

4. Resultados y discusión

Esta investigación confirma que los responsables de los museos construyen entornos virtuales y pretenden favorecer espacios de intercambio con los visitantes online. Sin embargo, como se detalla en la bibliografía científica desde 1997, con el estudio de las webs museísticas (Sabin, 1997), hasta las más recientes publicaciones (Oliveira; Capriotti, 2013) se observa que las organizaciones se encuentran desorientadas y que sus intervenciones no siempre favorecen el establecimiento de un museo social pleno. Así todas las instituciones han creado usuarios en los medios sociales más relevantes (tabla 3).

“ Para la consolidación de un museo social no sólo es importante el número de seguidores, sino la participación de éstos ”

Al comparar las cifras de los museos principales con las de las instituciones menores se descubren datos interesantes. La primera característica es que los espacios con menos recursos (filmotecas y centros pequeños, es decir, las instituciones de control en nuestro experimento) presentan unas estrategias comunicativas más elementales e, incluso, a veces nulas. Por ejemplo, la *Filmoteca Española (Cine Doré)* carece de web propia, espacio en *Facebook* y cuenta oficial en *Twitter*. Estas carencias, a veces, son suplidas por los usuarios más activos que crean *fakes* o perfiles falsos sobre las organizaciones (sería el caso de @cineDore que durante un año operó como la cuenta oficial sin serlo). Los encargados de comunicación de los centros menores reconocen la importancia de internet, pero al mismo tiempo carecen de los recursos necesarios para llevar a cabo sus planes. Esta actitud la confirman los responsables de la web del Museo

Tabla 3. Perfiles museísticos en medios sociales (sin Twitter) en 2013

	Institución	Visitantes físicos 2013	Facebook	Google plus	Pinterest	Flickr (nº fotos)	Instagram	Tumblr	YouTube	Storify
San Francisco	Asian Art Museum	200.458 (2012)	23.986	3	1.944	9.463	4.615	Sí	3.587	53
	De Young Museum	1.588.654 (2012)	53.328	1.073.858	1.039	no	6.230	Sí	591	No
	San Francisco Museum of Modern Art	562.676 (2012)	110.145	1.143.303	27.258	8.877	70.549	Sí	3.346	282
Londres	The British Museum	5.800.000	476.805	14.943	5.998	1.907	10.523	Sí	4.203	170
	The National Gallery	5.400.00	288.456	1.947	1.603	No	102	No	2.554	No
	Tate	5.318.688 (2012)	635.794	853.239	5.449	605	31.395	Sí	23.418	214
Madrid	Museo Nacional del Prado	2.306.966	319.686	466.858	1.594	No	No	No	3.422	170
	Museo Nacional Centro de Arte Reina Sofía	3.184.640	201.279	44	No	3.122	No	No	756	4
	Museo Thyssen-Bornemisza	944.827	80.927	553	1.637	215	912	Sí	463	115

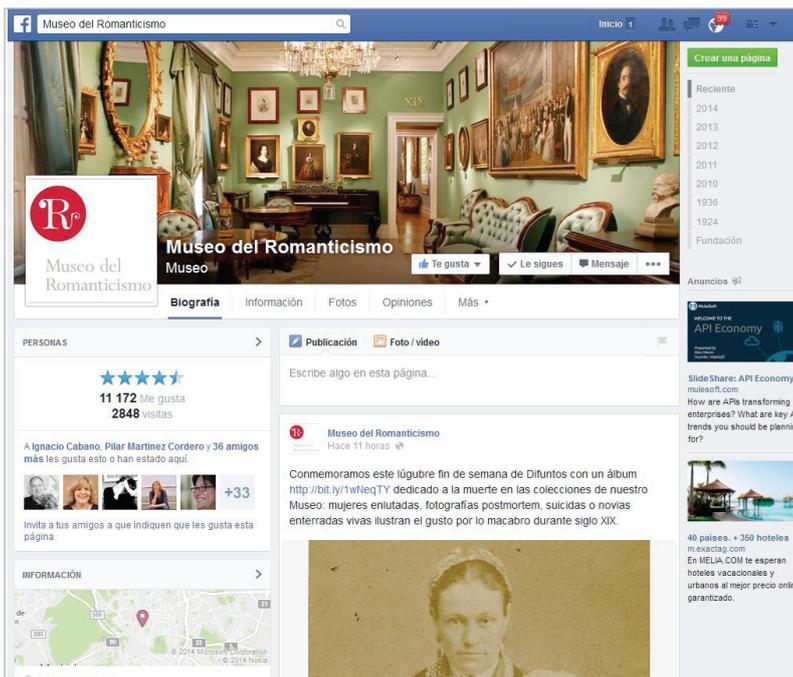
Nacional del Romanticismo que si bien son conscientes del valor de internet no disponen de presupuesto, lo que les obliga a utilizar las versiones gratuitas de Facebook y otras redes.

El ejemplo de Museo Nacional del Romanticismo es interesante como indican Cabrera-Bravo y Cabrejas-Almena, que han colaborado en los departamentos de difusión de varios museos (incluido éste), el centro era “absolutamente consciente de la importancia” de los medios sociales (Cabrera-Bravo; Cabrejas-Almena, 2013, p. 89). Así el museo ha desarrollado una atractiva actividad en Facebook, Insta-

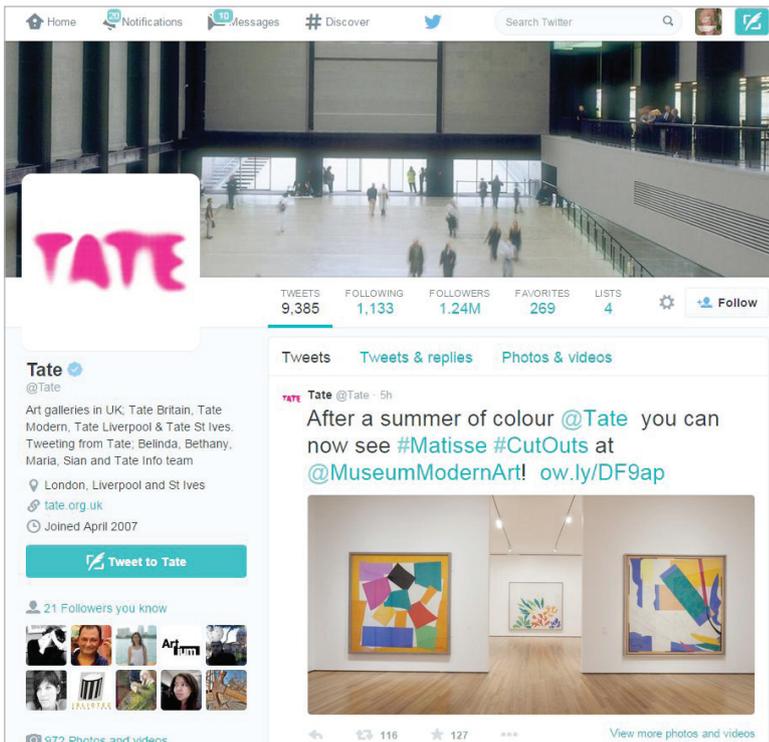
gram, Slideshare, Twitter y otros. Pero no tienen recursos económicos para crear apps y juegos propios. Incluso, en 2013, como indican las autoras citadas, ni siquiera disponían de “una persona especializada y formada a priori en esos temas, ni que tenga dedicación plena a los mismos, de modo que las personas encargadas de la web 2.0 se han ido formando, en muchas ocasiones, de manera autodidacta y a medida que iban realizando estas funciones” (Cabrera-Bravo; Cabrejas-Almena, 2013, p. 90).

Por el contrario los centros con mayores recursos (Museo Nacional del Prado, British Museum, Tate y Sfmoma) son los que disponen de mejores espacios online y mayor impacto en la Red. Todos ellos se esfuerzan en realizar una comunicación amplia y más equitativa con el visitante; un ejemplo de ello es la principal pinacoteca matritense. Hasta hace algo más de un año mantenía una estrategia clásica de información: “se monopoliza rigidamente la dirección y toma la batuta de la orquesta en el discurso digital. Sólo se habla de lo que el museo propone y es el museo quien elige los temas” (Viñarás-Abad; Cabezuelo-Lorenzo, 2012). Ahora el museo establece conversaciones con los visitantes virtuales y participa activamente en debates no generados por la propia organización.

Sin embargo existen museos pequeños que han logrado un excelente resultado en las redes sociales. El caso más destacado es el de Saatchi Gallery. Los perfiles de este centro, tanto en Facebook como en Twitter, gozan de un gran número de seguidores. Varias son las estrategias comunicativas y curatoriales que permiten este éxito:



Página del Museo del Romanticismo en Facebook <https://www.facebook.com/pages/Museo-del-Romanticismo/311260926718>



Página de Tate en *Twitter*
<https://twitter.com/tate>

- un diseño y una gran agilidad en los contenidos (que emula a las mejores campañas publicitarias de *Saatchi & Saatchi*);
- el carácter gratuito de sus exposiciones;
- el centrar su estrategia curatorial en exposiciones temporales.

El éxito online del espacio creado por Charles Saatchi también se repite en los visitantes físicos: cuatro de las siete exposiciones más vistas en 2012 en Londres fueron de dicho centro.

<http://www.museum-analytics.org>
<http://www.theartnewspaper.com>

“ El crecimiento de visitantes físicos no equivale, de forma directa, al aumento en el número de visitantes virtuales ”

Otra estrategia curatorial que comparten algunos de los grandes museos es la *gamificación* o ludificación. Así desde hace una década cuando en la sede central de *Tate* se instalaron los grandes toboganes del artista belga Carsten Höller, el centro británico ha apostado por el juego y por la gamificación ofreciendo al espectador actividades lúdicas. Esta estrategia curatorial se ha reforzado con las redes sociales. Las sedes de *Tate* ofrecen juegos y aplicaciones en red (hasta un total de 16 en diciembre de 2013). Incluso los museos de contenidos de arte no contemporáneo también se han sumado a la gamificación, como el *Metropolitan* con “Murder at the Met” o el *Museo Nacional del Prado* que ha creado toda una sección de “juegos” en su web (*Dosdoce.com*; *EndeComunicación*, 2013, p. 8).

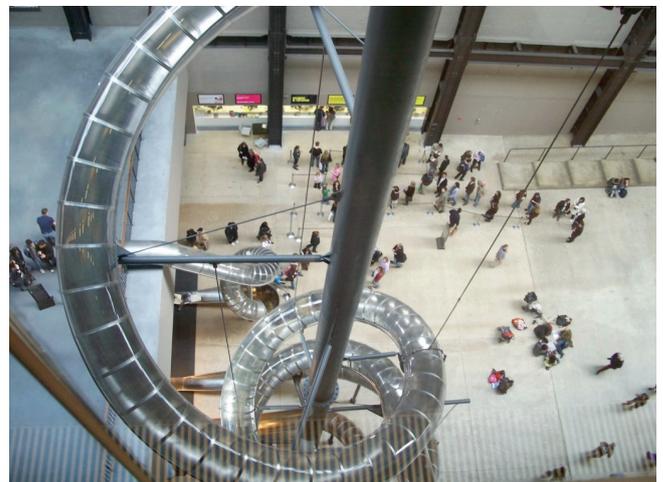
En el caso del *Mncars* resulta interesante su estrategia curatorial. Su actitud como museo físico es claramente de puertas abiertas y de negociación con visitantes y artistas. Durante el curso 2012-2013 ha realizado invitaciones a asociaciones vecinales, propuestas de residencia y actividades de intercambio con centros menores. Sin embargo, su web y sus perfiles en *Facebook* y *Twitter* muestran una actitud dominante y vertical, opuesta a su ideario curatorial. El museo virtual no participa en los debates y no genera conversaciones equitativas con sus visitantes virtuales. A pesar de esta actitud existe un considerable aumento de visitantes físicos y de seguidores en los medios sociales. Pero este crecimiento se debe únicamente a la celebración de una exposición tradicional dedicada al pintor Salvador Dalí (*Koch*, 2013).

Una de las preguntas centrales de esta investigación es si los museos interactúan con los usuarios, es decir, si mantienen un *feedback* con ellos. Todos los perfiles museísticos tanto en *Facebook* como en *Twitter* responden y comentan algunos de los tuits de los usuarios

aunque en ningún caso se trata de un diálogo igualitario. Sólo dos museos (tabla 4) mantienen un *feedback* mínimo o casi inexistente en estas redes sociales.

Uno de los datos más reveladores de nuestra muestra es el comportamiento en *Twitter* de los perfiles museísticos más importantes en los años 2012 y 2013 (tabla 4).

La tabla 4 muestra que todos los perfiles crecen en número de seguidores durante el período del experimento. No obstante, el aumento es significativamente distinto en las diversas ciudades. En San Francisco se percibe un freno en el incremento, incluso dos de sus museos llegan a perder puntos en su factor *Klout*. Esta cifra es importante ya que en dicha área metropolitana se encuentran las sedes de *Twitter* y *Facebook* y estos museos fueron los primeros en apostar de forma significativa por los medios sociales.



Toboganes de Carsten Höller en la *Tate Modern*, Londres

Tabla 4. Comparación *Twitter* años 2012 y 2013

	San Francisco			Londres			Madrid		
	<i>Asian Art Museum</i>	<i>De Young Museum</i>	<i>San Francisco Museum of Modern Art</i>	<i>The British Museum</i>	<i>The National Gallery</i>	<i>Tate</i>	<i>Museo Nacional del Prado</i>	<i>Museo Nacional Centro de Arte Reina Sofía</i>	<i>Museo Thyssen Bornemisza</i>
Visitantes 2012	187.000	1.588.654	632.516	5.600.000	5.400.000	5.318.688	2.711.651	2.565.000	1.256.209
Visitantes 2013	----	----	----	5.800.000	5.400.000	----	2.306.966	3.184.640	944.827
Apertura perfil	03/12/08	13/12/08	30/09/08	16/01/09	27/04/10	19/04/07	03/10/08	01/02/09	06/04/09
Total días en activo (23 dic 2013)	1.846	1.836	1.910	1.802	1.336	2.440	1.907	1.786	1.722
Seguidores 2012	155.024	114.760	366.665	130.441	54.199	771.417	139.341	37.151	28.173
Seguidores 2013	214.849	148.912	422.597	244.545	175.492	998.926	215.430	92.859	69.939
Seguidores adquiridos en 2013	59.825	34.152	55.932	114.104	121.293	227.509	76.089	55.708	41.766
Nuevos seguidores diarios hasta 2012	105	78	237	91	56	372	90	26	21
Nuevos seguidores diarios en 2013	164	94	153	313	332	623	208	153	114
Seguidos 2012	1.053	1.057	244.270	677	2.781	893	321	729	2.963
Seguidos 2013	1.477	1.205	240.191	14.016	3.184	1.001	615	884	1.743
Actualizaciones (hasta 2012)	3.025	4.507	5.926	3.778	4.131	4.700	7.223	2.037	2.177
Actualizaciones (hasta 2013)	3.763	6.674	8.849	6.639	6.445	7.300	14.255	4.214	8.129
Actualizaciones en 2013	738	2.167	2.923	2.861	2.314	2.600	7.032	2.177	5.952
Publicaciones diarias hasta 2012	2	3	4	3	4	2	5	1	2
Publicaciones diarias en 2013	2	6	8	8	6	7	19	6	16
Seguidores nuevos por tuit en 2013	81	16	19	40	52	88	11	26	7
Factor Klout 2012	68	63	82	90	67	87	86	64	63
Factor Klout 2013	63	63	80	92	66	87	89	65	68
¿Interactúan con los usuarios?	Mínimo	Sí	Sí	Sí	Sí	Sí	Sí	Mínimo	Sí
Generan hashtags	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí

La disminución en el ritmo del crecimiento en *Twitter* se puede deber a diversos motivos:

- La situación catastrofista que auguran para *Facebook* algunos profesores de *Princeton University*; según esta teoría, la red social perderá el 80% de sus usuarios antes del 2016 (Cannarella; Spechler, 2014). Sin embargo, excede a nuestra investigación comprobar su negación o aceptación.
- Que en San Francisco se haya alcanzado el máximo de usuarios potenciales. Es decir, en ese ámbito geográfico todos los internautas interesados ya seguirían a los perfiles museísticos. Esto podría justificarse con los datos de la tabla 1. El límite poblacional-región supondría una explicación plausible para esta tendencia al estancamiento pero, a la vez, mostraría el fracaso del museo global como

un espacio virtual sin fronteras. Ya que el museo virtual sólo sería seguido mayoritariamente por aquellos usuarios que viviesen en la proximidad del centro, por ello la creación de un museo global no se materializaría.

Existe una excepción en nuestro experimento a esta desorientación generalizada que observamos entre los museos. Es el caso *Tate*, que ofrece un comportamiento casi pleno de museo social. La propia estrategia curatorial del museo ya revelaba la transformación de este espacio en un museo 2.0:

“Lo digital solía ser la preocupación de sólo un departamento en la *Tate*, pero pronto impregnó todas las áreas de trabajo del museo. Esta transición requerirá el nivel adecuado de dotación de recursos, implicación de la dirección y el compromiso de toda la organización” (Stack, 2013).

Es cierto que, a parte de la estrategia curatorial, el posicionamiento de *Tate* como organización virtual lo favorecen muchos factores: una ciudad con una gran población, un volumen elevadísimo de turistas y la gratuidad de la mayoría de las exposiciones. A estos factores se debe añadir su excelente uso de los medios sociales, su presencia en casi la totalidad de los mismos y una brillante estrategia de comunicación (Villaespesa, 2013). Ya en 2010 sentaron las bases de *Tate online strategy* y para alcanzar sus objetivos expusieron que la web y el uso de los medios sociales componían “la quinta galería *Tate*” (Stack, 2010). La tabla 5 muestra los datos de *Tate* en los medios.

Huelga indicar que los contenidos de *Tate* no son similares a los de los otros museos. Este centro parece beneficiarse no sólo de una estrategia más abierta, sino también de un público más receptivo a esta nueva comunicación (museo-visitantes). Como observa Nina Simon, el nuevo museo lo formarán la institución y, también, los visitantes físicos y virtuales.

5. Conclusiones

Se muestra cómo todos los museos y centros de arte estudiados evidencian un gran interés en los medios sociales. Sin embargo los resultados de la monitorización de los perfiles de estas organizaciones son muy distintos.

Los centros menores, estudiados como casos de control, presentan estrategias de comunicación más sencillas. En la mayoría de ellos sus proyectos virtuales son ambiguos y contradictorios y no reflejan claramente el ideario curatorial. Al contrario, los grandes museos realizan una gran actividad en los medios sociales, aunque no todos muestran una estrategia coherente con sus idearios.

En la actualidad el gran museo social o museo 2.0 es *Tate*

El análisis de los perfiles museísticos entre 2012 y 2013 revela su gran crecimiento en la casi totalidad de las organizaciones. El aumento en el factor *Klout* y el número de seguidores manifiestan un mayor impacto social.

Sin embargo, este aumento parece alcanzar un estancamiento en San Francisco, donde se puede intuir que se ha llegado al listón máximo de seguidores. Por ello parece que actualmente estos museos no consolidan la estrategia de un museo social y global. El impacto online de estas organizaciones se debe relacionar directamente con el territorio donde se ubican.

De todos los perfiles museísticos estudiados destaca el de *Tate*, que está logrando un museo virtual y social pleno. No sólo por el elevado número de seguidores sino, sobre todo,

Tabla 5. Datos de los medios sociales de *Tate*

Twitter (19 cuentas)	Seguidores	Facebook (8 cuentas)	Me gusta
<i>Tate</i>	1.040.000	<i>Tate</i>	643.237
<i>Tate Liverpool</i>	43.500	<i>Tate Liverpool</i>	25.015
<i>Tate St. Ives</i>	12.200	<i>Tate St. Ives</i>	6.831
<i>Tate Kids</i>	3.472	<i>Tate Members</i>	7.689
<i>Tate Collectives</i>	3.617	<i>Tate Film</i>	18.000
<i>Tate Shop</i>	5.082	<i>Tate Shop</i>	4.506
<i>Artist Rooms</i>	1.912	<i>Tate Etc.</i>	2.851
<i>Tate Etc.</i>	16.700	<i>Tate Collectives</i>	151
<i>Tate Film</i>	1.619	Pinterest	Seguidores
<i>TateShots</i>	36.500	<i>Tate Gallery</i>	5.567
<i>Tate Publishing</i>	4.243	Google+	Seguidores
<i>Tate Teachers</i>	1.398	<i>Tate</i>	868.311
<i>Tate Live</i>	2.404	Instagram	Seguidores
<i>Tate Young Patrons</i>	1.484	<i>Tate</i>	33.705
<i>Tate Food</i>	374	<i>Tate Collectives</i>	2.107
<i>Tate Research</i>	1.703	Flickr	Fotos
<i>Martin Myrone</i>	216	<i>Tate</i>	605
<i>Helen Little</i>	1.591	Tumblr	
<i>Elvira Dyangani Ose</i>	484	<i>Tate y Tate Collectives</i>	
YouTube (2 cuentas)	Suscriptores	Reproducciones	
<i>Tate</i>	23.580	4.644.977	
<i>Great British Art Debate</i>	70	25.627	

por la interacción, la participación, la negociación y la creación que se genera entre visitantes virtuales e institución.

5. Bibliografía

Cabrera-Bravo, María-Jesús; Cabrejas-Almena, María-Carmen (2013). “Del XIX al XXI: el Museo del Romanticismo en internet”. *Revista de museología*, n. 56, pp. 88-100.

Cannarella, John; Spechler, Joshua A. (2014). *Epidemiological modeling of online social network dynamics*. Paper, Department of Mechanical and Aerospace Engineering, Princeton University.
<http://arxiv.org/pdf/1401.4208v1.pdf>

Castells, Manuel (2009). *Comunicación y poder*. Madrid: Alianza Editorial. ISBN: 978 8420684994

Congosto, Mari-Luz; Deltell, Luis; Claes, Florencia; Osteoso, José-Miguel (2013). “Análisis de la audiencia social por medio de *Twitter*. Caso de estudio: los premios Goya 2013”. *Icono 14*, v. 11, n. 2, pp. 53-82.
<http://www.icono14.net/ojs/index.php/icono14/article/view/577>
<http://dx.doi.org/10.7195/ri14.v11i2.577>

Del-Río-Castro, José-Nicolás (2011). “Museos y redes sociales, más allá de la promoción”. *Redmarka: Revista académica de marketing aplicado*, v. 3, n. 7, pp. 111-123.
<http://dialnet.unirioja.es/servlet/articulo?codigo=4126657>

- Deltell, Luis; Claes, Florencia; Osteso, José-Miguel** (2013). "Audiencias televisivas y líderes de opinión en Twitter. Caso de estudio: *El barco*". *Estudios sobre el mensaje periodístico*, v. 19, n. 1, pp. 347-364.
<http://revistas.ucm.es/index.php/ESMP/article/view/42526>
- Dosdoce; EndeComunicación (2013). "Los museos en la era digital. Uso de nuevas tecnologías antes, durante y después de visitar un museo, centro cultural o galería de arte". *Dosdoce.com*, 14 mayo.
<http://www.dosdoce.com/articulo/estudios/3820/museos-en-la-era-digital>
- Art newspaper* (2013). "Visitors figures 2012. Exhibition & museum attendance survey". *The art newspaper*, section 2, n. 245, April.
<http://www.theartnewspaper.com/attfig/attfig12.pdf>
- Gómez-Vílchez, Soledad** (2012). "Museos españoles y redes sociales". *Telos, cuadernos de comunicación e innovación*, n. 90, enero-marzo, pp. 1-8.
<http://goo.gl/QjqgBj>
- Gürel, Eda; Kavak, Bahtisen** (2010). "A conceptual model for public relations in museums". *European journal of marketing*, v. 44, n. 1/2, pp. 42-65.
- Hedrick-Wong, Yuwa; Choog, Desmond** (2013). "Global destination cities index". *MasterCard*.
<http://insights.mastercard.com/position-papers/top-20-global-destination-cities-in-2013>
- Ishii, Akira; Arakaki, Hisashi; Matsuda, Naoya; Umemura, Sanae; Urushidani, Tamiko; Yamagata, Naoya; Yoshida, Narihiko** (2012). "The 'hit' phenomenon: a mathematical model of human dynamics interactions as a stochastic process". *New journal of physics*, v. 14, n. 6.
http://iopscience.iop.org/1367-2630/14/6/063018/pdf/1367-2630_14_6_063018.pdf
<http://dx.doi.org/10.1088/1367-2630/14/6/063018>
- Koch, Tommaso** (2013). "El *Reina Sofía* prevé alcanzar los 3,2 millones de visitas en 2013". *El país*, 3 diciembre.
http://cultura.elpais.com/cultura/2013/12/03/actualidad/1386090272_552967.html
- Kotler, Neil G.; Kotler, Philip; Kotler, Wendy I.** (2008). *Museum marketing and strategy: designing missions, building audiences, generating revenue and resources*. San Francisco: John Wiley & Sons. ISBN: 978 0 7879 9691 8
- Martínez-Sanz, Raquel** (2012). "Estrategia comunicativa digital en el museo". *El profesional de la información*, v. 21, n. 4, pp. 391-395.
<http://eprints.rclis.org/17311/1/391-395.pdf>
<http://dx.doi.org/10.3145/epi.2012.jul.10>
- Oliveira, Andrea; Capriotti, Paul** (2013). "Gestión estratégica de los públicos en museos. De la identificación a la comunicación". *El profesional de la información*, v. 22, n. 3, pp. 210-214.
<http://www.kulturklik.euskadi.net/wp-content/uploads/2013/10/E130202.pdf>
<http://dx.doi.org/10.3145/epi.2013.may.03>
- Sabin, Richard** (1997). "Museums and their websites: An examination and assessment of how museums are coping with the challenge of the world wide web". *Journal of conservation and museum studies*, v. 11, pp. 6-10.
<http://www.jcms-journal.com/article/view/jcms.2972/7>
<http://dx.doi.org/10.5334/jcms.2972>
- Serrano-Puche, Javier** (2012). "Herramientas web para la medición de la influencia digital: análisis de Klout y PeerIndex". *El profesional de la información*, v. 21, n. 3, pp. 298-303.
<http://dadun.unav.edu/handle/10171/22464?mode=simple>
<http://dx.doi.org/10.3145/epi.2012.may.11>
- Simon, Nina** (2010). *The participatory museum*. Santa Cruz, California: Museum 2.0. ISBN: 978 06153446502
- Soler-Herrerros, Chimo** (2013). "Marketing global low cost de la cultura y el arte". *Revista de museología*, n. 56, pp. 19-24.
- Sookhanaphibarn, Kingkarn; Chatuporn, Utaiwan** (2013). "Expanding the experience of museum visitors with a social application on Facebook". En: *ICDS 2013 The 7th Intl conf on digital society*, pp.74-80.
http://www.thinkmind.org/index.php?view=article&articleid=icds_2013_3_40_10179
- Stack, John** (2010). "Tate online strategy 2010-2012". *Tate.org*, 1 April.
<http://www.tate.org.uk/research/publications/tate-papers/tate-online-strategy-2010-12>
- Stack, John** (2013). "Tate digital strategy 2013-15: Digital as a dimension of everything". *Tate.org*, 24 April.
<http://www.tate.org.uk/research/publications/tate-papers/tate-digital-strategy-2013-15-digital-dimension-everything>
- Tumasjan, Andranik; Sprenger, Timm O.; Sandner, Philipp G.; Welpe, Isabell M.** (2010). "Predicting elections with Twitter: What 140 characters reveal about political sentiment". En: *4th Intl AAAI conf on weblogs and social media*, pp. 178-185.
<http://cs.wellesley.edu/~cs315/Papers/predicting%20elections%20with%20twitter.pdf>
- Valcárcel-Medina, Isidoro** (2003). *Ir y venir de Valcárcel Medina*. Barcelona: Fundación Tàpies.
- Villaespesa, Elena** (2013). "El museo en la web social y su impacto en los visitantes". *Revista de museología*, v. 56, pp. 57-61.
- Viñarás-Abad, Mónica; Cabezuelo-Lorenzo, Francisco** (2012). "Claves para la participación y generación de contenido en las redes sociales: estudio de caso del Museo del Prado en Facebook". *adComunica*, v. 3, pp. 87-103.
<http://www.adcomunicarevista.com/ojs/index.php/adcomunica/article/view/49>
<http://dx.doi.org/10.6035/2174-0992.2012.3.6>



ARE LIBRARIES SUSTAINABLE IN A WORLD OF FREE, NETWORKED, DIGITAL INFORMATION?



Lluís Anglada

Nota: Este artículo puede leerse traducido al español en:
http://www.elprofesionaldelainformacion.com/contenidos/2014/nov/07_esp.pdf



Lluís Anglada is the director of the *Department of Libraries, Information and Documentation* of the *Consortium of University Services of Catalonia (CSUC)* after having been the director of the *Consortium of Academic Libraries of Catalonia (CBUC)* from 1997 to 2013. Previously he was director of the libraries of the *Polytechnic University of Catalonia* and professor at the *Jordi Rubió i Balaguer School of Library and Information Science* at the *University of Barcelona*. He holds a degree in philosophy and a diploma in library and information science. He was a member of the *OCLC Global Council* and currently is part of the *Liber Executive Committee*. He is the author and promoter of *Blok de Bid* and *Bdíg* blogs. He regularly delivers lectures and workshops and publishes articles on libraries and library consortia.

<http://orcid.org/0000-0002-6384-4927>

Consortium of University Services of Catalonia
Gran Capità, 2-4, Nexus bldg., 4th floor. 08034 Barcelona, Spain
langlada@gmail.com

Abstract

The evolution of libraries through three stages –modernization, automation and digitization- is analysed. A formula is presented to evaluate the importance of libraries to our society, and it is applied both retrospectively and futuristically, extrapolating a 2030 scenario. The conclusion is that if the current generation of librarians does not introduce radical changes in the role of libraries, their future is seriously threatened.

Keywords

Analysis, Sustainability, Values, Roles, Functions, Libraries, Extrapolation, Future developments, Prospects, Trends, Forecast.

Título: ¿Son las bibliotecas sostenibles en un mundo de información libre, digital y en red?

Resumen

Se analiza la evolución de las bibliotecas en sus tres etapas de modernización, automatización y digitalización. Se presenta una fórmula para evaluar el grado de interés que las bibliotecas tienen para la sociedad, y se aplica tanto retrospectivamente como para extrapolar el posible escenario del año 2030. Se concluye que si la generación actual de bibliotecarios no es capaz de introducir cambios radicales en el papel de las bibliotecas, su continuidad está seriamente amenazada.

Palabras clave

Análisis, Sostenibilidad, Valor, Rol, Funciones, Bibliotecas, Extrapolación, Futuro, Evolución, Perspectivas.

Anglada, Lluís (2014). "Are libraries sustainable in a world of free, networked, digital information?". *El profesional de la información*, v. 23, n. 6, November-December, pp. 603- 611.

<http://dx.doi.org/10.3145/epi.2014.nov.07>

1. Introduction¹

Sustainability is the capacity to endure over time. Libraries are among the institutions with the greatest longevity in our society, but the great technological changes that have affected them, especially since the advent of the internet and the digitization of information, have led many researchers to wonder about the future of libraries and whether such institutions continue to make sense. Recent studies include

Blumenthal (2014), **Kamat and Pujar** (2014), **Inouye** (2013), **Lewis** (2013), **Rasmus** (2013), **Cassi** (2012), **Coffman** (2012), **Dempsey** (2012b), **Laudano et al.** (2012), **Lynch, Greifeneder and Seadle** (2012), **Varela-Prado and Baiget** (2012), **Powell** (2010), **Dillon** (2008). The professional community began to worry when several studies (**Housewright et al.**, 2013) showed that users are now much more autonomous than before to seek information on their own, allowing them to dispense with using the library to meet their information needs.

Article received on 28-09-2014

Final acceptance: 15-10-2014

The sustainability of libraries can be evaluated according to the relationship established between the values added and their costs. Their physical persistence has been associated with physical documents, and particularly with printed books (Gauder, 2010), but our society is rapidly moving to an environment in which the information will be –or already is– primarily digital, probably more fragmented and flowing freely in the Network (Dempsey, 2012; Dillon, 2008). In an environment in which the object perceived as most associated with libraries –the printed book– tends to disappear, will the libraries that have preserved and made available those printed books to whoever wanted to use them also disappear?

The sustainability of libraries can be evaluated from the relationship established between the values added and their costs

This paper aims to present a reflective exercise on the factors that make libraries interesting and attractive to the society that funds and sustains them. The first version of these ideas was presented at the *Bobcatsss* conference (Anglada, 2014). In the following sections, a mathematical formula is proposed that reveals the sustainability of libraries in various historical periods based on the values assigned to its variables. The formula –which has explanatory and reflective, but not scientific, pretensions–, makes it possible to determine, depending on the historical stage, if libraries are sustainable, i.e. if their social value exceeds their cost. The final result is highly correlated with the perception of the library as an institution, and we can conclude that the sustainability of the library lies in the ability of librarians to show that the library provides values not limited to printed books or to physical documents.

2. Methodology: the sustainability formula

The proposed formula for the sustainability of libraries is the following:

$$\text{Sustainability} = \frac{\text{Value}}{\text{Cost}} = \frac{\text{Use} - \text{Dysfunctions} + (2 * \text{Perception})}{\text{Cost}}$$

Or, in an abbreviated form: $S = (U - D + 2P) / C$

Of course, this formula is not intended to be an expression of an exact law such as those of physics; it is rather a pretext, an intellectual exercise or an academic amusement (whatever you want to call it), which helps us to make a serious analysis of the role of libraries in society. While we can easily understand the use (U) of libraries, we must specify what we mean by dysfunctions (D): we can define them as unfulfilled expectations, a contrast between expectations and reality (E - R). In the proposed exercise, variables can be assigned values between 1 and 5, in the most objective way possible considering each stage or historical moment. Inevitably, the assignment of values to variables always has

a significant degree of subjectivity, but here we have based these values on facts and actual situations, duly reported in the professional literature.

Another methodological note before continuing. A division of the recent history of libraries into three major phases over the past 50 years is proposed:

- Mechanization, a period characterized by the construction of new buildings and mechanization of processes;
- Automation, in which networks of libraries were created, OPAC was introduced and union catalogues were created; and
- Digitization, in which electronic journals and books appeared and documents stored in libraries have been digitized (Buckland, 1992)².

We set off from the hypothesis that the same stages and library events have taken place in all countries and in all library types, but not at the same time. Therefore, the different phases that libraries passed through to become automated are similar in each country, but they occurred in different years (Anglada, 2006).

3. Recent values assigned to the library sustainability formula

3.1. Mechanization stage

The action program for the phase that we have named Mechanization was relatively simple: to contain many books and modern furnishings in a renovated or new building and to use “mechanical” rather than “manual” technology (although in truth the technologies were electrical and electronic). The library incorporated new media (music records and cassettes, audio books, audiovisual materials for teaching, etc.), microformats were used to increase collections with reproductions of documents previously not held, and technological innovations helped to modernize services (photocopiers for document delivery, telex and later fax to send interlibrary loan requests).

The Mechanization phase of library modernization that developed in Europe during the '60s –but did not reach Spain until the end of the '70s– assumed generally that libraries were the only place where everyone could find information (and besides for free). Public perception was highly positive and the operating cost was relatively low. In these circumstances, we estimate the values of the variables in the formula as follows: fairly high usage (U = 4), no dysfunctions because no one expected anything specific from libraries (D = 0), very good user perception (P = 4) and relatively low cost to funders (C = 3). The result of applying the sustainability formula (S = 4) argues that the sustainability of the modern (1970) mechanized library was high:

$$S(1970) = (U - D + 2P) / C = (4 - 0 + 2*4) / 3 = 12/3 = 4$$

3.2. Automation stage

This stage, which could also be called the computerization stage, is defined by the use of computers in libraries, which occurred experimentally in the early '70s and was definitively introduced in the '80s (Anglada, 2006). Libraries created networks (like OCLC) or cooperatives (PICA in Holland,

Scolcap in Scotland...) or national centres (*Bibsys* in Norway) to share the high costs of computers at the time. National or regional union catalogues were created, interlibrary loan agreements were put in place, and eventually commercial programs for integrated management (integrated library systems, ILS) became available. Thus libraries began to offer online access to databases (**Baiget**, 1998), online public access catalogues (OPACs) were launched, and CD-ROMs were introduced.

In this stage, librarians were the protagonists –30 or 40 years ago they were young people applying “new technologies” to library services at all times: electric typewriters to duplicate catalogue cards or CD-ROMs to provide access to bibliographic information.

If, as it has been said, the main steps of library automation occurred in the 70s and 80s in the developed world, in a Spain that was coming out of the Franco era the change occurred about 10 years later.

In the '90s, libraries were still the only place to find free information, and although their operating costs had increased, they were still not overly expensive institutions. Today we can say that the library automation programs –a dream that once seemed far-off in the future– are virtually ubiquitous in university libraries, and most other libraries as well (**Hernández-Sánchez; Arroyo-Vázquez**, 2012).

Let's assign values to the formula for this period: fairly high usage ($U = 4$), some dysfunctions due to technological limitations ($D = 1$) as, for example, online searches of remote databases incurred costs associated with connection time, as well as expensive and complicated telephone connections (**Baiget**, 1998). User perception was still good ($P = 4$) as they experienced improvements, but the cost was considerably higher due to the essential computer technology ($C = 4$). The result of applying the formula to the automated library of this stage still yields an acceptably high value ($S = 2.75$), although lower than that of the preceding stage.

$$S(1990) = (U - D + 2P) / C = (4 - 1 + 2 \cdot 4) = 11/4 = 2.75$$

3.3. Digitization stage

The third stage began its development between the late 20th and early 21st century. The digitization stage, like all others, is characterized on the one hand by a disruptive innovation (Internet) and, on the other, by the developments of the preceding stage. It is redundant and obvious to point out that the Internet has caused and still is causing major changes in the way society is organized and people interact, but perhaps it is less evident that these changes are so recent that we cannot yet know how they will be consolidated.



Figure 1. Less space for books and more for people (*Amsterdam Public Library*)

As in the past, librarians have been proactive in the face of change and have been able to create new uses for libraries. At the same time, the speed of change has been so great that dysfunctions (previously defined as the difference between expectations and realities) have continued to increase. Let's look at this in a little more detail.

“ Libraries are changing from being a space to store, locate and use books to places where people interact and socialize ”

We can start with the positive aspects, the legacy that our generation has been able to build. Starting with modernized/mechanized and computerized/automated libraries, librarians have explored new uses for library buildings and new roles for the profession. It is never easy to recreate the functionality of a space that has well-established applications, and even less easy to do so when these spaces are newly created. Perhaps the most intense changes have occurred in the public libraries, which were created as places for quiet, lonely, individual work and study, and today are offered as meeting places for socializing and sharing and active learning. Libraries are changing from being a space to store, locate and use books to places where people interact and socialize. This should transform the perception that citizens have of their libraries, seeing them as places to ‘change lives by giving people the tools they need to succeed’³.

Recent professional literature has shown that the contributions of the public library are not limited to lending books. In economic crises, libraries emphasize their role in helping people (**Gauder**, 2010), fighting the digital divide by offering culture, entertainment and Internet connectivity for

all (Vinjamuri, 2014; Fleck, 2013, Houghton *et al.*, 2014.; Goedekena, 2014), or contributing to the creation of social cohesion and impartial public spaces (Jochumsen *et al.*, 2012; Vårheim *et al.*, 2008).

The fact that libraries carry out functions of assistance and intermediation that do not produce final results impose on them a permanent invisibility; this is very negative because they depend on public funding, and their future depends on the perception or mental image of libraries held by administrators and policy makers who allocate budgets (Varela-Pardo; Baiget, 2012). Therefore, the efforts that libraries made in the past to show statistical data on resources (“holdings”) and usage (“circulation data”) must now show the value that libraries have for the entities that support them financially (Oakleaf, 2012; Gómez-Yáñez, 2014, Paños-Paños, 2011) and for the citizens who use them (Vakkari; Serola, 2012; Huysmans; Hillebrink, 2008).

“ The fact that libraries carry out functions of assistance and intermediation that do not produce final results impose on them a permanent invisibility ”

University libraries, meanwhile, have seen that their modernized buildings and computerized collections are not sufficient to continue to attract users and to be useful to their institutions. This has led, on the one hand, to the physical transformation of the space but also to changes in the ways they support research. Traditionally this was done by building collections, but changes in scientific communication (both in methods and in practice) have required libraries to reorganize services and more actively support researchers. Librarians have created new fields of professional activity by promoting Open Access, maintaining institutional repositories, helping researchers to make their work more visible on the Internet, facilitating the adoption of identifiers for authors and for documents, etc. Other areas in which the profession has created new activities include information literacy, digitization of resources and the creation of digital libraries (such as *Memòria Digital de Catalunya*) and portals (such as *Europeana* or *DART-Europe*) using a federated distribution of tasks and the adoption of standards...

We have tried to show that the problem of libraries is not immobility, as they have resilience and ability to adapt to new environments. At a time of profound change they have known how to reorient and to propose new options that create new uses. But the emergence of new roles for libraries does not mean that all library services have evolved over time.

Previously we defined dysfunctions as unfulfilled expectations, as the difference between the expected and the reality.⁴ We have accumulated too many dysfunctions. Some are economic and make libraries less competitive (Coffman, 2012; Dempsey, 2012, Lewis, 2007); others are related to service design and make them less useful. I will not dwell too much on these dysfunctions but will mention three examples:

- How many times is the same book catalogued? More than once is too many. Librarians spent decades developing common cataloguing standards and practices, and standards for data transmission, but this has not prevented the same document from being catalogued by several people in different ways. While the former “only” has economic impact, the latter generates duplications in catalogues, resulting in system recovery failures.
- Library catalogues and automated systems were innovative in the '80s, but have been stuck in outmoded practices. Users have adapted quickly to the ‘googlization’ of information and do not understand why they should have to look in different places to get a unique solution to an information need.
- Finally, despite the many social changes in recent decades, libraries have not sufficiently adapted their services to new expectations. Library hours should better respond to citizens’ nonwork and festive days, as well as adapting to this new form of communication that is social networking.

Given these considerations, we can assign values to the present moment. The usage is still quite high ($U = 4$) even though the library is no longer seen as the only place to find (free) information. As we have seen, dysfunctions increased ($D = 2$) and also increased the technology promises that cannot be met in the library at an everyday level. Perception is also good, in part because of past image ($P = 3$). Costs have increased but libraries have been able to increase productivity and savings thanks to cooperation and library consortia; we are doing more with less money, but what we do costs more ($C = 4$). The result of applying the formula to the Digitized library ($S = 2$) shows a clear downward trend:

$$S(2010) = (U - D + 2P) / C = (4 - 2 + 2 \cdot 3) / 4 = 8/4 = 2$$

4. Two key elements for future library sustainability: perception and adaptation to a new paradigm

Two of the elements involved in the formula –costs and use– are fairly intuitive, but we have to justify the introduction of the element of ‘perception’ and the importance we assigned to it. Daniel Kahneman received the 2002 Nobel Prize in economics for showing, along with Amos Tversky, that many of the decisions people take are not based on probability calculation or reason but on cognitive biases or prejudices. These biases or stereotypes construct perceptions that are not always grounded in reality. In the words of Kahneman (2011, ch. 16, p. 223):

“When the categories are social, these representations are called stereotypes. Some stereotypes are perniciously wrong, and hostile stereotyping can have dreadful consequences. But the psychological facts cannot be avoided: stereotypes, both correct and false, are how we think of categories.”

The contributions of cognitive psychology indicate that people make decisions motivated by mental frames (Lakoff, 2004) or perceptions, and not only by the facts or reality. This is also apparent in an OCLC study (De-Rosa; Johnson, 2008; Anglada, 2008) showing that the predisposition to

vote for taxes to create libraries has little to do with their actual use:

“Significantly and surprisingly, library funding support is not driven by demographics, i.e., income, age, gender, race, political affiliation, etc.

[...] Library support is more about a mindset or an attitude than a traditional demographic profile”.

However, the general public mental image (perception) of libraries was forged over decades and, although they have pioneered ample use of new technologies, the image of libraries remains attached to the printed book. In 2005, 69% of Americans had this perception (**De-Rosa, Cantrell, OCLC, 2005**), and it had increased to 75% in 2010 (**Gauder, 2010, p. 39**).

In the past, this perception of the libraries gave them a high social value because the printed book was the only means of disseminating information and recording culture, and the library –a books warehouse- received the same value associated to information and culture.⁵ The development of the Internet has established a dissociation between the information itself and its supporting document. Information flows in the Network independent of its support, and its access cost (in user time and money) does nothing but decrease.

In the print world (which we are leaving) the public perception equating libraries with books was positive for libraries because books, or rather access to them, were limited (and the user’s time to access the information was abundant) (**Dempsey, 2012**). The change in the digital world is not in the carrier, but in the huge, nearly free flows of information in the Network. The information is displayed to the public as plentiful, and abundant things tend to be considered less valuable, less important. Libraries may end up being seen as useful only to preserve the past, ie the printed book, and consequently of little use to handle digital information. In the new environment, some traditional strengths of libraries are weakening.

The library has been steadily declining in importance in university budgets. According to the annual reports of 40 members of the *American Research Libraries (ARL)*, the percentage of university spending in libraries decreased from

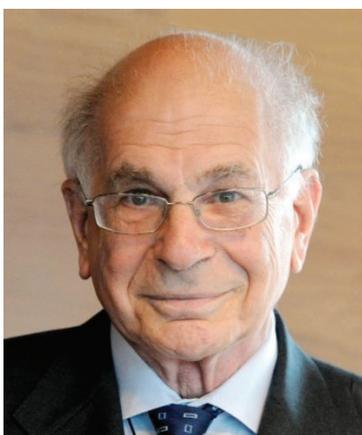


Figure 2. Daniel Kahneman: *People’s decisions are motivated by perceptions and not only by the reality*

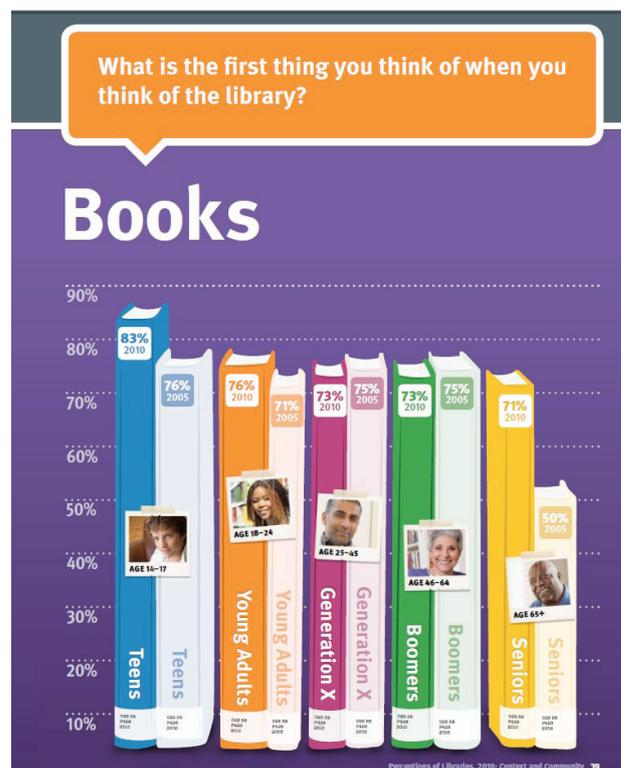


Figure 3. OCLC. *Perceptions of libraries: context and community*, p. 39. http://www.oclc.org/reports/2010perceptions/2010perceptions_all_singlepage.pdf

3.7% in 1984 to slightly less than 2% in 2009 (**Davis, 2012**). Surely, many libraries never had such a large university budget share as the *ARL* members, but surely they also are experiencing decreased spending as a trend.

“Internet has consolidated the dissociation between the information itself and its supporting document”

At the same time, some classic library services are experiencing significant declines. Annual statistics from the *Association of College and Research Libraries (ACRL)* show a decrease of more than 30% in loan transactions between 1998 and 2008. In the same period, the decline was significantly greater in reference inquiries made to librarians, a service traditionally associated with the university library (*University Leadership Council, 2011*).

Several studies have indicated a displacement of the starting point for literature searches. While the physical library or its catalogue were the classical mode, between 2003 and 2012 both lost importance in comparison with beginning searches in Internet search engines or specialized databases (**Housewright, 2013**). A recent survey of teachers and researchers from the universities of Catalonia showed that half of the respondents went first to a specialized database to start a literature search; the second option was Internet search engines (30%), followed by the library catalogue (17%) and the option of physically visiting the library, a residual 3% (**Borrego, 2014**).

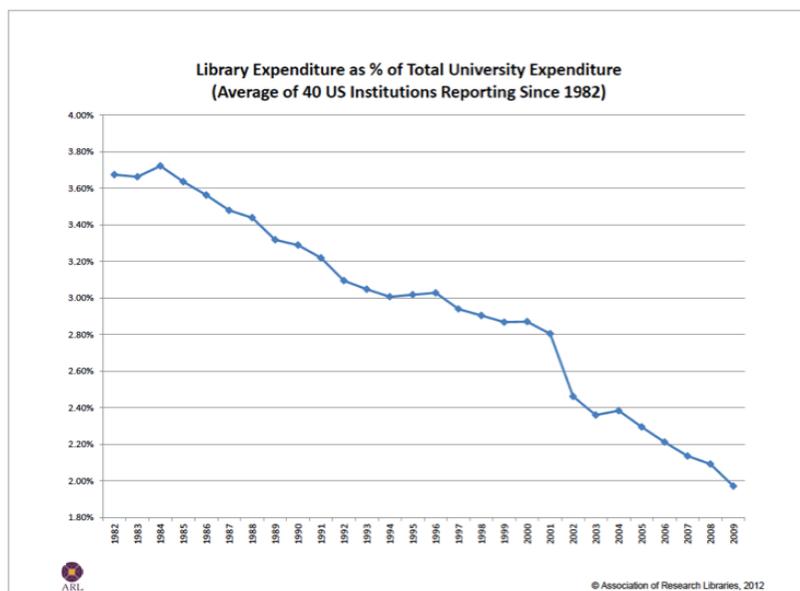


Figure 4. Library expenditure as % of total university expenditure (average of 40 US institutions reporting since 1982). Association of Research Libraries.

We can summarize the current situation as a complex framework in which libraries have been able to play new roles, although they could not prevent the existence of dysfunctions. The key element, however, is the perception that libraries are associated with the print world, giving rise to a mental image that assigns less value to the library and its services.

“ Libraries have been able to play new roles, although they could not prevent the existence of dysfunctions ”

5. Are libraries sustainable in a world of free, networked, digital information?

Let's imagine that 15 years have elapsed and we are making a trip to a library in 2030. We ask one of the librarians to calculate the S value for 2020.

That librarian of the future will probably tell us that in 2020 a large part of the professional and scientific journals were already available for free in digital format, there was a lot of quality information on the Internet, services provided through the Net were better than those provided by libraries, and libraries were becoming marginal resources with very high costs. The values of the variables could be: $U = 3$, $D = 2$, $P = 2$, $C = 4$, with a result of $S = 1.25$ –namely, low sustainability.

$$S(2020) = (U - D + 2P) / C = (3 - 2 + 2 \cdot 2) = 5/4 = 1.25$$

What if we asked our librarian to tell us about the situation in 2030? Perhaps we would hear that open access is a complete reality in the scientific world, that many services provide access to scientific, literary, promotional and leisure works for a clearly acceptable flat fee, and that *Google's* successor finds needed documents with almost no margin of

error... In this situation, the library stands as a resource with relatively high marginal costs. Let's assign values: uses declining ($U = 2$), a similar degree of dysfunction ($D = 2$), low perception of usefulness ($P = 1$), and costs remain stable ($C = 4$). The value of S –less than one– indicates that libraries are unsustainable.

$$S(2030) = (U - D + 2P) / C = (2 - 2 + 2 \cdot 1) = 2/4 = 0.5$$

Examining our recent past more closely, we can see that libraries have had some ability to reinvent themselves (participating in open access or information literacy, for example), attracting new uses. Costs have been contained by creating networks and consortia, and making joint purchases. They have had less influence in changing how society perceives them. It is difficult to accept the paradox that, while the public is fully immersed in the Information Society, the information professional is ignored. The name of the profession –'librarian'– inevitably confines it to what its name suggests: the four walls of the library⁶.

The generation that started the modernization and automation stages of libraries made (ie, we made) the mistake of trying to raise our profile by emphasizing objects, the containers of information and data. We thought we were indispensable and dreamed of new names for a profession empowered by a time of change: information scientist, information architect, knowledge manager... We forgot that our true role is modest, that we are little more than 'helpers'⁷ in *Infoland*, a world of abundant and free-flowing information in which our role is not major, but secondary, although still important.

Rephrasing the title question: Who sustains libraries and why? Libraries are sustained by people through institutions and society in general because they believe, feel, intuit or think that libraries are important to them, because they have a positive perception of them. With the aforementioned studies seeking to show the added value of libraries, we can influence the rational thinking of citizens but changing their perception of libraries requires that we address their emotions.

But, are libraries really important to people? And how about the professionals who work in and out of them? Yes, we represent a unique point of view regarding information. We are not interested in the results (to get information), but in the process (to help people to find and use it). We do not want just any information (the most used) or only some users (those who can best use it): a library (the organization, not the building) seeks to collect and preserve all information for all and forever (**Gorman, 2007**). The library is closer to the people than to the document, because it seeks to put information at the service of people by providing tools and skills to be citizens (not lackeys) in a world where access to and use of information can be key elements of success or social exclusion.

We believe that society (still) needs the functions performed by libraries (and librarians), but does this make them immediately sustainable? The answer is no. And it will remain so unless we can soon establish a new stereotype of 'library' in people's minds, one that is not based on the physicality of the buildings or books, but focuses on the role of support and assistance in the difficult process of using information and transforming it into knowledge. The creation of perceptions of a library and librarian that are associated with assistance regarding information is a contribution that has not yet been made.

The creation of this new perception certainly cannot be done by the players who participated in the modernization and automation of the libraries of the late 20th century. It remains to be performed by the current generation of young librarians, those who are inheriting renovated libraries but also a mental image that is associated with becoming less powerful for society. This is the challenge and responsibility for young librarians: to create a new perception of our profession.

“ We must establish a new stereotype of 'library' in peoples' minds, one that is not based on the physicality of the buildings and books, but on the role of support and assistance in the difficult process of using information and transforming it into knowledge ”

Acknowledgements

This article would have not gone to press without the temper, encouragement and aid of **Tomàs Baiget** who, besides that spiritual companionship, generated the conference transcript. I also want to thank **Cristóbal Urbano** for reading a previous version of the article, and his on-target comments.

Notes

1. This article is a transcript and expansion of the inaugural lecture delivered by the author at the 22nd *Bobcatsss* conference in Barcelona, 29-31 January 2014, which had the motto of *Library (r)evolution: Promoting sustainable information practices*. *Bobcatsss* is organized by Library and Information Science students, and this paper was addressed to them.
2. Buckland terminology is 'paper library', 'automated library', and 'electronic library'.
3. 'Libraries: a place that changes lives by giving people the tools they need to succeed', John Szabo, director of Los Angeles Public Library, in: *Libraries as a resource in the digital age*, post *Citizens of culture*, Sept. 17th 2014. <http://citizensofculture.com/2014/09/17/libraries-as-a-resource-in-the-digital-age>
4. Perhaps this definition of "dysfunction" is insufficient.

Cristóbal Urbano proposed the term "friction differential", which would be the difference in the costs of obtaining a piece of information in a library and getting the same information from other agents in the document chain. Thus, D (friction differential) would increase over time in relative terms, not absolute terms: libraries have improved accessibility, availability, etc., but at the same time other agents in the information chain (both commercial and social and collaborative on the Internet) have offered shortcuts and greater flexibility.

5. See, for example, "The one place where books have been free is libraries, whose existence is instructive in another way: We've had libraries for centuries and fund them with public dollars because we view books not just as entertainments, but as repositories of culture and knowledge." Hughes, Evan. "Books don't want to be free: How publishing escaped the cruel fate of other culture industries". *New republic*, October 8th, 2013.

http://www.newrepublic.com/article/115010/publishing-industry-thriving?utm_content=bufferad667&utm_source=buffer&utm_medium=twitter&utm_campaign=Buffer

6. "No one can understand that a 'librarian' can act outside the library, and while this is repeated again and again, stubbornly librarians do not want to leave that label that will marginalize them for life, preventing them, for example, from working in companies that have no library." Conversation with Tomàs Baiget, 2014.

7. I do not know what other word to use to refer to those professionals (like the librarian, or whatever one chooses to call them) whose function is to help others to achieve a given result. The library should be the organization that allows the citizen to be self-sufficient in information use. A seemingly paradoxical role of the librarian is to become unnecessary, to empower people to be autonomous in their search for the information they need.

Bibliographic references

- Anglada, Lluís** (2006). "Veinticinco años de automatización de bibliotecas en Cataluña". *BiD: textos universitaris de biblioteconomia i documentació*, junio, n. 16. <http://bid.ub.edu/16angla2.htm>
- Anglada, Lluís** (2008). "Aprendiendo de Lakoff ¿Quién apoya a las bibliotecas?". *BDig*, 28 septiembre. <http://bdig.blogspot.com.es/2008/09/aprendiendo-de-lakoff-quin-apoya-las.html>
- Anglada, Lluís** (2012). "Críticas excelentes, alternativas insuficientes". *BDig*, lunes, 26 noviembre de 2012 <http://bdig.blogspot.com.es/2012/11/criticas-excelentes-alternativas.html>
- Anglada, Lluís** (2014). "Are libraries sustainable in a world of free, digital and networked information?" [not published]. In: 22nd *Bobcatsss* conference. *Library (r)evolution: Promoting sustainable information practices*. Barcelona, 29th-31st January 2014. <http://bobcatsss2014.hb.se>
- Baiget, Tomàs** (1998). "25 años de teledocumentación en España". *Revista española de documentación científica*, v. 21, n. 4, pp. 373-387.

<http://dx.doi.org/10.3989/redc.1998.v21.i4.359>

Blumenthal, Jane (2014). "Creating the future". *Journal of the Medical Library Association*, v. 102, n. 1, pp. 2-4.
<http://dx.doi.org/10.3163/1536-5050.102.1.002>

Borrego, Ángel (2014). "Comportament informatiu del professorat de les universitats catalanes: estudi realitzat per encàrrec del Consorci de Serveis Universitaris de Catalunya (CSUC)". Barcelona: CSUC: Àrea de Biblioteques, Informació i Documentació (not published).

Buckland, Michael K. (1992). *Redesigning library services*. Chicago: American Library Association.
<http://sunsite.berkeley.edu/Literature/Library/Redesigning/html.html>

Cassi, Marco (2012). "Il futuro della biblioteca accademica di ricerca: alcuni temi chiave". In: *Oltre la biblioteca e i bibliotecari: professioni necessarie per la biblioteca accademica di ricerca*, Roma (Italia), 5 nov.
<http://eprints.rclis.org/17950>

Coffman, Steve (2012). "The decline and fall of the library empire". *Searcher*, v. 20, n. 3.
<http://www.infotoday.com/searcher/apr12/Coffman--The-Divide-and-Fall-of-the-Library-Empire.shtml>

Davis, Phil (2012). "Libraries receiving a shrinking piece of the university pie". *The scholarly kitchen*, Feb 15, 2012,
<http://scholarlykitchen.sspnet.org/2012/02/15/a-shrinking-piece-of-the-university-pie>

Dempsey, Lorcan (2012). "Thirteen ways of looking at libraries, discovery, and the catalog: scale, workflow, attention". *Educause review online*, Monday, Dec. 10, 2012.
<http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention>

Dempsey, Lorcan (2012b). "Libraries and the informational future: some notes". *Information professionals 2050: Educational possibilities and pathways*, Eds. Gary Marchionini and Barbara Moran (Chapel Hill: School of Information and Library Science, University of North Carolina at Chapel Hill), pp. 113-125. ISBN: 978 1 300 20486 2
<http://sils.unc.edu/sites/default/files/publications/Information-Professionals-2050.pdf>

De-Rosa, Cathy; Johnson, Jenny (2008). *From awareness to funding: A study of library support in America*. A report to the OCLC membership, 212 pp.
<http://www.oclc.org/reports/funding/fullreport.pdf>

De-Rosa, Cathy; Cantrell, Joanne et al. (2005). *Perceptions of libraries and information resources*. A report to the OCLC membership, 290 pp.
<http://oclc.org/reports/2005perceptions.en.html>

Dillon, Dennis (2008). "A world infinite and accessible: Digital ubiquity, the adaptable library, and the end of information". *Journal of library administration*, v. 48, n. 1, pp. 69-83.
<http://dx.doi.org/10.1080/01930820802035034>

Fleck, Alissa (2013). "Libraries bridge digital divide". *News our town downtown*, July 3rd.
<http://nypress.com/libraries-bridge-digital-divide>

Gargiulo, Paola (2012). "L'electronic resources librarian tra competenze vecchie, nuove e future". In: *Oltre la biblioteca e i bibliotecari: professioni necessarie per la biblioteca accademica di ricerca*, Roma, Italia, 5 nov.
<http://eprints.rclis.org/17947>

Gauder, Brad (editor) (2010). *Perceptions of libraries: context and community. A report to the OCLC membership*. Dublin, Ohio: OCLC. ISBN: 978 1 55653 395 2
http://www.oclc.org/reports/2010perceptions/2010perceptions_all_singlepage.pdf

Goedekena, Ed (2014). "Digital libraries and information access: Research perspectives". *Technical services quarterly*, v. 31, n. 2, pp. 210-212.
<http://dx.doi.org/10.1080/07317131.2014.875403>

Gómez-Yáñez, José-Antonio (coord.) (2014). *El valor económico y social de los servicios de información: bibliotecas: Informe de resultados*. Fesabid.
<http://www.fesabid.org/system/files/repositorio/fesabid-valor-economico-social-servicios-informacion-bibliotecas.pdf>

Gorman, Michael (2007). "The wrong path and the right path: The role of libraries in access to, and preservation of, cultural heritage". *New library world*, v. 108, n. 11/12, pp. 479-489.
<http://dx.doi.org/10.1108/03074800710838236>

Hernández-Sánchez, Hilario; Arroyo-Vázquez, Natalia (2014). "Efectos de la crisis económica en las bibliotecas españolas". *El profesional de la información*, v. 23, n. 2, marzo-abril, pp. 158-164.
<http://dx.doi.org/10.3145/epi.2014.mar.08>

Hernández-Sánchez, Hilario; Arroyo-Vázquez, Natalia (2012). "Las bibliotecas en España en 2011". *Anuario ThinkEPI*, v. 6, pp. 48-54.
<http://www.thinkepi.net/anuario-thinkepi/anuario-thinkepi-2012>

Houghton, Kirralie; Miller, Evonne; Foth, Marcus (2014). "The local library bridging the digital and physical city: opportunities for economic development". *Commonwealth journal of local governance* (in press).
<http://eprints.qut.edu.au/66194>

Housewright, Ross; Schonfeld, Roger C.; Wulfson, Kate (2013). *Ithaka S+R US faculty survey 2012*. New York: Ithaka S+R, 2013.
<http://www.sr.ithaka.org/research-publications/us-faculty-survey-2012>

Huysmans, Frank; Hillebrink, Carlien (2008). "The future of the Dutch public library: ten years on". The Hague: Netherlands Institute for Social Research.
http://www.scp.nl/english/Publications/Publications_by_year/Publications_2008/The_future_of_the_Dutch_public_library_ten_years_on

Inouye, Alan S. (2013). "The future of libraries at thirty thousand feet: strategy and public policy". *Young adult library services*. Fall, v. 12, n. 1, pp. 9-12.
<http://www.questia.com/library/journal/1G1-349902998/the-future-of-libraries-at-thirty-thousand-feet-strategy>

- Jochumsen, Henrik; Rasmussen, Casper-Hvenegaard; Skot-Hansen, Dorte** (2012). "The four spaces - a new model for the public library". *New library world*, v. 113, n. 11/12, pp. 586-597.
<http://dx.doi.org/10.1108/03074801211282948>
- Kahneman, Daniel** (2011). *Thinking, fast and slow*. Farrar, Strauss & Giroux. ISBN: 978 0 374 27563 1
- Kamat, Rajanish K.; Pujar, Shamprasad M.** (2014). "Inside out and outside in: future legend of library and information science". In: *Recent trends in library and information science: a forward for the future*. LIS Academy International, pp. 206-216.
<http://eprints.rclis.org/22730>
- Kepple, Sarah** (2013). "Intentionally backwards, the future of learning in libraries". *Young adult library services*. Fall, v. 12, n. 1, pp. 33-37.
<http://www.questia.com/library/journal/1G1-349903004/intentionally-backwards-the-future-of-learning-in>
- Lakoff, George** (2004). "Don't think of an elephant!". White River Junction, Vermont, USA: Chelsea Green Publishing, 174 pp. ISBN: 1 931498 71 7
- Laskowski, Mary S.; Maddox Abbott, Jennifer A.** (2014). "The evolution of technical services: learning from the past and embracing the future". *Technical services quarterly*, v. 31, n. 1, pp. 13-30.
<http://dx.doi.org/10.1080/07317131.2014.844619>
- Laudano, Claudia-Nora; Planas, Javier; Corda, María-Cecilia** (2012). "Un futuro cargado de tecnología: aproximaciones a los cambios imaginables en el campo de la bibliotecología en una década". *Palabra clave*, v. 2, n. 1.
<http://eprints.rclis.org/17924>
- Lewis, David W.** (2007). "A strategy for academic libraries in the first quarter of the 21st century". *College & research libraries*, v. 68, n. 5, pp. 418-434.
<http://crl.acrl.org/content/68/5/418.full.pdf>
- Lewis, David W.** (2013). "From stacks to the web: the transformation of academic library collecting". *College & research libraries*, v. 74, n. 2, pp. 159-176.
<http://hdl.handle.net/1805/3252>
- Lynch, Clifford; Greifeneder, Elke; Seadle, Michael S.** (2012). "Interactions between libraries and technology over the past thirty years: an interview with Clifford Lynch, 23 June 2012". *Library hi tech*, v. 30, n. 4.
- Oakleaf, Megan** (2010). *The value of academic libraries: A comprehensive research review and report*. Chicago: Association of College and Research Libraries.
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/value/val_report.pdf
- O'Connor, Steve** (2014). "Leadership for future libraries". *Library management*. Feb., v. 35, n. 1/2, pp. 78-87.
<http://dx.doi.org/10.1108/LM-06-2013-0047>
- Paños-Paños, Antonio** (2011). "Análisis de los beneficios económicos y el perfil de innovación de las bibliotecas públicas". *Boletín de la Asociación Andaluza de Bibliotecarios*, n. 102, pp. 65-77.
<http://dialnet.unirioja.es/servlet/articulo?codigo=4004296>
- Pattinson, Damian; MacCallum, Catriona J.** (2014). "The future is open: opportunities for publishers and institutions". *Insights: the UKSG journal*, v. 27, n. 1, Mar, pp. 38-44.
<http://dx.doi.org/10.1629/2048-7754.139>
- Powell, Allen** (2010). "Times of crisis accelerate inevitable change". *Journal of library administration*, v. 51, n. 1, pp. 105-129.
<http://dx.doi.org/10.1080/01930826.2011.531644>
- Rasmus, Daniel W.** (2013). "Uncertainty and the future of libraries". *Computers in libraries*. Dec., v. 33, n. 10, pp. 4-32.
<http://pqasb.pqarchiver.com/infotoday/doc/1468554366.html>
- University Leadership Council (2011). *Redefining the academic library*. Washington, DC: Education Advisory Board, November.
- Vakkari, Pertti; Serola, Sami** (2012). "Perceived outcomes of public libraries". *Library and information science research*, v. 34, n. 1, pp. 37-44.
<http://dx.doi.org/10.1016/j.lisr.2011.07.005>
- Varela-Prado, Carmen; Baiget, Tomàs** (2012). "El futuro de las bibliotecas académicas: incertidumbres, oportunidades y retos". *Investigación bibliotecológica*, v. 26, n. 56, enero/abril, pp. 115-135.
<http://www.revistas.unam.mx/index.php/ibi/article/view/33175>
- Vårheim, Andreas; Steinmo, Sven; Eisaku, Ide** (2008). "Do libraries matter? Public libraries and the creation of social capital". *Journal of documentation*, v. 64, n. 6, pp. 877-892.
<http://dx.doi.org/10.1108/00220410810912433>
- Vinjamuri, David** (2014). "How public libraries are solving America's reading problem". *Forbes*, 11 March.
<http://www.forbes.com/sites/davidvinjamuri/2014/03/11/how-public-libraries-are-solving-americas-reading-problem>
- Walton, Graham** (2014). "Will smart phones and other bring your own devices (byod) dominate how academic library services are developed and delivered for the foreseeable future? *New review of academic librarianship*, Jan., v. 20, n. 1, pp. 1-3.
<http://dx.doi.org/10.1080/13614533.2014.875294>
- Young Adult Library Services Association (Yalsa) (2014). *The future of library services for and with teens: a call to action*. Institute of Museum and Library Services, Jan., 59 pp.
http://www.ala.org/yaforum/sites/ala.org.yaforum/files/content/YALSA_nationalforum_final.pdf

CRITERIOS DE EVALUACIÓN Y APLICACIÓN DE FUENTES DE INFORMACIÓN WEB EN CENTROS DE DOCUMENTACIÓN PERIODÍSTICA

Paz Orero y Bernardino J. Cebrián-Enrique



Paz Orero, documentalista en *Radiotelevisión Valenciana*, es licenciada en periodismo por la *Universidad CEU Cardenal Herrera* y en documentación por la *Universidad Politécnica de Valencia*. Es colaboradora docente de la asignatura de gestión de contenidos en los *Estudios de Ciencias de la Información y de la Comunicación* de la *Universitat Oberta de Catalunya* y profesora en el título de *Experto en creación de contenidos audiovisuales online* de la *Universidad CEU Cardenal Herrera*.
<http://orcid.org/0000-0001-8520-5512>

Radiotelevisión Valenciana
Polígono Acceso Ademuz, s/n. 46100 Burjassot (Valencia), España
mpaz@rtvv.es



Bernardino J. Cebrián-Enrique es doctor en comunicación pública (*Universidad de Navarra*, 1994) y profesor de documentación y técnicas de investigación social en la *Universidad CEU Cardenal Herrera*. Su investigación académica se centra en el valor periodístico de las fuentes documentales, el uso de internet como fuente, la verificación y otros elementos del periodismo de calidad. Forma parte del grupo de investigación impulsor del *Observatorio para la Investigación en Medios Digitales (Oimed)*.

<http://www.oimed.com>
<http://orcid.org/0000-0001-9722-5481>

Universidad CEU Cardenal Herrera
Facultad de Humanidades y Ciencias de la Comunicación
Luis Vives, 1. 46115 Alfara del Patriarca (Valencia), España
dcebrian@uch.ceu.es

Resumen

Se proponen los criterios documentales que, a juicio de investigadores y profesionales, resultan esenciales para la evaluación de fuentes de información en internet y la aplicación de las mismas en los centros de documentación periodística. Se muestran los resultados de una investigación realizada en 2013 mediante un estudio Delphi con ocho expertos en documentación informativa y una encuesta a 81 documentalistas de medios de comunicación generalistas españoles. Se analiza el impacto de los recursos digitales en estos servicios, la frecuencia de uso, el tipo de recurso consultado y las consecuencias en las rutinas laborales de los profesionales. El 86,4% de los encuestados utiliza internet cada día para buscar información. Los centros de documentación periodísticos españoles de medios generalistas y con difusión nacional no han establecido una política para la selección de fuentes de información en la Red.

Palabras clave

Fuentes de información, Evaluación de fuentes digitales, Criterios de selección, Documentación periodística, Documentalistas de medios, Centros de documentación, Medios de comunicación, Internet, Periodismo, Estudio Delphi, Encuesta.

Title: Criteria for evaluation and implementation of web information sources in news media libraries

Abstract

The documentary criteria that researchers and professionals think are essential for assessing Internet information sources and for implementing these sources in news libraries are presented. Results of a Delphi study conducted in 2013 with eight experts in news librarianship and a survey of 81 news librarians working in Spanish media are discussed. Aspects such as the impact of digital sources on news libraries, their frequency of use, the type of resource searched, and the consequences for professional routines in these centres are analyzed. 86.4% of respondents use the Internet every day to search for information. Documentation centres of the Spanish mainstream media have not established a policy for the selection of information sources on the Web.

Keywords

Information sources, Evaluation of digital sources, Criteria for selection, News librarianship, News librarians, News libraries, Mass media, Internet, Journalism, Delphi study, Survey.

Artículo recibido el 17-06-2014
Aceptación definitiva: 05-09-2014

Orero, Paz; Cebrián-Enrique, Bernardino J. (2014). "Criterios de evaluación y aplicación de fuentes de información web en centros de documentación periodística". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 612-617.

<http://dx.doi.org/10.3145/epi.2014.nov.08>

1. Introducción

El volumen ingente de información presente en internet debido a la facilidad de edición e inmediatez en la publicación, junto con la falta de garantías sobre la autoría y calidad de los contenidos junto con los problemas de localización, exigen la aplicación de criterios de evaluación y selección de fuentes.

La evaluación de la información no depende de los soportes en los que está contenida, aunque las características de algunos de ellos, como los digitales, influyen en su valoración. "Criterios tradicionales como la exactitud, la posibilidad de verificación o la vigencia del contenido son aplicables a cualquier entorno o tipo de formato. Sin embargo, los que se refieren al formato, la forma o el diseño son los que más pueden influir en un determinado medio. Ahora bien, la Web y los documentos electrónicos, debido a sus propias características, acentúan la importancia de determinados aspectos como la interactividad o la rápida puesta al día" (Jiménez-Piano; Ortiz-Repiso-Jiménez, 2007).

Se ha realizado¹ un estudio descriptivo (criterios de evaluación de fuentes digitales) y correlacional (vinculación y consecuencias para los centros de documentación) basado en el conocimiento de docentes investigadores y documentalistas de medios de comunicación generalistas con el fin de conocer el impacto de las fuentes digitales en los servicios de documentación periodística españoles en la era *Google* y las rutinas de las unidades de información en este nuevo contexto.

2. Objetivos y metodología

El objetivo principal es determinar los criterios de selección documental en fuentes de información online, concretamente en webs, blogs y redes sociales, en el ámbito periodístico.

Los objetivos específicos son:

- Descubrir qué tipo de fuentes digitales consultan los documentalistas de los medios para satisfacer las demandas informativas de las redacciones.
- Conocer la frecuencia de acceso a internet para la búsqueda de información en los servicios de documentación periodística.
- Detectar si en los centros existen políticas de tratamiento documental de las fuentes disponibles en la Red.
- Enumerar los cambios que está produciendo el uso de los recursos digitales en las rutinas y perfil profesional de los documentalistas.
- Indagar la opinión de expertos y profesionales de los medios sobre las consecuencias de la abundancia de recursos y facilidad de acceso a los contenidos de la Red para el periodismo.

Los estudios previos que se han tenido en cuenta son los de

metodología de análisis y evaluación de recursos digitales de **Codina** (2000; 2006a; 2006b) y los criterios de calidad web propuestos por **Jiménez-Piano** y **Ortiz-Repiso-Jiménez** (2007). En el ámbito periodístico cabe señalar el trabajo de **Abadal** y **Guallar** (2010) sobre los parámetros para la evaluación de prensa digital; el método de análisis y evaluación de hemerotecas en prensa digital de **Guallar**, **Abadal** y **Codina** (2012), revisado sucesivamente desde 2007; y la evaluación de la calidad en cibermedios de **Rodríguez-Martínez**, **Codina** y **Pedraza-Jiménez** (2012).

Respecto a los análisis sobre las consecuencias de las fuentes de información en internet en las rutinas y perfiles profesionales en los centros de documentación periodística destacan especialmente **Rubio-Lacoba** (2007 y 2010), **Rubio-Lacoba** y **Blanco-García** (2010) y **Micó-Sanz**, **Masip** y **García-Avilés** (2007 y 2009). También han abordado el uso de las fuentes digitales y la selección de material en estos servicios **Cebrián-Enrique** (1998 y 2012), **Castillo-Blasco** (2002), **Meso-Ayerdi** (2006), **Giménez-Rayó** (2007), **Marcos-Recio** (2008), **López-de-Quintana-Sáenz** (2010) y **Rodríguez-Mateos** (2011).

Los rasgos considerados esenciales en las fuentes digitales son autoría, actualización, objetividad, exactitud, pertinencia, coherencia del contenido y calidad de los enlaces

La metodología de la investigación es cualitativa y cuantitativa. La parte cualitativa del estudio fue un panel de expertos mediante técnica Delphi como fase preliminar para definir y justificar las variables. El tipo de panelista requerido para la investigación fue el de especialista en documentación informativa con publicaciones en evaluación y/o selección de fuentes, recursos informativos en internet y servicios de documentación en el periodismo digital.

La selección de expertos concluyó en el mes de febrero de 2013 con ocho especialistas que aceptaron participar en la dinámica Delphi: Lourdes Castillo-Blasco, Lluís Codina, Eulàlia Fuentes-Pujol, Gabriel Galdón, Mabel Giménez-Rayó, Javier Guallar, Tony Hernández-Pérez y María Rubio-Lacoba. El panel finalizó en junio de 2013 después de tres rondas de preguntas, al lograr el consenso de los participantes y obtener un informe final con las conclusiones que sirvió para preparar la parte cuantitativa del estudio: una encuesta autoadministrada por internet a documentalistas de medios de comunicación españoles.

El universo de la encuesta lo constituyen los documentalistas de medios de comunicación de información general y con cobertura nacional. En este punto cabe realizar una

aclaración esencial: el universo lo forman los profesionales que trabajan en esos medios y no las cabeceras y emisoras. Es decir, no todos los medios de comunicación generalistas españoles con difusión nacional disponen de un centro de documentación. Para determinar el universo, se contactó con diversos medios y se comprobó la existencia del servicio en trece de ellos, de prensa: *ABC*, *El mundo*, *El país*, *La razón* y *La vanguardia*; de televisión: *Antena 3*, *TVE*, *Cuatro*, *La sexta* y *Telecinco*; y de radio: *Cadena COPE*, *Cadena SER* y *RNE*. Quedaron descartados por la ausencia de documentalistas otros medios contactados como la emisora *Onda cero*, o diarios digitales que no editan en papel.

Una vez determinados los medios, en abril de 2013 se solicitó la colaboración de los responsables o antiguos responsables de los centros para distribuir el url de la encuesta. Todos los medios aceptaron participar en la investigación excepto el diario *La vanguardia*, que como consecuencia de la reducción de personal ha optado por restringir las peticiones externas.

Colaboraron los siguientes profesionales: Ángeles Afuera (*Cadena SER*), Concha Algarra (*Antena 3*), Federico Ayala (*ABC*), María-Carmen Bernal (*Cadena COPE*), Juan-Carlos Blanco (*El país*), Beatriz Cardoso (*La razón*), Ana García-de-Viedma (*El mundo*), Daniel García-Hernández (*La sexta*), Mónica Pérez-Carabías (*RNE*), José-Ángel Renedo (*TVE*) y Javier Solano (*Mediaset*).

Los datos proporcionados por los medios permiten concluir que el tamaño de la población o universo es de 189 profesionales². La muestra la conforman los 81 documentalistas que contestaron a la encuesta. El nivel de confianza es de 95,5% (2 sigmas), $p=q=0,5$ y un error muestral atribuible de 10,5% (Mateu; Casal, 2003). El tipo de encuesta fue autoadministrada en la web en los meses de junio y julio de 2013.

Se presenta un resumen de los resultados obtenidos en la investigación en los apartados siguientes.

Las fuentes más utilizadas son webs de medios, páginas institucionales, buscadores y hemerotecas digitales

3. Criterios de evaluación de fuentes digitales

Desde el punto de vista específico de la documentación informativa, los criterios tradicionales de evaluación documental de fuentes siguen vigentes para internet puesto que son independientes del soporte o modo de edición, según

Tabla 1. Vigencia de los criterios de evaluación de fuentes de información tradicionales

	Número de respuestas	%
Sí	49	60,5
No	22	27,2
No sabe	9	11,1
No contesta	1	1,2

todos los expertos del Delphi y el 60,5% de los documentalistas encuestados. Las características propias de la Red (acceso libre, amplificador de pseudofuentes, heterogeneidad caótica, desinformación, etc.) provocan que la valoración documental tenga que preservarse si cabe con más rigor.

Autoría, actualización, objetividad, exactitud, pertinencia y coherencia del contenido junto con la calidad de las referencias siguen considerándose esenciales para las tareas de los documentalistas de medios (búsquedas directas, recomendaciones a usuarios, filtrado, etc.). Otros aspectos a tener cuenta son: grado de reelaboración de la información, cobertura, lenguaje y derechos de uso.

Internet y las fuentes de elaboración propia son los recursos más consultados por los documentalistas de los medios

Por su parte, los expertos participantes en el Delphi enumeran nuevos criterios documentales para valorar las fuentes de información online: accesibilidad, aprovechamiento de la dimensión hipermedia, acceso a la información de archivo (búsqueda por interrogación, por navegación, etc.), usabilidad, organización de la información en el sitio web (arquitectura de la información) y visibilidad de la fuente en los motores de búsqueda.

En el caso de los documentalistas de los medios, los criterios que citan como fundamentales para valorar las fuentes de información de internet según orden de preferencia son:

- Webs: objetividad, exactitud, pertinencia y coherencia del contenido (79% de los encuestados); actualización (79%); autoría (69%); calidad de referencias y enlaces (60%); y organización de la información en la web/arquitectura de la información (51%). El resto de items propuestos no logra alcanzar un resultado mayoritario (lenguaje, acceso a la información de archivo, claridad de los objetivos y propósitos, usabilidad, grado de reelaboración de la información, visibilidad, derechos de uso y diseño web).
- Blogs: autoría (69%), actualización (62%), coherencia del contenido con sus partes (43%), lenguaje (35%) y organización de la información web/arquitectura de la información (33%).
- Cuentas de *Twitter*: autoría (64%), calidad de los enlaces y recursos adjuntos (49%), coherencia de las informaciones (44%), comportamiento de la cuenta a lo largo del tiempo (37%) e interacción con los usuarios (24%).

Tabla 2. Soporte de la fuente consultada

Soporte de la fuente	Respuestas
Fuentes de elaboración propia	63
Internet	67
Fuentes en papel	26
No sabe	0
No contesta	0
Otro	7

Si se comparan los cinco primeros criterios de evaluación obtenidos para cada tipo de fuente, se observa una coincidencia en la autoría y la objetividad, exhaustividad, pertinencia y coherencia del contenido. En el caso de las webs y los blogs existe aún una mayor similitud al coexistir la actualización y la organización de la información. Webs y blogs comparten cuatro de los cinco primeros rasgos considerados necesarios para evaluar las fuentes online.

4. Tipo y frecuencia de fuentes consultadas

Pervive la función de intermediación del documentalista en las búsquedas de información para los periodistas. Las fuentes digitales online junto con las de elaboración propia son las más consultadas por los documentalistas para responder a estas consultas, aunque se observa una tendencia a priorizar los recursos de internet.

Los documentalistas consultan diariamente la Red para responder a las demandas de los periodistas. El 86,4% de los encuestados dice que utiliza internet cada día para buscar información. Las webs de organismos oficiales e instituciones junto con las de medios de comunicación son las fuentes más consultadas.

Los centros de documentación periodística de medios generalistas y con difusión nacional no han establecido una política para la selección de fuentes de internet

La búsqueda de contenidos en redes sociales es muy baja. Sólo el 17% de los encuestados dice que las utiliza para resolver las peticiones de los periodistas. Las dos más consultadas y en una proporción idéntica son *Facebook* y *Twitter*. El uso de estas fuentes no está relacionado con la edad del profesional.

5. Políticas de selección y cambio en las rutinas y perfiles profesionales

A pesar del acceso masivo a internet y de la saturación informativa por la enorme cantidad de recursos disponibles, los centros de documentación de medios generalistas españoles y con difusión nacional no han establecido una política para la selección de fuentes de información en la Red.

El 63% de los documentalistas que ha contestado a la encuesta ha reconocido que no existen políticas y estrategias de selección de fuentes digitales en sus centros. Sin embargo, esta cuestión requiere un análisis más detallado para determinar la situación en cada uno de los medios.

La comparativa de las variables de política de selección y centro de procedencia de los profesionales permite concluir que no existen estrategias para las fuentes informativas digitales online en 6 de los 11 centros. El 100% de los documentalistas que han respondido a la encuesta y que trabajan en estos medios lo ha confirmado.

En el resto de los medios no se ha obtenido unanimidad en las respuestas en uno u otro sentido, lo que impide confir-

Tabla 3. Fuente online específica

Tipo de fuente	Respuestas
Webs de medios de comunicación	69
Webs institucionales	69
Buscadores	62
Hemerotecas digitales	53
Bases de datos	43
Blogs	11
Redes sociales	11
No sabe	0
No contesta	0
Otro	0

mar con certeza la existencia de este tipo de políticas. Sin embargo, de los porcentajes elevados de respuestas afirmativas en dos casos (medio 10 y 11) se deduce que sí que deben de existir estrategias o al menos algún tipo de pauta. En el caso contrario, de la mayoría de respuestas negativas del medio 7 se deduce que no existen tales recomendaciones. Resulta imposible determinar la presencia o ausencia de normas de selección en otros dos centros (medio 8 y 9).

Expertos y documentalistas consideran que las fuentes digitales han mejorado la calidad del servicio de las unidades de información por el crecimiento exponencial de los recursos y la rapidez en la respuesta de las consultas. Sin embargo, se confirma que el uso de internet en las redacciones ha provocado también una pérdida de control de los documentalistas sobre las búsquedas de los usuarios. Se ha producido una situación crítica para los profesionales que ven su función tanto valorizada (la abundancia de fuentes y la buena formación en competencias informacionales hace que puedan seguir ofreciendo servicios de calidad) como desvalorizada (los periodistas pueden acceder a las fuentes sin necesidad de intermediarios). Por ello, se propone un reciclaje profesional a causa de la gran cantidad de nuevas herramientas 2.0 (social media, sistemas de monitorización, RSS, buscadores...) que aparecen constantemente, ya que para responder a las demandas informativas se hace necesario combinar el uso de fuentes clásicas con fuentes digitales.

Tabla 4. Porcentaje de respuestas sobre políticas de selección por medios

	Sí	No	Ns	Nc
Medio 1	0	100	0	0
Medio 2	0	100	0	0
Medio 3	0	100	0	0
Medio 4	0	100	0	0
Medio 5	0	100	0	0
Medio 6	0	100	0	0
Medio 7	28,5	71,5	0	0
Medio 8	20,0	50,0	20,0	10,0
Medio 9	33,3	33,3	33,3	0
Medio 10	62,5	12,5	12,5	12,5
Medio 11	50,0	25,0	0	25,0

6. Periodismo documentado

Finalmente, los expertos participantes en la dinámica Delphi apuntan que la abundancia y facilidad de acceso a las fuentes de internet debería contribuir a un periodismo más documentado, pero eso está condicionado por el grado de profesionalidad de los periodistas. En el caso de los documentalistas, la proporción similar de respuestas (43% sí, 39% no) impide determinar si las fuentes digitales está favoreciéndolo o no. Entre aquellos que opinan que los recursos digitales no están ayudando, se enumeran las razones que siguen:

- Hay un contraste entre las posibilidades potenciales de la información digital y el uso real en la práctica periodística mayoritaria en nuestro país.
- Se considera que el periodismo actual es de mucha inmediatez, pero de poca documentación, verificación e interpretación.
- En algunos medios, teniéndose a mano buenos servicios y fondos, siguen sin utilizarse adecuadamente.
- Internet ha provocado el acceso directo a las fuentes por parte de los periodistas sin la colaboración de los documentalistas y aquellos no dedican el tiempo necesario a la búsqueda en internet, con lo que no se quedan con la mejor información, sino con la primera que encuentran.
- La falta de periodismo documentado también se explica, según expertos y profesionales, por la situación actual de crisis de las empresas periodísticas y la reducción de las plantillas en los medios.

“ Sólo el 17% de documentalistas consulta las redes sociales para resolver las peticiones de periodistas ”

7. Conclusiones

Los documentalistas de medios de comunicación consultan diariamente la Red para responder a las demandas de los periodistas. El 86,4% de los encuestados dice que utiliza internet cada día para buscar información. Se observa la tendencia a priorizar los recursos de internet, siendo las webs de medios, las páginas institucionales, los buscadores y las hemerotecas digitales las fuentes más utilizadas. La consulta de información en redes sociales es muy baja. Las más utilizadas y en una proporción idéntica son *Facebook* y *Twitter*. El uso de estas fuentes no está relacionado con la edad del profesional.

Los criterios principales para evaluar los recursos y otras fuentes en internet siguen siendo los específicos de la ciencia documental (autoría, actualización, pertinencia y exhaustividad del contenido, etc.). Nos encontramos pues ante otra oportunidad para subrayar la importancia del documentalista en su tarea de ayudar al periodista a descubrir, identificar y evaluar los recursos más valiosos, a pesar de que los profesionales encuestados reconocen que los periodistas consultan las fuentes digitales sin la participación de los documentalistas y que se ha reducido la labor de intermediario en el acceso a las fuentes.

A pesar de la saturación informativa, de la enorme cantidad de recursos de fácil acceso y del uso masivo de internet, los centros de documentación periodísticos españoles de medios generalistas y con difusión nacional no han establecido una política para la selección de fuentes de información en la Red. En este sentido, sería recomendable que los servicios establecieran esas políticas y estrategias, que conducirán al mayor acercamiento y colaboración con los periodistas, a aprovechar las posibilidades potenciales de la información digital y, en último término, a un uso periodístico más eficaz.

Notas

1. Esta investigación se ha realizado en el marco del proyecto “Supervivencia del periodismo en la era post-digital”, del que es miembro el investigador Bernardino J. Cebrián-Enrique (referencia CSO2011- 29510-C03-03), financiado por el *Ministerio de Economía y Competitividad* de España.

2. El cuestionario se distribuyó a la totalidad de la plantilla de los servicios de documentación, excepto en el caso de *TVE*, *RNE* y *Antena 3* que únicamente se envió a los profesionales que trabajan en las áreas relacionadas con las fuentes digitales

8. Bibliografía

Abadal, Ernest; Guallar, Javier (2010). *Prensa digital y bibliotecas*. Gijón: Trea. ISBN: 978 84 9704 446 2

Castillo-Blasco, Lourdes (2002). “La selección de las noticias en el entorno de un centro de documentación de información de actualidad”. *Revista española de documentación científica*, v. 25, n. 2, pp. 182-189.
http://bddoc.csic.es:8080/basisbwdocs_rdisoc/rev0001/2002_vol25-2/2002_vol25-2_pp182-189.htm

Cebrián-Enrique, Bernardino J. (1998). “Nuevos enfoques de la misión del documentalista en los medios de comunicación”. *Documentación en ciencias de la información*, n. 21, pp. 91-100.
<http://dialnet.unirioja.es/servlet/articulo?codigo=51416>

Cebrián-Enrique, Bernardino J. (2012). “Al rescate de la verificación periodística”. *Zer: Revista de estudios de comunicación = Komunikazio ikasketen aldizkaria*, v. 17, n. 33, pp. 227-241.
<http://www.ehu.es/ojs/index.php/Zer/article/view/10633>

Codina, Lluís (2000). “Evaluación de recursos digitales en línea: Conceptos, indicadores y métodos”. *Revista española de documentación científica*, v. 23, n. 1, pp. 9-44.

Codina, Lluís (2006a). “Información documental e información digital”. En: López-Yepes, José (Ed.). *Manual de ciencias de la documentación* (2ª ed.). Madrid: Pirámide, pp. 301-315. ISBN: 978 84 368 2032 4

Codina, Lluís (2006b). *Metodología de análisis y evaluación de recursos digitales en línea* (v. 2006). Barcelona: UPF. Área de Biblioteconomía y Documentación. Departamento de Periodismo y Comunicación Audiovisual.
<http://www.digidocweb.net/metodos/procedimientos2006.doc>

García-Avilés, José-Alberto; Masip, Pere; Micó-Sanz, Jo-

- sep-Lluís** (2007). "La redefinición del perfil y funciones del documentalista en las redacciones digitales de medios españoles". *IX Jornadas de gestión de la información. Informar y difundir: servicios documentales y de comunicación*. Madrid, pp. 28-42.
<http://eprints.rclis.org/10845/1/SedicJGI07-Garcia-Masip-Mico.pdf>
- Giménez-Rayó, Mabel** (2007). *Documentación audiovisual de televisión: la selección del material*. Gijón: Trea. ISBN: 978 84 9704 320 5
- Guallar, Javier; Abadal, Ernest; Codina, Lluís** (2012). "Sistema de análisis de hemerotecas en prensa digital". *Trípodos*, n. 31, pp. 37-64.
http://www.tripodos.com/index.php/Facultat_Comunicacio_Blanquerna/article/view/37
- Jiménez-Piano, María; Ortiz-Repiso-Jiménez, Virginia** (2007). *Evaluación y calidad de sedes web*. Gijón: Trea. ISBN: 978 84 9704 318 2
- López-de-Quintana-Sáenz, Eugenio** (2010). "Tecnología, convergencia y cambio: un viaje sin retorno". *Revista general de información y documentación*, v. 20, pp. 435-453.
- Marcos-Recio, Juan-Carlos; Edo-Bolós, Concha; Flores, Jesús; García-Alonso, Pedro; Parra, David** (2008). "Estrategias y perspectivas documentales en la información digital". *Ámbitos: Revista internacional de comunicación*, n. 17, pp. 9-23.
<http://eprints.ucm.es/9550/1/%C3%81mbitosNro.17-2008-EstrategiasDocu-InformacionDigital.pdf>
- Mateu, Enric; Casal, Jordi** (2003). "Tamaño de la muestra". *Revista de epidemiología y medicina preventiva*, n. 1, pp. 8-14.
<http://videodigitals.uab.es/cr-vet/www/21216/TamanoMuestra3.pdf>
- Meso-Ayerdi, Koldo** (2006). *Introducción al ciberperiodismo: Breve acercamiento al estudio del periodismo en internet*. Bilbao: Servicio Editorial de la Universidad del País Vasco. ISBN: 978 84 8383 861 0
- Micó-Sanz, Josep-Lluís; Masip, Pere; García-Avilés, José-Alberto** (2009). "Periodistas que ejercen de documentalistas (¿y viceversa?). Nuevas relaciones entre la redacción y el archivo tras la digitalización de los medios". *El profesional de la información*, v. 18, n. 3, pp. 284-290.
<http://www.elprofesionaldelainformacion.com/contenidos/2009/mayo/05.pdf>
<http://dx.doi.org/10.3145/epi.2009.may.05>
- Rodríguez-Martínez, Ruth; Codina, Lluís; Pedraza-Jiménez, Rafael** (2012). "Indicadores para la evaluación de la calidad en cibermedios: Análisis de la interacción y de la adopción de la web 2.0". *Revista española de documentación científica*, v. 35, n. 1, pp. 61-93.
<http://repositori.upf.edu/handle/10230/16361>
<http://dx.doi.org/10.3989/redc.2012.1.858>
- Rodríguez-Mateos, David** (2011). "Internet y su influencia sobre la documentación audiovisual". En: Caridad-Sebastián, Mercedes; Hernández-Pérez, Tony; Rodríguez-Mateos, David; Pérez-Lorenzo, Belén (Eds.). *Documentación audiovisual: nuevas tendencias en el entorno digital*. Madrid: Síntesis, pp. 87-110.
- Rubio-Lacoba, María** (2007). *Documentación informativa en el periodismo digital*. Madrid: Síntesis. ISBN: 978 84 9756 878 4
- Rubio-Lacoba, María** (2010). "Documentalistas de prensa, ¿cuál es vuestro oficio?". *El profesional de la información*, v. 19, n. 6, pp. 645-651.
<http://eprints.rclis.org/15121>
<http://dx.doi.org/10.3145/epi.2010.nov.11>
- Rubio-Lacoba, María; Blanco-García, Juan-Carlos** (2010). "Mejor que un buscador, un encontrador". *Documentación de las ciencias de la información*, v. 33, pp. 273-287.

Próximos temas centrales

Enero 2015	Marketing de información
Marzo 2015	Servicios bibliotecarios
Mayo 2015	Nuevos medios
Julio 2015	Bases de datos bibliográficas
Septiembre 2015	Difusión científica en internet
Noviembre 2015	Periodismo computacional
Enero 2016	Nuevos profesionales de la información

Los interesados por favor consulten detenidamente las Normas para autores:

<http://www.elprofesionaldelainformacion.com/autores.html>

y luego envíen sus artículos a través del gestor de manuscritos OJS de la plataforma del Recyt:

<http://recyt.fecyt.es/index.php/EPI/index>



NEWS OVERLOAD IN SPAIN: THE ROLE OF DEMOGRAPHIC CHARACTERISTICS, NEWS INTEREST, AND CONSUMER PAYING BEHAVIOR



Manuel Goyanes



Manuel Goyanes has a PhD in communication science from the *University of Santiago de Compostela* and master's degree in statistics (*Universidad Politécnica de Madrid*). Currently, he is a post-doc in media management at *Carlos III University of Madrid*. He has presented at several international congresses and published in international journals such as *Journalism practice*, *International journal on media management*, and *First Monday*. His research interests include information overload, consumer culture, and entrepreneurial intention.
<http://orcid.org/0000-0001-8329-0610>

Carlos III University of Madrid
Madrid, 133. 28903 Getafe (Madrid), Spain
manuel.goyanes@uc3m.es

Abstract

This article explores factors that correlate with the perception of news overload and the extent to which Spaniards feel overloaded, or not, with the amount of news available. Specifically, this study explores the following factors: demographics (age, gender, and income), news interest, and consumer paying behavior for online news. The research method employs a quantitative, explorative approach with data collected from an online survey of 1,439 Spanish adults, 18 years and older, in January 2013. Finally, managerial and theoretical implications are discussed.

Keywords

News overload, news interest, consumer paying behavior, online newspapers, online news.

Título: Sobrecarga informativa en España: el rol de variables demográficas, el interés en las noticias y el comportamiento de pago del consumidor

Resumen

La presente investigación explora factores noveles que presumiblemente afectan a la percepción de sobrecarga informativa, así como analiza su nivel de expansión en España. Particularmente, el estudio analiza el rol de diferentes variables demográficas (sexo, edad e ingresos), el interés por las noticias y el comportamiento de pago del consumidor por su periódico digital favorito. Para ello la investigación aplica una aproximación cuantitativa y exploratoria a través de una encuesta a 1.439 adultos españoles realizada en enero de 2013. Por último, se discuten las implicaciones teóricas y gerenciales.

Palabras clave

Sobrecarga informativa, Interés por las noticias, Comportamiento de pago, Diarios digitales, Noticias online.

Goyanes, Manuel (2014). "News overload in Spain: the role of demographic characteristics, news interest, and consumer paying behavior". *El profesional de la información*, v. 23, n. 6, November-December, pp. 618-624.

<http://dx.doi.org/10.3145/epi.2014.nov.09>

Introduction

Over the last two decades the amount of information available has dramatically increased (**Serra**, 2004). This trend has been exacerbated by the rise of the Internet; more specifically by the World Wide Web in the 1990s and later by the advent of social networking tools (**Bawden**; **Robinson**, 2009). The large increase of available information is exemplified with a few statistics: everyday more than 3,000 books

are published worldwide; the number of documents on the Internet doubled from 400 million to 800 million from 1998 to 2000; more information has been created in the past 30 years than in the previous 5,000 years; and the weekly edition of the *New York Times* contains more information than an average person was likely to come across in his lifetime during the seventeenth-century in England (**Bawden**, 2001). This vast oversupply of information is increasingly causing psychological and physiological problems, ranging from

Article received on 19-09-2014

Final acceptance: 29-10-2014

increased levels of stress when working through email inboxes to anxiety about keeping up with multiple social networking sites such as *Facebook* and *Twitter* (Swartz, 2011; Bregman, 2012).

The problems associated with information overload have been addressed in several disciplines including consumer behavior, information science, management information systems, marketing, and organizational communication. Communication scholars are now interested in information overload; specifically how news consumption behavior might trigger or complement problems associated with information overload. Nordenson (2008) points out that some newspapers' readers are suffering increased levels of news fatigue. A clear example of this trend is the US pattern of news consumption: despite the fact that the majority of the American adults get news content from multiple news outlets, an important segment of consumers report problems associated with an incapacity to process information and, even more surprisingly, with discerning between important and useless information. As a result, some scholars (Hermida, 2010) have theorized that news consumption behavior could be described as "ambient news," suggesting that news content is ubiquitous and omnipresent.

“The problems associated with information overload have been addressed in several disciplines”

News consumption of traditional media is on the decline while its digital counterpart is on the rise. Therefore, media managers are focusing their efforts on maintaining and increasing audiences while also monetizing the digital product. Since the profitability and relevance of news organizations are being threatened by new—and practically free—content, studying how to reduce or decrease the actual or perceived level of news overload for online newspaper consumers deserves an empirical examination. Previous empirical research in this area has focused only on the analysis of information overload by consumers in the US. Therefore, we can only speculate about how the findings can be generalized to other populations such as those in European countries where the attitudes and news consumption patterns are different. Thus, the main objective of the present research is to analyze news content as a form of information overload, examining the extent to which Spaniards feel overloaded with the amount of available news. Also, this study explores factors that correlate with the perception of news overload and the extent to which Spaniards feel overloaded, or not, with the amount of news available. Specifically, this study explores the following factors: demographics (age, gender, and income), news interest, and consumer paying behavior for online news. Theoretical and managerial implications are discussed.

Theoretical framework

News consumption and delivery has dramatically changed since the arrival of the Internet. Now, more than ever, tra-

ditional media have been challenged by new digital media, from both inside and outside the journalism profession, as the main and central providers of information and entertainment (Picard, 2003). This competition has created a saturation of news content and platforms (Holton; Chyi, 2012) and can contribute to the psychological feeling of information overload. Despite the fact that there is no single generally accepted definition of information overload, the term usually represents a state of affairs where an individual's efficiency in using information is hampered by the large amount of relevant, and potentially useful, information available to them (Bawden; Robinson, 2009). According to Bawden, Hotman, and Courtney (1999), information overload occurs when the amount of available content becomes difficult to process, often causing negative feelings for the end user. In addition, some other psychological conditions have been associated with information overload, such as continuous partial attention (Stone, 2008), attention deficit (Hallowell, 2005), cognitive overload (Kirsh, 2000), and technostress (Thornton, 2007).

The omnipresent nature of information in the current digital panorama is intensified by the popularity of different digital devices (tablets, smartphones, e-readers, etc.) that enable the user to continuously consume updated information. Some segments of the US population rely on multiple news media sources (Diddi; LaRose, 2006, Pew Research, 2010), and according to Pew Research (2014), 92 percent of American adults get news content from multiple platforms on a typical day—half of those using four to six platforms daily¹. While news viewership and readership in traditional media are declining, when online and mobile consumption are included, consumers actually spend more time with news products (Pew Research Center, 2010). Consumers may prefer to read the news online, but they continue to read the news (Chyi; Chadha, 2011). In this context, where the digital content is omnipresent and its access is at hand to any user by a simple click at anytime, the number of cases of people that suffer news overload is increasing.

“The main objective of the present research is to analyze news content as a form of information overload”

Previous research studies have analyzed the impact of different news delivery platforms on news overload. For example, Holton and Chyi (2012) found news access through platforms and outlets such as computers, e-readers, and Facebook had a positive association with overload, whereas other platforms such as television and the iPhone were negatively associated with overload. On the other hand, one demographic factor, gender (being female), was positively related to news overload whereas two other demographic factors, age and income, were negatively associated with news overload (Ji, Ha; Sypher, 2014). Personal preference, such as news interest, was a significant and negative predictor (Holton; Chyi, 2012), that is, people who enjoy keeping up with news are less likely to suffer news overload.

In regards to perceived news overload, 141 (24.5 percent) of respondents in one study said they had never experienced information overload; while 434 (75.5 percent) replied they had experienced some level of news overload (Ji, Ha; Sypher, 2014).

In spite of the theoretical and empirical contributions of these previous studies, there is a surprising lack of research focused on non-US samples. Therefore, we can only speculate about how the findings can be generalized to other populations such as those in European countries where the attitudes and news consumption patterns are different. For example, in Spain various online media outlets are used for news consumption but are not as widespread as in the US (Reuters, 2013). Furthermore, tablet penetration in Spain is not as high as in the US, despite the fact that smartphone use for information consumption in Spain is one of the highest in the world. In this context, with the aim of measuring how Spaniards feel with the amount of news content now available and how demographic variables and news interest influence the perception of news overload, this study asked the following research questions:

RQ1: To what extent do Spaniards feel overloaded with the amount of news available?

RQ2: How does news interest influence the perceived degree of news overload?

In addition, this study aims to determine if there is a relationship between consumers' paying behavior and news overload. This relationship has not been studied previously; however, the recent increase of digital paywalls in various news organizations allows us to empirically test whether there are significant differences between paid and non-paid readers with respect to the level of news overload. According to the *Pew Research Center* (2014), digital pay plans are being adopted at 450 US news outlets including major news organizations (such as *The New York Times* and *Wall Street Journal*) and also small and mid-sized papers (Mitchell, 2013). With regard to Spain, digital pay plans are exponentially increasing, although notable is the complete absence of the freemium and metered model (Goyanes; Dürremberg, 2014).

The exploratory analysis of this relationship is important for several reasons. First, because it will compare consumer behavior patterns between free vs. paid news consumption with respect to news overload. Second, because the analysis may provide managerial insights that help reduce news overload among Spanish news consumers. Finally, the results may help managers choose appropriate business models and design effective revenue-generation methods. Therefore, this study includes a third research question:

RQ3: Are Spanish consumers who are paying for their favorite online newspaper more or less likely to suffer news overload when compared to those who consume free news?

Method

A randomly selected sample of 1,439 Spanish adults (18 years and older) were administered an online survey between January 14 and 17, 2013² to examine the perception

of news overload in Spanish Internet users. The participants were recruited by snowball sampling (Atkinson; Flint, 2001). Snowball sampling may simply be defined as a technique for finding research subjects. One subject gives the researcher the name of another subject, who in turn provides the name of a third, and so on (Vogt; Johnson, 2001). This technique is a widely used method, especially for research with an exploratory nature in which the descending methodologies (such as household surveys) can suffer a lack of responses from particular groups—usually young people and unemployed males (Atkinson; Flint, 2001). Ascending methodologies, such as the use of snowball techniques, can be used to work upwards and locate those on the ground who are needed to fill in the gaps in our knowledge on a variety of social contexts. In this sense, snowball sampling can be considered as an alternative or as a complementary strategy for attaining more comprehensive data on a particular research question (Atkinson; Flint, 2001).

“ In spite of the theoretical and empirical contributions of previous studies, there is a surprising lack of research focused on non-US samples ”

The survey was distributed via web communities and social networking sites such as *Twitter* and *Facebook*. Personal networks of the researchers were employed and also the accounts and pages of Spanish newspapers. The sample size of 1,439 yielded a standard sampling error of ± 3.1 percentage points at the 95 percent confidence level. The sample included slightly more women (57.1 percent, $N = 822$) than men (42.9 percent, $N = 617$) with an age range between 18 and 85. The survey took an average of five minutes to complete. Each participant was asked to enter his or her account number and password before participating in the survey to ensure the sample included no duplications. The participation³ (or completion rate) was 70.45 percent. In this study, in terms of statistical power, the overall sample size (1,439 respondents) contributes to the stability of statistical analysis since larger sample sizes tend to produce more reliable sample estimates. The survey sample of 1,439 was compared with the Internet population of Spain for control variables (i.e. age, gender, and income) to ensure representation. In this case, the survey sample overrepresented females and people with lower income. The survey sample was weighted to these two particular variables. The final weighted sample size was 1,445. In general, the weighted sample is comparable to the population in terms of gender, age, and income but underrepresented those with an age between 24 and 44 years old.

The survey focused on the factors that trigger the perception of news overload. A questionnaire was created by three university academics and then pilot-tested on four graduate students and on six academics (one post-doctoral fellow, two assistant professors, one associate professor, and two full professors). Their comments on the content of the survey instrument, item wording, terminology, and clarity were

incorporated into a revised instrument. Next, the refined instrument was piloted again on a larger sample of 50 university students (which are not part of the sample), and final revisions were made.

The measurement of news overload was operationalized asking respondents the level of agreement with the following statement “I often feel overload with the amount of news available these days” on a four-point Likert scale with anchors ranging from 1 = *Completely disagree*; 4 = *Completely agree*. Despite being a construct measured by a single item, previous empirical research has employed the same operationalization (Chyi; Chadha, 2012; Ji, Ha; Sypher, 2013). News interest was measured by asking respondents: “In general, how much do you enjoy keeping up with the news?” on a four-point Likert scale with anchors ranging from 1 = *Not at all* and 4 = *A lot*. This measurement was adopted from Pew Research Center as well as other empirical research studies (Holton; Chyi, 2012; Goyanes, 2014). Consumers’ paying behavior was measured asking respondents: “Are you currently paying for your favorite online newspaper?” 0 = *No*; 1 = *Yes*. Finally, controls (demographics: age, gender, and income) were collected by standard survey measurements.

“The analysis revealed age as a significant and negative predictor of news overload”

The model constructed in this research is based on a hierarchical multiple regression with the aim of predicting readers’ perceptions of news overload. Based on the theoretical framework developed, the predictor variables were entered in the equation in three different blocks: demographics (gender, age and, income), news interest, and finally consumers’ paying behavior. A multicollinearity test confirmed that no independent variables were highly correlated with each other.

Results

Regression analysis was conducted to answer the three research questions posed in the theoretical framework. Only the RQ1 can be answered with a frequency analysis. The descriptive statistics and correlations of the quantitative variables are presented in Table 1. From the total sample of 1,439 Spanish adults, only 6.7 percent ($N = 97$) and 19.0 percent⁴ ($N = 273$) of respondents indicated they completely or mostly disagreed with the statement “I often feel overload

with the amount of news available these days”. Therefore, most of the respondents’ perceptions of news overload were very high. Most of the respondents mostly agree (44.1 percent; $N = 635$) while 27.3 percent ($N = 393$) completely agree. The answers to the rest of the research questions (RQ2 and RQ3) are presented in table 2.

“It might be the case that the relationship between news satisfaction and news overload is moderated by the fact that the reader is actually paying for content or not”

Hierarchical regression analysis was conducted to examine the relationship between demographics, news interest, and consumers’ paying behavior on news overload. All three models were statistically significant as well as the R^2 change ($p < 0.05$). Demographics (age, gender, and income) were entered on the first step of the regression analysis. Only age had a significant and negative association with news overload ($\beta = -0.008$; $p < 0.05$). Therefore, younger people felt more overloaded with the news available today when compared with older people. Both gender and income did not statistically influence the perception of news overload. This first model only accounted for 2 percent of the variance in news overload ($R^2 = 0.02$).

News interest was entered at the second step of the hierarchical regression analysis. In addition to age ($\beta = -0.007$; $p < 0.05$), news interest was also a significant and negative predictor of news overload ($\beta = -0.168$; $p < 0.05$). Therefore, the higher the users’ news interests were the lower the perception of news overload was. This second model accounted for the 7 percent of the variance of news overload ($R^2 = 0.07$). Finally, consumers’ paying behavior was entered on the fourth step of the hierarchical regression analysis. In addition to age ($\beta = -0.009$; $p < 0.05$) and news interest ($\beta = -0.187$; $p < 0.05$) consumers’ paying behavior was a significant and negative predictor of news overload ($\beta = -0.545$; $p < 0.05$). Therefore, people who were paying for access to a favorite online newspaper suffered less news overload with respect to those readers who consumed free information. This fourth model accounted for 15 percent of the variance of news overload ($R^2 = 0.15$).

Table 1. Descriptive statistics and correlations of quantitative variables. * $p < 0.05$; ** $p < 0.01$

	M	SD	1	2	3	4
Age	50.93	17.65	1.00	0.09**	-0.14**	-0.05*
Income	4.84	2.31	0.09**	1.00	0.09**	-0.01
News interest	3.41	0.85	-0.14**	0.09**	1.00	-0.06*
Overload	2.95	0.86	-0.05*	-0.01	-0.06*	1.00

Table 2. Hierarchical regression analysis of predictors of news overload. * $p < 0.05$; ** $p < 0.01$

	Block 1	Block 2	Block 4
Gender	0.063	0.067	0.018
Age	-0.008**	-0.007**	-0.009**
Income	0.001	0.003	0.002
News interest		-0.168*	-0.187*
WTP			-0.545**
F statistic	8.67**	7.85**	3.41**
R^2	0.02	0.07	0.15
Adjusted R^2	0.00	0.03	0.11

Discussion and conclusions

The research presented in this article explored factors that negatively or positively correlated with the perception of news overload and the extent to which Spaniards feel overloaded with the amount of news available. Specifically, the study explored the following factors: demographics (age, gender, and income), news interest, and consumer paying behavior. To that end, the research employed a quantitative, explorative approach with data collected from an online survey of 1,439 Spanish adults, 18 years and older, in January 2013. Using a hierarchical multiple regression analysis, the results of the study indicate a perception of news overload among Spaniards and a significant and negative relationship between age, news interest, and consumers' paying behavior.

“ Younger generations have considerable exposure to various information outlets, and can be expected to be more susceptible to news overload ”

The results of this study point to several possible theoretical and managerial implications. First, the analysis revealed age as a significant and negative predictor of news overload in all three models. To a certain extent these results replicated **Williamson and Eaker's** (2012), **Honton and Chyi's** (2012), and **Ji, Ha and Sypher's** (2014) findings that age negatively correlates with the perception of news overload; young adults are generally more likely to suffer news overload. However, this study is unlike the previous studies, which were conducted in the US, because it did not find gender (being female) or income (positive) as a significant predictor of news overload.

What are some possible explanations for our findings? Currently there are multiple sources from which people can obtain their information and communication (*iPad, iPhone, PC, Netbook, etc.*).

Younger generations, in both Spain and the US, have had considerable exposure to various information outlets (*Pew Research Center, 2013*), and can be expected to be more susceptible to news overload (**Ji, Ha; Sypher, 2014**). On the other hand, the data indicated a non-significant relationship when correlating gender and income to news overload (which is surprising when compared to previous studies in the US). This finding and comparison with previous studies highlights the difficulty of generalizing research from one country's population to another when they have differing cultural, economic, and social backgrounds. Future research should explore the cultural, gender, and socio-economic differences between countries as they relate to the feeling of news overload.

This study also found news interest was a negative predictor of news overload, suggesting that people who are interested in news are actually less, not more, overloaded with the information available. The results are supported by previous empirical research (**Holton; Chyi, 2012**). A possible explanation for this result may be the readers' consumption

habits; perhaps people who are interested in news are capable of filtering exactly what they are looking for or want to consume (**Savolainen, 2007**) through different channels and time schedules and are, therefore, able to decrease the perception of news overload. In comparison, perhaps "sporadic" readers who do not particularly enjoy reading the news feel overloaded with the amount of information and consequently lost in decoding and processing complicated news stories (**Thompson, 2009**). To some extent, what this research indicates is that filtering (**Shirky, 2008**) and organizing the news through different software programs and according to ones' personal capacity would lead to a decrease in the level of news overload. The personalization of information delivery by news organizations would be, therefore, a suitable strategy in assisting consumers in avoiding the perception of news overload.

This study also provides an empirical foundation for future research into news overload by analyzing a statistically significant and negative predictor of news overload, consumer paying behavior. Readers who are currently paying for their favorite online newspaper are less likely to feel overloaded than those who do not pay and instead consume only free news. This finding suggests that online newspapers with paid content strategies serve as filters for the readers' consumption of digital information. The development of computer software programs that personalize (and also interact and help to customize) the news to fit the needs and wants of the readers may help explain this finding.

This result proposes a new research avenue for future research since, it is more than probable that people who feel satisfied with the information provided by news organizations (no matter whether the organizations develop free or paid content strategies) also suffer more news overload than those who are not. However, in online newspapers with paid content strategies, despite the fact that it is more than probable also that the satisfaction with the information provided is high, the news overload experimented (by the paid readers) is low. Maybe, users who have high desire and not paying for news subscription will feel greater news overload, as compared to those with high desire but paid for news subscription. Therefore, it might be the case that the relationship between news satisfaction and news overload is moderated by the fact that the reader is actually paying for content or not.

“ Filtering and organizing the news through software programs and according to ones' personal capacity would lead to a decrease in the level of news overload ”

Future studies and limitations

This study has weaknesses that should be addressed in future research. Because there is no generally embraced operationalization of news overload and news interest, the present study used single-item measures for these two variables, which may limit the construct validity of the study.

One consequence of the use of single items might be the low R^2 showed in the hierarchical regression analysis, since only costumers' paying behavior considerable increment the explained variance in news overload. However, in both cases (news overload and news interest) the measurement technique was adapted from previous empirical research (Holton; Chyi, 2012; Ji; Ha; Sypher, 2013).

Future studies focused on the analysis of news overload should expand the selection of predictor variables used in the regression models, and more specifically, incorporate personal characteristics or personality factors that might explain a large percentage of variance in news overload. Also, researchers should empirically measure possible interaction effects, since this would extend our knowledge of the phenomena and open new avenues of research (for example, the interaction between consumers' satisfaction and paying behavior). Finally, the development of a more reliable scale to measure news overload that incorporates items such as the feeling of being overwhelmed, perceived manageability of information, and so forth, would be highly beneficial.

Notes

1. In Spain, various online media are used for news consumption: 56 percent of Spanish adults access news via computer, 35 percent via smartphone and 13 percent via tablets (Reuters, 2013; n = 979). Multiplatform use might not be as prevalent in Spain as in the US, but can reasonably be expected to grow.
2. Timing was not a factor to consider.
3. That is, the number of people who completed the survey divided by the number of people who responded to the survey invitation.
4. A total of 25.7 percent of respondents.

References

- Atkinson, Rowland; Flint, John (2001). "Accessing hidden and hard-to-reach populations: Snowball research strategies". *Social research update*, v. 33, n. 1, pp. 1-4. <http://sru.soc.surrey.ac.uk/SRU33.pdf>
- Bawden, David (2001). "Information overload". *Library & information briefings*, n. 92, pp. 1-15.
- Bawden, David; Holtham, Clive; Courtney, Nigel (1999). "Perspectives on information overload". *Aslib proceedings*, v. 58, n. 8, pp. 249-255. <http://dx.doi.org/10.1108/EUM000000006984>
- Bawden, David; Robinson, Lyn (2009). "The dark side of information: overload, anxiety and other paradoxes and pathologies". *Journal of information science*, v. 35, n. 2, pp. 180-191. http://www.bollettinoadapt.it/old/files/document/21976david_b-2008.pdf <http://dx.doi.org/10.1177/0165551508095781>
- Bregman, Peter (2012). "Coping with email overload". *Harvard business review*, 26 April. <http://blogs.hbr.org/2012/04/coping-with-email-overload>
- Chyi, Hsiang-Iris; Chadha, Monica (2012). "News on new devices: Is multi-platform news consumption a reality?". *Journalism practice*, v. 6, n. 4, pp. 431-449. <http://www.tandfonline.com/doi/pdf/10.1080/17512786.2011.629125> <http://dx.doi.org/10.1080/17512786.2011.629125>
- Didi, Arvind; LaRose, Robert (2006). "Getting hooked on news: Uses and gratifications and the formation of news habits among college students in an Internet environment". *Journal of broadcasting & electronic media*, v. 50, n. 2, pp. 193-210. http://dx.doi.org/10.1207/s15506878jobem5002_2
- Goyanes, Manuel (2014). "An empirical study of factors that influence the willingness to pay for online news". *Journalism practice*, pp. 1-16. <http://www.tandfonline.com/doi/full/10.1080/17512786.2014.882056#.VFCz64twsug> <http://dx.doi.org/10.1080/17512786.2014.882056>
- Goyanes, Manuel; Dürrenberg, Catherina (2014). "A taxonomy of newspapers based on multi-platform and paid content strategies: Evidences from Spain". *International journal on media management*, v. 16, n. 1, pp. 27-45. <http://dx.doi.org/10.1080/14241277.2014.900498>
- Hallowell, Edward (2005). "Overloaded circuits". *Harvard business review*, January. <http://www.integrity-plus.com/eStore/WP/overload%20circuitsR0501Ef2.pdf>
- Hermida, Alfred (2010). "Twittering the news: The emergence of ambient journalism". *Journalism practice*, v. 4, n. 3, pp. 297-308. <http://dx.doi.org/10.1080/17512781003640703>
- Holton, Avery; Chyi, Hsiang-Iris (2012). "News and the overloaded consumer: Factors influencing information overload among news consumers". *Cyberpsychology, behavior, and social networking*, v. 15, n. 11, pp. 619-624. <http://dx.doi.org/10.1089/cyber.2011.0610>
- Ji, Qihao; Ha, Louisa; Sypher, Ulla (2014). "The role of news media use and demographic characteristics in the possibility of information overload prediction". *International journal of communication*, v. 8, pp. 699-714. <http://ijoc.org/index.php/ijoc/article/viewFile/2419/1093>
- Kirsh, David (2000). "A few thoughts on cognitive overload". *Intellectica*, v. 1, n. 20, pp. 19-51. http://adrenaline.ucsd.edu/kirsh/Articles/Overload/Cognitive_Overload.pdf
- Mitchell, Bill (2013). "How two small family-owned newspapers in Vermont had success with a paywall". <http://www.poynter.org/latest-news/top-stories/222073/how-two-small-family-owned-newspapers-in-vermont-had-success-with-a-paywall>
- Nordenson, Bree (2008). "Overload!". *Columbia journalism review*, 30 Nov. http://www.cjr.org/feature/overload_1.php?page=all
- Pew Research Center. (2010). *State of the news media 2010*. <http://stateofthemedias.org/2010>

Pew Research Center (2014). *State of the news media 2014*. <http://www.journalism.org/packages/state-of-the-news-media-2014>

Picard, Robert (2003). "Cash cows or entrecote: Publishing companies and disruptive technologies". *Trends in communication*, v. 11, n. 2, pp. 127-136. <http://wiiu4.free.fr/pdf/CashCows.pdf>
http://dx.doi.org/10.1207/S15427439TC1102_04

Reuters (2013). *Reuters digital news report*. <http://www.digitalnewsreport.org>

Savolainen, Reijo (2007). "Filtering and withdrawing: strategies for coping with information overload in everyday contexts". *Journal of information science*, v. 33, n. 5, pp. 611-621. <http://dx.doi.org/10.1177/0165551506077418>

Serra, Marcello (2012). "Wikileaks: el poder entre bastidores". *Estudios sobre el mensaje periodístico*, v. 18, n. 1, pp. 211-220. http://dx.doi.org/10.5209/rev_ESMP.2012.v18.n1.39366

Shirky, Clay (2008). "It's not information overload. It's filter failure". *Web 2.0 Expo NY*. <http://blip.tv/web2expo/web-2-0-expo-ny-clay-shirky-shirky-com-it-s-not-information-overload-it-s-filter-failure-1283699>

Stone, Linda (2008). "Continuous partial attention. Not the same as multi-tasking". <http://lindastone.net/qa/continuous-partial-attention>

Swartz, Jon (2011). "Social media users grapple with information overload". *USA today*, 2 Febr. http://usatoday30.usatoday.com/tech/news/2011-02-01-tech-overload_N.htm

Thompson, Matt (2009). "An antidote for web overload". *Nieman reports*. <http://niemanreports.org/articles/an-antidote-for-web-overload>

Thornton, Glenda (2007). *Information tomorrow: Reflections on technology and the future of public and academic libraries*. Medford, NJ: Information Today. ISBN: 978 1 57387 303 1
<http://dx.doi.org/10.1080/01462670802368232>

Vogt, Paul; Johnson, Burke (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*. Thousand Oaks: Sage. ISBN: 978 1 4129 7109 6

Williamson, Jeanine; Eaker, Christopher (2012). "The information overload scale". *Asist 2012*, Baltimore, MD.

El profesional de la información
The information professional

INTERNATIONAL JOURNAL ON INFORMATION AND COMMUNICATION

EPI

EPI deja la edición en papel

Adaptándose a los nuevos tiempos, a partir del próximo mes de enero de 2015 esta revista *El profesional de la información* va a publicarse sólo online.

Al mismo tiempo, debido al cese de actividad en abril de 2015 de la empresa *MetaPress*, del grupo *Ebsco*, la revista se instalará en el *Repositorio Español de Ciencia y Tecnología (Recyt)* de la *Fundación Española para la Ciencia y la Tecnología (Fecyt)*:

<http://recyt.fecyt.es/index.php/EPI>

Desde ahí los suscriptores podrán acceder a los contenidos desde el año 2000 hasta la actualidad.

La web <http://elprofesionaldelainformacion.com> continuará funcionando como ahora, con toda la información sobre la revista, y los contenidos 1992-2010.



APORTACIONES DE UN CLUB DE LECTURA ESCOLAR A LA LECTURA POR PLACER



Carmen Álvarez-Álvarez y Julián Pascual-Díez



Carmen Álvarez-Álvarez es doctora en pedagogía y profesora ayudante doctor en la *Facultad de Educación* de la *Universidad de Cantabria*. Sus líneas de investigación son las relaciones teoría-práctica, la animación a la lectura, la innovación escolar y las comunidades de aprendizaje.
<http://orcid.org/0000-0002-8160-2286>

Universidad de Cantabria, Departamento de Educación
Av. Castros, s/n. 39005 Santander, España
carmen.alvarez@unican.es



Julián Pascual-Díez es doctor en pedagogía y profesor titular de didáctica de la lengua y literatura en la *Facultad de Educación* de la *Universidad de Oviedo*. Sus líneas de investigación son: didáctica de la literatura, promoción de la lectura y formación del profesorado.
<http://orcid.org/0000-0003-2694-9647>

Universidad de Oviedo, Departamento de Educación
Aniceto Sela, s/n. 33005 Oviedo, España
jpascual@uniovi.es

Resumen

Uno de los principales retos en el trabajo de la lectura en los centros educativos es el cultivo de la lectura por placer en los niños. Se presenta un estudio de caso único realizado en un club de lectura escolar en España que ha logrado conseguir este objetivo y ha recibido una valoración muy positiva por parte de los participantes. Las razones de su éxito residen fundamentalmente en cuatro claves: planificación, lectura dialógica, formación crítica y educación en valores.

Palabras clave

Animación a la lectura, Hábito lector, Club de lectura, Lectura dialógica, Estudio de caso, Escuela primaria, Enseñanza primaria.

Title: Contributions of a school book club to encourage reading for pleasure

Abstract

Currently one of the main challenges in the work of promoting reading in schools is the cultivation of children's enjoyment of reading. We present a unique case study in a school reading club in Spain that has managed to achieve that goal and has received a very positive assessment from participants. The reasons for its success mainly reside in four key areas: planning, dialogic reading, critical thinking and values education.

Keywords

Encourage to reading, Reading habit, Book club, Reading club, Dialogic reading, Case study, Children, Primary school.

Álvarez-Álvarez, Carmen; Pascual-Díez, Julián (2014). "Aportaciones de un club de lectura escolar a la lectura por placer". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 625-631.

<http://dx.doi.org/10.3145/epi.2014.nov.10>

1. Introducción

La educación lectora no sólo trata de formar personas para que sepan leer cada vez mejor, sino también para que, en lo posible, quieran leer, estén motivadas hacia la lectura y puedan integrarla en su vida. A pesar de las interesantes propuestas

didácticas, materiales formativos, portales digitales, cursos, seminarios, convocatorias de premios y publicaciones de buenas prácticas que se han elaborado durante los últimos años a nivel autonómico y estatal, es mucho el camino que queda por recorrer tanto en políticas educativas como en el ámbito de la investigación e innovación en los centros educativos.

Artículo recibido el 30-07-2014
Aceptación definitiva: 06-10-2014

Las actividades literarias más tradicionales plantean tareas posteriores a la lectura habitualmente orientadas al análisis formal y estructural de los textos con preguntas de comprensión no siempre significativas, sin contemplar el comentario de las emociones o actitudes que suscitan los textos.

Uno de los aspectos de las bibliotecas escolares más valorados por la comunidad educativa es su capacidad para estimular el gusto por la lectura (Baró; Mañà, 2013). Uno de los medios más adecuados para estimular la lectura por placer y las actividades de aprendizaje comunicativas son los clubs de lectura escolares (Arana-Palacios; Galindo-Lizaldre, 2009; Jiménez-Fernández; Cremades-García, 2013). Un club de lectura es “una obra en movimiento que se recorre de la mano de un libro, un lector y un conductor” (Carreño, 2012, p. 31); supone “una ocasión de leer con los demás, un deseo de compartir lecturas, un modo dialogal de extender la mirada y afinar los oídos” (Mata, 2008, p. 82). La interacción del grupo y la escucha a los compañeros favorece la implicación personal de los participantes y discusiones que ayudan a conectar la literatura con la vida (Gritter, 2011). Los clubs de lectura escolares pueden favorecer una lectura crítica (Atwell, 2007; Cassany, 2008) y contribuir al desarrollo de niveles de alfabetización superior (Olson, 2009). Además, pueden potenciar la función transformadora que la biblioteca debería desempeñar en la educación, al posibilitar la interacción y el aprendizaje dialógico (Gómez-Hernández, 2013). A pesar de ello, la mayor parte de los centros educativos españoles carecen de clubs de lectura.

Las sesiones de los clubs de lectura escolares siguen unas formas habituales: se selecciona un libro, se determina un tiempo para su lectura y se fija una reunión posterior para comentarlo

Las sesiones de los clubs de lectura escolares siguen unas formas habituales: se selecciona un libro, se determina un tiempo para su lectura y se fija una reunión posterior para comentarlo. En estas reuniones suelen abordarse pasajes destacados del libro, se cuestionan comportamientos de los protagonistas, se plantean experiencias personales y se genera un intercambio rico en pensamientos y reflexiones.

La formación de niños lectores puede investigarse siguiendo numerosos enfoques metodológicos. En este artículo se presentan los resultados del estudio de caso de un club de lectura escolar en el marco de una investigación-acción. La relevancia de la investigación-acción radica en su interés por conocer y entender los problemas cotidianos reales que viven los docentes, con el objetivo de analizar la propia práctica para comprenderla y transformarla, favoreciendo el desarrollo profesional del profesorado, que se va emancipando a medida que va llevando a efecto estos procesos (Whitehead, 2009).

El objetivo general de la investigación es analizar la contribución de un club de lectura, organizado en la biblioteca del centro, al desarrollo de la lectura por placer del alumnado. La hipótesis de partida es que ello puede ocurrir como consecuencia de la interrelación que se establece entre las ideas del lector, las ideas que transmite la lectura de la obra en solitario y las ideas de los compañeros expuestas en las tertulias que se organizan.

2. Método

2.1. Participantes y lecturas

En este estudio de caso participó un grupo de 24 estudiantes de tercer ciclo de educación primaria (10 niños y 15 niñas de 11-12 años y clase social media-baja), que pertenecían al club de lectura escolar de un colegio público de educación infantil y primaria de Asturias, durante el curso 2009-2010.

Los coordinadores fueron una profesora universitaria de la *Facultad de Pedagogía* externa al centro (que ya había participado en dos ediciones anteriores del club de lectura) y un alumno de esa misma *Facultad*. Ellos seleccionaron cinco obras narrativas, motivados por su carácter eminentemente dialógico. Alzola (2007, p. 155), entre otros, defiende estas lecturas porque la narración “incorpora muchas voces diferentes, que hablan unas con otras y con otras voces fuera del texto, los discursos de la cultura y de la sociedad en general”. Seleccionaron los libros tomando como criterio su calidad literaria, el interés de sus temas, su valor formativo, la capacidad lectora de los alumnos y otros criterios asumidos, en general, por la comunidad escolar y por la didáctica de la literatura: presencia de obras escritas por hombres y mujeres, autores nacionales y extranjeros, contemporáneos y antiguos; obras infantiles y juveniles actuales, clásicos de

Tabla 1. Obras leídas en el club de lectura

Obra	Síntesis
<i>El no de Marina</i> (Carmen Olaechea)	Marina es una niña muy tímida que va de vacaciones a un pueblo pequeño donde conoce a Andrés, un señor mucho más tímido que ella que, sin pretenderlo, le ayudará a vencer su problema, liberándose finalmente los dos del mismo.
<i>¡No es tan fácil ser niño!</i> (Pilar Lozano Carbayo)	Fernando es cambiado de clase a principios de curso y, en la nueva, Pedrito y sus seguidores le acosan incesantemente por el tamaño de sus orejas. Poco a poco, con la ayuda de nuevos amigos, logra superar esta situación.
<i>Don Quijote</i> (versión adaptada de Vicens-Vives)	El libro recorre con relativa fidelidad los principales pasajes que aparecen en <i>El Quijote</i> original.
<i>Luna de Senegal</i> (Agustín Fernández Paz)	Khoedi, una niña senegalesa, viaja a España a reencontrarse con su padre, de quien luego se descubre que ha sido inmigrante ilegal y las penurias por las que pasó.
<i>Las brujas</i> (Roald Dahl)	El niño protagonista de la historia y su abuela luchan contra todas las temidas brujas del mundo, reunidas en un hotel para celebrar su convención anual, donde traman acabar con todos los niños del planeta.

la literatura universal adaptados a la competencia lectora de los alumnos y clásicos internacionales de la literatura infantil y juvenil.

2.2. Instrumentos

En el proceso de investigación se generaron diversos documentos, que constituyeron la base para el análisis de datos:

- Grabaciones en audio y en vídeo de las reuniones.
- Transcripciones de las sesiones y notas de campo elaboradas durante y después de cada sesión.
- El blog del club de lectura, en el que se incluyó un apartado para comentar los libros.
- Memoria exhaustiva de la experiencia, que incluye una revisión sistemática de las sesiones y de las fases del proceso de trabajo: planificación, acción, reflexión y metarreflexión.
- Aplicación de un cuestionario de satisfacción al alumnado al final del curso.
- Entrevistas semi-estructuradas realizadas a los participantes, a sus familias y al profesorado.

También se contó con observadores externos puntuales que reflexionaron sobre el diseño y evolución de la actividad. Con estos elementos se pudo obtener el máximo caudal de información y triangular y saturar los datos registrados en los diferentes soportes.

Procedimiento

El presente estudio se ha planteado como un estudio de caso en el que se ha llevado a cabo un proceso de investigación-acción. El objetivo último de los estudios de caso es conocer una unidad de análisis en profundidad con el interés de ofrecer ideas que contribuyan a mejorar esa realidad local y tomar decisiones informadas (Stake, 2005) que, a su vez, pueden ser de utilidad para otros casos más o menos afines (Gill, 2011). Para ello se elige una muestra pequeña, pero significativa y se profundiza en ella mediante técnicas de recogida de información variadas, que garanticen el rigor.

Se siguió un proceso de trabajo que constó de cuatro situaciones:

- planificación (previa a cada reunión);
- acción (durante la sesión);
- reflexión (durante y después de la reunión);
- metarreflexión (después de la reunión).

Análisis

Con posterioridad a la recogida de información, se procedió a organizar y analizar los datos, siguiendo pautas propias del análisis de contenido para los datos cualitativos y calculando frecuencias, porcentajes, puntuaciones medias y desviaciones típicas para los datos cuantitativos. La mayor parte de los datos recogidos son de tipo cualitativo, dada la necesidad de conocer el punto de vista de los participantes respecto a su experiencia como miembros del club.

Los resultados se han organizado sobre dos ejes: el modo de llevar a cabo las reuniones y el contenido de las mismas. En el primer caso, se han distinguido dos dimensiones de análisis: la planificación y desarrollo de las sesiones y la lectura dialógica; es decir, la estrategia didáctica seguida en



Figura 1. Proceso seguido en la sistematización de la investigación-acción

las tertulias literarias. Desde el punto de vista formativo o del contenido, se han planteado dos dimensiones más: la formación crítica, que se produce en la tertulia-debate y la educación en valores, que se trata de promover a través de las lecturas seleccionadas y su comentario oral.

3. Resultados

Los resultados de la investigación se presentan a continuación articulados en torno a las cuatro categorías anteriormente citadas.

3.1. Planificación y desarrollo de las sesiones

Una de las claves para que un club de lectura pueda cumplir con sus objetivos es contar con una buena planificación, a diferentes niveles. Varios son los aspectos destacados por los participantes que contribuyen al éxito del club y a facilitar los procesos de reflexión-acción:

- preparación minuciosa de cada reunión por parte de los coordinadores;
- estabilidad en la coordinación del club y la buena formación de sus responsables;
- adecuada selección de libros;
- existencia de normas de funcionamiento claras y asumidas por todos;
- creación de un clima que estimula la participación;
- utilización de un blog que permite hacer comentarios personales de las obras.

3.2. Lectura dialógica

Desde el punto de vista formal, un aspecto especialmente significativo es la realización de prácticas de lectura dialógica. Éstas suponen ir más allá de las prácticas tradicionales de lectura íntima y solitaria que merman la posibilidad de

Tabla 2. Ejes y dimensiones de análisis de los datos

Ejes	Dimensiones
Forma	-Planificación y desarrollo de las sesiones -Lectura dialógica
Contenido	-Formación crítica -Educación en valores



- "A mí los libros me gustan más sin imágenes, para imaginármelo"

discutir sobre las obras con otras personas. Así lo han expresado algunos alumnos al responder qué es lo que más les gusta del club¹: "Los comentarios, porque hablamos de los libros" (O2,11,C); "Las reuniones, porque me parecen interesantes" (O9,11,C). Los casos de lectura dialógica son muy numerosos y están presentes en las tres fases principales de las sesiones.

Primera fase

Tras la apertura de la sesión por parte de los coordinadores, se daba la palabra al alumnado, que opinaba libremente sobre la obra leída: calidad literaria, agrado personal, ilustraciones, etc. Después de un primer momento en el que intervenían varios participantes ofreciendo sus reflexiones, los coordinadores animaban al resto a que expusieran también su valoración de la obra.

La lectura dialógica favorece el desarrollo del pensamiento propio y del pensamiento divergente del alumnado

Segunda fase

Más tarde se pedía a los participantes que comentasen los momentos de la obra que les resultaron más interesantes o más llamativos, argumentando el por qué. Esto es muy importante, porque propicia una postura activa y curiosa en el lector:

Alumno: -Me causó mucha intriga cuando el protagonista estaba detrás del biombo y no sabía si lo iban a pillar, no se sabía si les iba a oler mal a las brujas...

[Risas]

Coordinadora: -¿Tú querías que lo pillasen?

Alumno: -No.

Coordinadora: -¿Por qué no querías que lo pillasen?

Alumno: -Porque me estaba gustando cómo era él y con su abuela y pensaba que si lo pillaban y lo convertían en ratón igual lo mataban o algo, o su abuela no lo podía distinguir.

En este caso, con la lectura de *Las brujas*, un miembro del club pone de relieve el interés que le generó un momento de tensión del libro: tras identificarse con el protagonista, teme por la vida del personaje en una situación tensa.

Tercera fase

En un momento posterior se pedía a los niños que expusieran las enseñanzas para la vida que aportaba el libro. Es posible destacar gran número de situaciones ricas en formación en valores y en educación crítica, pero, como guardan estrecha relación con el contenido de las lecturas, serán abordadas posteriormente.

3.3. Formación crítica

La lectura dialógica favorece el desarrollo del pensamiento propio y del pensamiento divergente del alumnado. El primero aparece en aquellas intervenciones de los niños en las que exponen libremente su opinión y la justifican a partir de la lectura, su experiencia, sus ideas o conocimientos previos. El pensamiento divergente se produce cuando los niños realizan reflexiones estimuladas por las aportaciones de otros compañeros y toman postura a partir de ellas.

Pensamiento propio. En este caso se han encontrado dos situaciones diferentes:

a) Expresión de la propia opinión y comprensión autónoma: el alumnado manifiesta libremente lo que piensa. Este ejemplo fue tomado de *El no de Marina*.

Alumno: -A mí los libros me gustan más sin imágenes, para imaginármelo. Y en este libro, la imagen que más me gusta es cuando está Andrés cantando y Marina lo ve.

b) Toma de conciencia: el alumno se apropia de una idea manifestada en la obra para hacerla suya; en este caso, es ofrecida por la coordinadora en el análisis del libro *¡No es tan fácil ser niño!*

Coordinadora: -¿Por qué crees que el autor habla de una paliza?

Alumno: -Para que los niños vieran que por tener un defecto no hay que pegar a nadie.

Coordinadora: -Puede ser. Yo creo que el autor lo hace con otra intención, que es la de decir que peleando no se solucionan las cosas. ¿Os disteis cuenta de ese mensaje que trata de transmitir el libro? (...)

Varios alumnos: -Sí.

Pensamiento divergente. Dentro de esta categoría, se detectan dos situaciones distintas:

a) Construcción de un discurso alternativo: el alumno plantea un discurso alternativo al que se recoge en la obra o al manifestado por los compañeros o coordinadores, como en este caso, con *El no de Marina*.

Alumna: -El tema del libro también puede ser la apariencia que tiene Andrés de fuerte y cuando está con Marina creen que no puede ser tan tímido, que puede ser agresivo, pero, en cambio, es muy tímido y cuida bien de su hija y ella es más valiente siendo más pequeña; él tiene más fuerza física, pero es más valiente Marina.

Coordinador: -Las apariencias engañan, ¿no?

Alumna: -Sí.

b) Defensa de posiciones personales: el alumno argumenta sobre una postura personal que ha creado a partir de la lectura o del debate. Este ejemplo está tomado de *Las brujas*.

Coordinadora: -Cosas que criticarías al libro.

Alumna: -A mí no me gustaron las ilustraciones (...). Me parecieron muy infantiles.

En todos estos casos se aprecia que al alumnado se le da libertad para pensar por sí mismo, generando un pensa-

miento crítico. Esto ha sido valorado muy positivamente por todos los participantes: -“Lo que más me gusta del club es que nos dejan expresar con libertad lo que nos parecen las cosas” (A4,12,C).

3.4. Educación en valores

En un club de lectura confluyen los valores que tienen los participantes con los que se encuentran implícitos y explícitos en las obras. Este contraste genera un crecimiento axiológico (valores) en el niño, que se ve nuevamente influenciado por los testimonios de los compañeros y coordinadores. Es algo que el alumnado aprecia y valora en las reuniones, y con lo que, además, aprende: -“Yo con ese libro entendí que me tenía que comportar mejor con todos mis compañeros de clase y del colegio” (O14,11,E).

A continuación reproducimos un momento de debate en una reunión que demuestra el interés de los estudiantes por las dimensiones éticas de las lecturas.

Alumna: -Cuando los chicos esos vienen y destrozan el jardín, están destrozándolo todo y Marina dice “no”.

Coordinadora: -¿Por qué ella grita “no”? ¿Qué sentimiento es el que tiene que le hace gritar “no” y superar su timidez? O por lo menos en ese momento parece que supera su timidez. ¿Cómo se llama ese sentimiento? ¿Cómo se siente ella?

Alumno: -Yo creo que se siente enfadada porque le da pena de Andrés, porque le están destrozando el jardín y él está ahí en la ventana; yo creo que está llorando y esperando a que se fueran.

[Tratan de acertar un adjetivo para ese pensamiento, hasta que una alumna lo dice].

Alumna: -Impotente.

En este momento del comentario del libro *El no de Marina*, se percibe el interés y el sentimiento de rechazo que ha despertado en una alumna un acto vandálico llevado a

cabo por unos adolescentes. En todas las obras leídas se han producido situaciones similares que han permitido intensos debates sobre lo que está bien o no, generando una red de interacciones en las que se funde la formación transversal y la literaria y lingüística. La tabla 3 recoge aspectos relacionados con la educación en valores abordados en los diferentes coloquios.

El club de lectura desde la biblioteca escolar puede contribuir al desarrollo de la lectura por placer y favorecer el pensamiento crítico ante la lectura y la educación en valores

4. Discusión y conclusiones

Los resultados obtenidos apoyan la hipótesis planteada. La implementación de un club de lectura desde la biblioteca escolar puede contribuir al desarrollo de la lectura por placer y, al mismo tiempo, puede favorecer el pensamiento crítico ante la lectura y la educación en valores. Aunque es cierto que en el proceso de lectura individual los niños se suelen identificar con los protagonistas de las aventuras literarias (Patte, 2008), en el proceso de lectura colectivo, en la tertulia, tienen la oportunidad de tomar la palabra y aportar sus valores a la lectura (Chambers, 2007). Así se educan el sentido estético, la dimensión ética y las emociones (Camps, 2011). El siguiente testimonio de una participante puede ser suficientemente ilustrativo: -“Me gusta mucho venir al club de lectura porque ahora leo más y me gusta leer. Antes leía con mi madre, pero no me gustaba” (A15,12,E).

Tabla 3. Aspectos de educación en valores abordados

Obra	Aspectos abordados
<i>El no de Marina</i>	-La timidez y cómo la superan los protagonistas -La dificultad para expresar los sentimientos y lo positivo que resulta hacerlo -La impotencia ante situaciones que se consideran injustas -Las apariencias pueden engañar -La capacidad de cambio de las personas y el potencial de la autocrítica para la mejora -El amor y las buenas relaciones en familia
<i>¡No es tan fácil ser niño!</i>	-El sufrimiento de un niño acosado y sus sensaciones y preocupaciones -Papel de los acosadores en los grupos y el de los que observan y animan cómplicemente estos comportamientos -Cómo se puede prevenir y corregir el maltrato en la escuela -Papel de los educadores ante situaciones de acoso -Papel de la violencia en la resolución de conflictos -El respeto y la convivencia como elementos de cohesión social
<i>Don Quijote</i>	-La obsesión como algo negativo -La locura y sus problemas -Uso de expresiones malsonantes en libros de literatura infantil -El valor de la perseverancia
<i>Luna de Senegal</i>	-La dureza de la inmigración, legal e ilegal -La necesidad de las personas de trabajar y vivir con dignidad -Interculturalidad -Autodisciplina para leer un libro cuando no resulta agradable -La capacidad de reflexión de las personas como cualidad que se puede cultivar -La importancia de ayudar a las personas vulnerables -El inigualable cariño que proporcionan los abuelos
<i>Las brujas</i>	-El amor en familia, sea cual sea su forma -Los caprichos de los niños hoy y cómo debe actuarse ante ellos -El engaño como algo negativo -La importancia de seguir siendo lectores en la etapa de educación secundaria



La discusión en grupo estimula la participación de los alumnos

La interrelación de las cuatro dimensiones que articulan los resultados de la investigación puede constituir una base adecuada para construir una práctica exitosa de animación a la lectura, que fomenta el gusto lector por las razones siguientes:

En primer lugar, la planificación y estructuración minuciosa de las sesiones, siguiendo procesos de investigación-acción, permite configurar un clima en el grupo que estimula la participación de los alumnos, la selección adecuada de los textos y la formación constante de los coordinadores de la actividad. Estos aspectos han sido elementos clave para favorecer la asistencia y la implicación creciente de los participantes en una actividad de carácter voluntario. Por su parte, el blog ha permitido a los miembros del club interactuar entre sí compartiendo su experiencia lectora: comentando impresiones, destacando pasajes de la obra, emitiendo valoraciones, etc. Esta tecnología, puesta al servicio del club de lectura, es una herramienta complementaria a las tertulias muy valiosa. En otros clubs de lectura, el empleo de medios tecnológicos puede ser un elemento determinante, como muestra el estudio de la *Fundación Germán Sánchez Ruipérez* (2012) con el uso de libros electrónicos o el estudio de *Manso-Rodríguez* (2012) con redes sociales. En el caso analizado, el blog ha jugado un papel complementario.

Los participantes han crecido axiológicamente, formulando juicios respecto a los valores de las obras y sus implicaciones sociales

En segundo lugar, se constata el gran potencial educativo del enfoque dialógico para favorecer la formación crítica de lectores y la lectura por placer. Aunque es posible que el alumnado pueda desarrollar individualmente y por escrito ciertas habilidades de lectura crítica, tiene un enorme interés la práctica de estrategias didácticas innovadoras asentadas sobre el diálogo crítico como método de trabajo (*Chambers*, 2007; *Fecho*; *Coombs*; *McAuley*, 2012), la cooperación y la interacción entre el alumnado y los coordinadores. Además, desde el punto de vista lingüístico-literario, una tertulia literaria dialógica permite acercar al alumnado a una actividad de ocio como la lectura que, en muchos ca-

sos, resulta novedosa. Esto tiene un valor especial cuando los estudiantes están privados de un ambiente motivador por razones socio-culturales, como es este caso, y cuando se constata que las prácticas dialógicas son las que tienen una mayor influencia transformadora en los participantes (*Robinson*, 2011) y que, habitualmente, las prácticas escolares no responden a este planteamiento (*Reznitskaya*, 2012).

En tercer lugar, se comprueba que dar la opinión sobre una lectura y debatir sobre la misma favorece la comprensión de una obra literaria, en el sentido en el que ésta se define en el estudio de la *Fundación Germán Sánchez Ruipérez* anteriormente citado: comprender las tramas, profundizar en los personajes y temas y enriquecer o ampliar conscientemente la lectura conectando los contenidos del relato con temas y competencias específicos suscitados por él. Estos tres niveles de lectura se potencian en las sesiones del club y favorecen el desarrollo del pensamiento crítico: los estudiantes pueden ponerse en el lugar de los personajes, criticar aspectos de un texto o una parte de la lectura con argumentos que están en un proceso de análisis y de reelaboración.

En cuarto lugar, se constata que los participantes han crecido axiológicamente (en valores), formulando juicios respecto a los valores de las obras y sus implicaciones sociales. Esto es importante ya que la reflexión y el debate sobre los valores implícitos y explícitos de los relatos no constituye algo complementario a la lectura, sino que, por un lado, la “enriquece o amplía” y, por otro, estimula la implicación e interés hacia la misma y hacia las interpretaciones de los compañeros y personas adultas (*Blum et al.*, 2010).

Además de los aspectos reseñados, hay que señalar la contribución positiva del club de lectura escolar a la mejora de la competencia lectora, tal y como pusieron de manifiesto los profesores del alumnado participante y sus familias en los cuestionarios aplicados. Finalmente, hay que destacar su aportación al desarrollo de la educación literaria, que supone una intervención en campos estrechamente vinculados: ayudar al alumnado a descubrir la lectura como experiencia satisfactoria, aprender a construir el sentido del texto y familiarizarse con sus características discursivas, textuales y lingüísticas (*Zayas*, 2011).

Notas

1. Los testimonios del alumnado permiten identificar el sexo del participante (O para alumnos; A para alumnas), su identidad (1, 2, 3...), edad (11, 12...) y la técnica empleada (O: observación; E: entrevista; C: cuestionario).

5. Bibliografía

Alzola, Nerea (2007). “Literatura infantil y educación ética: análisis de un libro”. *Revista de psicodidáctica*, v. 12, n. 1, pp. 153-166. <http://www.ehu.es/ojs/index.php/psicodidactica/article/view/215/211>

Arana-Palacios, Jesús; Galindo-Lizaldre, Belén (2009). *Leer y conversar. Una introducción a los clubes de lectura*. Gijón: Trea. ISBN: 978 84 9704 449 3

Atwell, Nancie (2007). *The reading zone: how to help kids become skilled, passionate, habitual, critical readers*. Nueva York: Scholastic. ISBN: 978 0 439 92644 7

<http://gec.kmu.edu.tw/~lc/ecorner/eBook/The%20Reading%20Zone.pdf>

Baró, Mònica; Mañà, Teresa (2013). "Bibliotecas escolares: ¿un valor en alza?". *El profesional de la información*, v. 22, n. 2, pp. 106-112.

<http://bd.ub.edu/pub/mana/sites/bd.ub.edu.pub.mana/files/EPI,106-112.pdf>

<http://dx.doi.org/10.3145/epi.2013.mar.02>

Blum, Irene H.; Koskinen, Patricia S.; Bhartiya, Purnima; Hluboky, Sandra (2010). "Thinking and talking about books: using prompts to stimulate discussion". *The reading teacher*, v. 63, n. 6, pp. 495-499.

<http://dx.doi.org/10.1598/RT.63.6.6>

Camps, Victoria (2011). *El gobierno de las emociones*. Barcelona: Herder. ISBN: 978 8 42542 744 2

Carreño, Óscar (2012). *Clubes de lectura*. Barcelona: editorial UOC, colección El profesional de la información, n. 11. ISBN: 978 84 9029 238 9

Cassany, Daniel (2008). "Llegir críticament, all llarg del currículum. Presentació". *Temps d'educació*, n. 34, pp. 7-10.

<http://www.publicacions.ub.edu/refs/indices/06977.pdf>

Chambers, Aidan (2007). *Dime*. México: FCE. ISBN: 978 96 8168 4532

Fecho, Bob; Coombs, Dewan; McAuley, Sean (2012). "Reclaiming literacy classrooms through critical dialogue". *Journal of adolescent & adult literacy*, v. 55, n. 6, pp. 476-482.

<http://dx.doi.org/10.1002/JAAL.00057>

Fundación Germán Sánchez Ruipérez (2012). *Ebook-18: Los lectores niños y jóvenes y los libros electrónicos*. Peñaranda de Bracamonte: Centro de Desarrollo Sociocultural.

http://www.territorioebook.com/recursos/vozelectores/menos18/informe_lectores_menos_18.pdf

Gómez-Hernández, José-Antonio (2013). "Unir educación y bibliotecas: la evolución de un reto permanente". *El profesional de la información*, v. 22, n. 2, pp. 101-105.

<http://www.elprofesionaldelainformacion.com/contenidos/2013/marzo/01.pdf>

<http://dx.doi.org/10.3145/epi.2013.mar.01>

Gill, T. Grandon (2011). *Book informing with the case method*.

Londres: Informing Science Press. ISBN: 978 1 932886 44 3

<http://grandon.com/publications/CaseMethod.pdf>

Gritter, Kristine (2011). "Promoting lively literature discus-

sion". *The reading teacher*, v. 64, n. 6, pp. 445-449.

<https://crmsliteracy.wikispaces.com/file/view/Promoting+Lively+Literacy+Discussion.pdf>

<http://dx.doi.org/10.1598/RT.64.6.7>

Jiménez-Fernández, Concepción; Cremades-García, Raúl

(2013). *Bibliotecas escolares*. Barcelona: EPI-UOC, n. 22. ISBN: 978 84 029 886 2

Manso-Rodríguez, Ramón-Alberto (2012). "Bibliotecas: fomento de la lectura y redes sociales: convirtamos amigos en lectores". *El profesional de la información*, v. 21, n. 4, pp. 401-405.

<http://eprints.rclis.org/17310/1/401-405.pdf>

<http://dx.doi.org/10.3145/epi.2012.jul.12>

Mata, Juan (2008). *Animación a la lectura. Hacer de la lectura una práctica feliz, trascendente y deseable*. Barcelona: Graó. ISBN 978 84 7 827681 3

Olson, David (2009). "Education and literacy". *Infancia y aprendizaje*, v. 32, n. 2, pp. 141-151.

<http://dx.doi.org/10.1174/021037009788001824>

Patte, Geneviève (2008). *Déjenlos leer. Los niños y las bibliotecas*. México: FCE. ISBN: 978 96 81676698

Robinson, Andrew (2011). "Thinking better, whatever one thinks: Dialogue, monologue and critical literacy in education". *Critical literacy: Theories and practices*, v. 6, n. 1, pp. 21-35.

<http://goo.gl/Z344Oz>

Reznitskaya, Alina (2012). "Dialogic teaching. Rethinking language use during literature discussions". *The reading teacher*, v. 65, n. 7, pp. 446-456.

http://www.montclair.edu/profilepages/media/370/user/2012-Dialogic_Teaching-RT.pdf

<http://dx.doi.org/10.1002/TRTR.01066>

Stake, Robert E. (2005). *Investigación con estudio de casos*. Madrid: Morata. ISBN: 978 84 7112 422 7

Whitehead, Jack (2009). "Generating living theory and understanding in action research studies". *Action research*, v. 7, n. 1, pp. 85-99.

<http://www.actionresearch.net/writings/jack/jwartheory0309.pdf>

<http://dx.doi.org/10.1177/1476750308099599>

Zayas, Felipe (2011). *La educación literaria. Cuatro secuencias didácticas*. Barcelona: Octaedro. ISBN: 978 84 9921 329 3



El profesional de la información tiene el sello de revista de calidad otorgado por la **Fecyt (Fundación Española para la Ciencia y la Tecnología)**, y está reconocida por la **Aeres (Agence d'évaluation de la recherche et de l'enseignement supérieur)** de Francia.

Suscríbete a

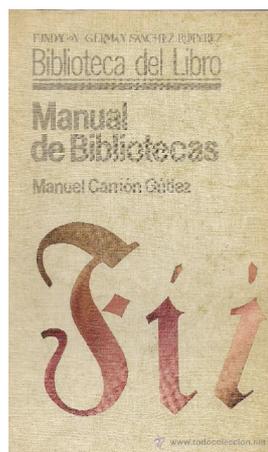
Mi Biblioteca

La revista de mundo bibliotecario

Ofrecemos la suscripción gratuita por un año a la revista *Mi Biblioteca* a todas aquellas bibliotecas que ofrezcan a sus usuarios y amigos la posibilidad de ser socios de la Fundación Alonso Quijano (entidad sin ánimo de lucro editora de la revista) y consigan al menos 5 socios.

Mi Biblio

Y recibirás de modo gratuito este manual de referencia para bibliotecarios



Válido para todas las suscripciones NUEVAS en 2014

www.mibiblioteca.org
952235405

ANÁLISIS



DATOS ENLAZADOS DE PUBLICACIONES, PROYECTOS Y HERRAMIENTAS PARA INVESTIGADORES EN HUMANIDADES DIGITALES: CATÁLOGO PILOTO DEL CENTRO CLARIN IULA-UPF



Silvia Arano-Poggi y Núria Bel



Silvia Arano-Poggi es técnica de investigación en el *Proyecto Clarin/Feder* en la *Universitat Pompeu Fabra (UPF)*, y colaboradora del grupo de investigación *Tecnologías de los recursos lingüísticos*. Sus líneas de investigación son organización del conocimiento, comportamiento informacional, aprendizaje virtual, y humanidades digitales. También es colaboradora docente en la *UOC*.
<http://orcid.org/0000-0001-9436-3301>

silvia.arano@upf.edu



Núria Bel es profesora agregada en la *Universitat Pompeu Fabra (UPF)* y dirige el grupo de investigación *Tecnologías de los recursos lingüísticos*. Sus líneas de investigación son las tecnologías lingüísticas, especialmente la generación de recursos para tecnologías como la traducción automática, extracción y minería de información. Ha participado en proyectos financiados por la UE como: *Panacea* (2010-2012, 7FP-ITC-248064) y *Metanet4u* (2010-2013, CIP-PSP-270893), este último proyecto de la red de excelencia europea *Multilingual Technology Alliance, Meta-net*.
<http://orcid.org/0000-0001-9346-7803>

nuria.bel@upf.edu

*Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada
Roc Boronat, 138. 08018 Barcelona, España*

Resumen

Los investigadores en Humanidades Digitales tienen dificultades para acceder y utilizar herramientas informáticas que les asistan en la explotación de los textos objeto de sus estudios. Esto se debe al hecho de que, en la mayoría de los portales y registros especializados, la información sobre dichas herramientas no está enlazada con la información sobre dónde encontrarlas y cómo utilizarlas. El *Centro de Competencias Clarin IULA-UPF* compila e interrelaciona la información necesaria en un catálogo de datos enlazados para ofrecer a los investigadores una forma integrada de acceder a toda la información. En este artículo se presentan los detalles del diseño y de la selección de materiales e instrumentos de descripción utilizados en la elaboración de dicho catálogo.

Palabras clave

Humanidades digitales, Catálogos, Centros de competencias, Datos enlazados, Servicios web, *Clarin IULA-UPF*.

Title: Linked data of publications, projects and computational tools for digital humanities researchers: the pilot catalogue of the *Clarin IULA-UPF* Center

Abstract

Digital humanities researchers experience some difficulties in accessing and using computational tools which are meant to assist them when carrying out queries of texts analysis in their studies. This is due to the fact that information concerning tools in this field and their effective usage is not linked to the information about their location and utilization. *Clarin IULA-UPF Center* aims at compiling and relating required information in a catalogue of linked data in order to offer researchers an integrated way of accessing it. This article presents details about the design and selection of the materials and description tools included in the catalogue.

Keywords

Digital humanities, Catalogues, Competence centres, Linked data, Web services, *Clarin IULA-UPF*.

Artículo recibido el 09-05-2014

Aceptación definitiva: 28-08-2014

Arano-Poggi, Silvia; Bel, Núria (2014). "Datos enlazados de publicaciones, proyectos y herramientas para investigadores en humanidades digitales: catálogo piloto del centro *Clarín IULA-UPF*". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 633-642.

<http://dx.doi.org/10.3145/epi.2014.nov.11>

1. Introducción

Las humanidades digitales (DH) han aumentado el interés por la utilización de herramientas informáticas para el análisis y explotación de textos, y en particular por las tecnologías de la lengua (Juola, 2008; Schreibman; Hanlon, 2010; Brier; Hopp, 2011; Wiedemann, 2013). Sin embargo el acceso a información sobre la disponibilidad y uso de estas aplicaciones puede resultar arduo para los investigadores del área. Las investigaciones están accesibles en bases de datos de artículos científicos o proyectos, portales de asociaciones profesionales, etc. (como *Scopus*, *Cordis*, *Association for Computational Linguistic - Anthology*); pero las descripciones de las aplicaciones se difunden en portales y registros especializados (por ejemplo *Bamboo DiRT - Digital Research Tools*; *Text Analysis Portal for Research: TAPoR 2.0*; *Virtual Language Observatory*). Esto es, no hay relación entre la documentación relativa a la aplicación de las herramientas a diferentes objetos de estudio y la documentación de las mismas.

<http://www.elsevier.com/online-tools/scopus>

http://cordis.europa.eu/fp7/projects_es.html

<http://www.aclweb.org/anthology>

<http://dirtdirectory.org>

<http://www.tapor.ca>

<http://catalog.clarin.eu/vlo>

En el marco de la iniciativa europea *Clarín*¹ que pretende facilitar el uso de aplicaciones informáticas en la investigación en humanidades y ciencias sociales (HSC), y gracias a financiación *Feder*² y *Universidad Pompeu Fabra*, se puso en marcha en España el *Centro de competencias Clarín IULA-UPF*³. Su misión es dinamizar el uso de las tecnologías de la lengua en la investigación en humanidades, y asesorar a los investigadores en materia de acceso a datos y aplicaciones informáticas para llevar a cabo investigación experimental, tanto a nivel académico como industrial, que tenga como objeto la extracción de datos en textos en cualquier lengua.

Las HD han aumentado el interés en la utilización de herramientas informáticas para el análisis y explotación de textos, en particular de las aplicaciones de las tecnologías de la lengua

El centro ofrece acceso a herramientas útiles para el análisis y la explotación de textos en forma de servicios web. La descripción de las mismas se relaciona, gracias a la tecnología de datos enlazados abiertos (*linked open data*, LOD), con información y enlaces a publicaciones y proyectos.

<http://www.clarin-es-lab.org>

Como forma de contribuir a la colaboración interdisciplinaria y a la reutilización y difusión de los resultados de investigación, se elaboró un catálogo con valor añadido para potenciales usuarios de las DH. Se ha considerado valor añadido no sólo la interrelación de información de publicaciones, proyectos y herramientas, sino también la puesta en práctica de una política singular de selección para el crecimiento del catálogo. La política, que tiene en consideración a los posibles usuarios, se basa en dos ideas fundamentales:

a) promover el razonamiento en casos análogos: que como resultado de la navegación en el catálogo, los usuarios puedan reproducir, emular o adaptar las experiencias relatadas en artículos, comunicaciones o proyectos con diferentes textos, similares herramientas o en otras disciplinas;

b) enfatizar el carácter "operativo" de la información: el concepto de información operativa (*actionable information*⁴), que se toma prestado de la gestión del conocimiento, hace referencia al uso dado a la información. La información operativa se presenta en un formato fácilmente interpretable que busca despertar la capacidad exploratoria del usuario con la finalidad de que utilice o no dicha información para la toma de decisiones.

Una vez incorporados los nuevos materiales seleccionados bajo esta política, el usuario será capaz de construir a partir de la navegación en el catálogo y gracias a su estructuración utilizando recursos de la web semántica, su propio itinerario para, por ejemplo, descubrir cómo usan otros una aplicación informática a partir de la lectura de publicaciones, y experimentar con una herramienta similar.

Por último se destaca el carácter experimental del proyecto. La versión del catálogo presenta unos contenidos y una estructura que no son definitivos, ya que está en período de pruebas por parte de usuarios potenciales⁵. El catálogo piloto muestra el trabajo de modelado de datos y la vinculación de información relativa a una de las posibles tareas incluida en los procesos de análisis de texto: el reconocimiento de entidades.

2. Metodología

Para la elaboración del catálogo piloto se han seleccionado artículos y descripciones de proyectos, sobre las aplicaciones informáticas más utilizadas en la investigación actual en DH, y quién y cómo las han utilizado. Durante la búsqueda y análisis preliminar de información se constata que en general los proyectos y las publicaciones refieren cómo unas herramientas particulares ayudan a los investigadores a identificar, anotar, extraer y visualizar la información requerida para un caso de estudio. La tarea más frecuente que desempeñan dichos productos es el reconocimiento de entidades (*named entity recognition*, NER): donde se identifican y clasifican segmentos de texto que son nombres de per-

sonas, organizaciones o instituciones, lugares geográficos, expresiones temporales, cantidades, valores monetarios, etc. Si bien es una tarea básica dentro del procesamiento del lenguaje natural, es recurrente en estudios que explotan *corpora* digitales (revistas, periódicos, blogs, tuits, libros, etc.) en diversas disciplinas de las HSC y las DH.

http://es.wikipedia.org/w/index.php?title=Reconocimiento_de_nombres_de_entidades&oldid=65336440

Para la búsqueda de publicaciones se consultan bases de datos y repositorios (*Scopus*, *Dialnet*, *ACL Anthology*, *arXiv.org*).

<http://www.elsevier.com/online-tools/scopus>

<http://dialnet.unirioja.es>

<http://www.aclweb.org/anthology>

<http://arxiv.org>

Los proyectos se identificaron en la base de datos *Cordis*, pero también en webs de instituciones y programas de financiación específicos de investigación en HSC:

- *German Federal Ministry of Education and Research* (Alemania);
- *Joint Information Systems Committee (JISC)* (Reino Unido);
- *National Science Foundation* (Estados Unidos);
- *Economic and Social Research Council* (Reino Unido);
- *National Endowment for the Humanities* (United States);
- *Social Sciences and Humanities Research Council* (Canadá);
- *Digging Into Data Challenge* (de carácter internacional);
- *Ministerio de Economía y Competitividad* (España); y
- *Ministerio de Educación Cultura y Deporte* (España).

En las búsquedas se utilizaron como palabras clave: análisis textual, análisis de texto/de texto asistido por ordenador/automático de texto, datos textuales, reconocimiento/extracción/identificación de entidades, minería de textos/de datos, herramientas informáticas/digitales, software, y aplicaciones informáticas. Todos ellos en combinación con los términos: ciencias sociales, humanidades o humanidades digitales, pero también con nombres específicos de disciplinas (historia, ciencia política, literatura, psicología, etc.).

Dado que el objetivo era permitir que el usuario estableciera un itinerario de información desde el descubrimiento de la herramienta hasta cómo utilizarla, se establecieron como criterios delimitadores:

a) rango temporal (2009-2014) para que fueran investigaciones recientes;

b) que hubiera información explícita sobre el uso de una aplicación informática identificada y se explicara para qué y cómo se había utilizado.

Este segundo criterio fue determinante para desestimar un gran número de trabajos, poniendo de

manifiesto que muchos investigadores no mencionan explícitamente esta información en la comunicación de sus resultados.

El centro *Clarín IULA-UPF* ofrece acceso a herramientas útiles para el análisis y la explotación de textos en forma de servicios web

3. Características del catálogo

La aplicación del modelo LOD a un catálogo genera un producto abierto de información al vincular datos que se encuentran distribuidos en la Web. En el catálogo piloto del centro *Clarín IULA-UPF*, los datos enlazados abiertos, gracias a su estructura de grafo, posibilitan al usuario durante la navegación, por ejemplo, el descubrimiento de un informe final de proyecto donde se usan aplicaciones informáticas para revisar y extraer información de los textos, pero también le permiten acceder a otros trabajos del mismo autor u otros proyectos en los que ha participado al estar estos datos relacionados en forma de enlaces visitables.

En el catálogo se publican datos generados específicamente para el mismo, como la descripción de un proyecto o de una herramienta en particular, pero también se reutiliza información de otros recursos ajenos al establecer relaciones con fuentes externas, por ejemplo utilizando el *Virtual international authority file (VIAF)* para identificar un autor, o recurriendo a la base de datos *Faceted digital bibliography and library project (DBLP)* para establecer el identificador único de una comunicación a un congreso.

<http://viaf.org>

<http://dblp.l3s.de>

En los siguientes apartados se explica con más detalle el contenido del catálogo y los instrumentos de descripción utilizados (metadatos y vocabularios).

Tabla 1. Proyectos seleccionados

N.	Disciplina	Acrónimo	Nombre	Fecha
1	Historia	<i>Chalice</i>	<i>Connecting historical authorities with linked data, contexts and entities</i>	2010 – 2011
2	Historia	<i>ChartEx</i>	---	2011 –
3	Historia	<i>DbyD</i>	<i>Digging by debating</i>	2012 – 2013
4	Multidisciplinar	<i>DMCI</i>	<i>Data mining with criminal intent</i>	2010 – 2011
5	Historia	<i>DPRM</i>	<i>Digital prosopography for renaissance musicians</i>	2013 – 2014
6	Historia	<i>DVE</i>	<i>Dynamic variorum editions</i>	2010 – 2011
7	Historia	<i>Isher</i>	<i>Integrated social history environment for research</i>	2012 – 2013
8	Historia	<i>MRL</i>	<i>Mapping the republic of letters</i>	2009 - ?
9	Historia	<i>MT</i>	<i>Mapping texts</i>	2010 – 2012
10	----	<i>Scrutiny</i>	----	2009 – 2010

3.1. El contenido

El catálogo relaciona información de proyectos, publicaciones y herramientas informáticas⁶. Cabe recordar que los contenidos presentados en este trabajo no son definitivos ya que el catálogo está en constante expansión, las cifras de publicaciones, proyectos y herramientas son efímeras, y en el momento de la consulta real al catálogo la información puede haber cambiado debido a la incorporación periódica de nuevos items.

Proyectos

Incluyen proyectos de investigación que utilizan o han utilizado tecnologías lingüísticas. En esta primera selección se incorporan diez que trabajan con el reconocimiento de entidades con varias finalidades (tabla 1).

A continuación se presentan dos ejemplos:

- **ChartEx**: mediante el uso de tecnologías de procesamiento del lenguaje natural (NLP), que incluyen el NER (con la herramienta *Brat*), el proyecto propone aumentar el saber sobre la vida doméstica y relaciones entre las personas en la Edad Media a través de la extracción de información sobre lugares, personas o eventos en contratos de compra/venta de propiedades en formato digital.

<http://www.chartex.org>

<http://brat.nlplab.org>

- **Isher**: tiene como objetivo investigar la aplicación de software para detectar, vincular y visualizar eventos, tendencias, nombres de personas y organizaciones, así como cualquier otra entidad de interés para la historia social. Utiliza la plataforma *U-Compare* para el procesamiento de textos.

<http://www.nactem.ac.uk/DID-ISHER>

<http://u-compare.org>

Publicaciones

Incluye información sobre artículos, comunicaciones a congresos, informes, etc., sobre las herramientas y cómo se han utilizado. Para ilustrar la tarea de reconocimiento de entidades se incorporan de momento 14 publicaciones (tabla 2).

A continuación se describen sintéticamente algunas de estas publicaciones⁷ a modo de ejemplo:

En el campo de la historia, **Torget et al.** (2011) y **Yang, Torget y Mihalcea** (2011) informan sobre los resultados del proyecto *MT*. En las publicaciones se explica cómo el proyecto explora la combinación de la minería de textos con la cartografía geoespacial para estudiar los periódicos históricos gracias a la frecuencia de palabras, el reconocimiento de entidades con nombre (con *Stanford NER*), y *Topic models*⁸ (con *Mallet*).

<http://nlp.stanford.edu/software/CRF-NER.shtml>

<http://mallet.cs.umass.edu>

En el campo de la ciencia política **Shor et al.** (2014) estudian la cobertura informativa de las mujeres políticas en periódicos de tendencia liberal y conservadora. En este caso se utiliza la identificación de nombres de personas, con el software *Lydia text analysis*⁹, como parte del proceso de investigación.

Waila, Singh y Singh (2013) presentan el análisis de blogs con contenido político, mediante el análisis textual asistido por ordenador en el modelado de temas (*topic models*), el análisis de sentimientos, y el reconocimiento de entidades

Tabla 2. Publicaciones seleccionadas

N.	Disciplina	Título	Autores	Año
1	Ciencias políticas	<i>Is there a political bias?</i>	Eran Shor et al.	2014
2	Cinematografía	<i>The movie mashup application MoMa</i>	Jean-Marc Finsterwald et al.	2012
3	Literatura	<i>Advanced visual analytics methods for literature analysis</i>	Daniela Oelke; Dimitrios Kokkinakis; Mats Malm	2012
4	Literatura	<i>Names in novels</i>	Karina Van-Dalen-Oskam	2013
5	Literatura	<i>The geographic imagination of civil war era American fiction</i>	Matthew Wilkens	2013
6	Literatura	<i>From speaker identification to affective analysis</i>	Elias Iosif; Taniya Mishra	2014
7	Literatura	<i>Structure based clustering of novels</i>	Mariona Coll-Ardanuy; Caroline Sporleder	2014
8	Literatura/Sistemas información geográfica	<i>Mapping the English lake district</i>	David Cooper; Ian Gregory	2011
9	Literatura/Sistemas información geográfica	<i>Visual GISting</i>	Ian Gregory; Andrew Hardie	2011
10	Historia	<i>Topic modeling on historical newspapers</i>	Tze-I Yang; Andrew Torget; Rada Mihalcea	2011
11	Historia	<i>Data mining with criminal intent</i>	Dan Cohen et al.	2011
12	Historia	<i>Mapping texts</i>	Andrew Torget et al.	2011
13	Sociología	<i>Analyzing androcentric focus & dehumanization in news headlines using advanced exploitation tools</i>	Marta Villegas, Maite Melero, Núria Bel	2011
14	Sociología/Ciencias políticas	<i>Blog text analysis using topic modeling, named entity recognition and sentiment classifier combine</i>	Pranav Waila; V. K. Singh; M. K. Singh	2013

nombradas. Utilizan la herramienta *Alchemy API named entity extraction* para reconocer nombres de personas, organizaciones, lugares y fechas.

<http://www.alchemyapi.com/products/features/entity-extraction>

En literatura y lingüística **Oelke, Kokkinakis y Malm** (2012) analizan un subconjunto del *Banco literario sueco* centrándose en la extracción de nombres de persona, su género y su forma normalizada, incluyendo menciones de seres teístas (dioses y figuras mitológicas), y examinan su aparición en 13 novelas con la herramienta *Swedish FS* para el reconocimiento de entidades en lengua sueca¹⁰.

Coll-Ardanuy y Sporleder (2014) investigan las redes sociales que surgen desde el personaje principal (Elizabeth) de la novela *Orgullo y prejuicio* de Jane Austen. Como primer paso para iniciar el trabajo se propone el reconocimiento de los nombres de los personajes de la novela utilizando la herramienta *Stanford NER*.

http://www.tekstlab.uio.no/nn/foredrag/fefor03/dimitris_fefor03.pdf

Otros autores aplican el análisis geoespacial a los textos literarios. Es el caso de **Wilkens** (2013) que estudia el uso de los lugares geográficos reales en la literatura de ficción norteamericana del siglo XIX. Utiliza también el *Stanford NER* para la identificación y extracción de nombres de lugares geográficos que posteriormente identifica en mapas con sistemas de información geográfica.

Herramientas

Incluyen aplicaciones de tecnologías lingüísticas para el análisis de texto. En esta primera versión del catálogo se decide incorporar solamente servicios web alojados en el centro *Clarín IULA-UPF*, aunque se está valorando la integración de productos de otros proveedores de servicios que también los ofrezcan en acceso libre. Para ejemplificar la aplicabilidad al reconocimiento de entidades se incorporan dos servicios web:

1) *Freeling*: se integran dos módulos en formato servicio web del kit de productos de análisis lingüístico *Freeling* que incluyen el desempeño de la tarea de reconocimiento de entidades nombradas: el *Analizador morfosintáctico Freeling* y el *Etiquetador morfosintáctico Freeling*.

http://ws04.iula.upf.edu/soaplab2-axis/#morphosyntactic_tagging.freeling3_morpho_row

http://ws04.iula.upf.edu/soaplab2-axis/#morphosyntactic_tagging.freeling3_tagging_row

2) *ContaWords*: aplicación web que analiza las palabras de un archivo de texto, clasificándolas por categorías (nombres, verbos, etc.), asignándoles su correspondiente lema, para después contar su frecuencia de aparición en el texto. También realiza el reconocimiento de entidades con nombre, por ejemplo nombres de personas, de empresas, de organizaciones, geográficos, etc.

<http://contawords.iula.upf.edu>

El catálogo relaciona información relativa a proyectos, publicaciones y herramientas informáticas

3.2. Instrumentos de descripción

Metadatos

Para la descripción de los diversos items se adopta una combinación de esquemas de metadatos que permite proporcionar siempre una descripción genérica, pero también una especializada para aquellos casos que la requieran.

En la descripción genérica se utilizan metadatos del conjunto básico de elementos del esquema *Dublin core (DC)*: identificador, autor, colaborador, título, descripción y materia. También del mismo esquema se aplican metadatos de especificación para incluir información sobre la citación bibliográfica (en formato *Vancouver*) y las relaciones entre publicaciones, herramientas y proyectos. Se identifican con el prefijo “dc” y “dcterms” respectivamente.

<http://dublincore.org>

Debido a que se apuesta por la inclusión de aplicaciones informáticas que no necesitan instalación, en formato de servicio web, se usa el esquema *Biocatalogue* para describirlas. Éste es el esquema de metadatos elaborado para la plataforma del mismo nombre donde se describen servicios web del ámbito de las ciencias de la vida. Los metadatos específicos que se han utilizado de este esquema son: versión, tecnología, tarea ejecutada, persona de contacto, proveedor del servicio, dirección web de acceso, dirección de ejecución (*wSDL*), ejemplo de texto de entrada y ejemplo de resultado. Se identifican con el prefijo “bio”.

Tabla 3. Vocabularios utilizados

Acrónimo	Nombre	Web
BIBO	<i>Bibliographic ontology</i>	http://bibliontology.com/specification
BIO	<i>Biocatalogue</i>	https://www.biocatalogue.org
Faceted DBLP	<i>Faceted digital bibliography and library project</i>	http://dblp.l3s.de/?q=&newQuery=yes&resTableName=query_resultucocuv
Dbpedia	<i>Dbpedia</i>	http://dbpedia.org/About
DCMI Types	<i>DCMI type vocabulary</i>	http://dublincore.org/documents/dcmi-type-vocabulary
FOAF	<i>Friend of a friend</i>	http://www.foaf-project.org
ISOCat	<i>Data category registry</i>	http://www.isocat.org
LC/NAF	<i>Library of Congress name authority file</i>	http://id.loc.gov
VIAF	<i>Virtual international authority file</i>	http://viaf.org

<http://purl.org/ms-lod/BioServices.ttl>
<https://www.biocatalogue.org>

En previsión de la futura incorporación y vinculación con recursos lingüísticos se integra además el esquema de metadatos *MetaShare (MS)*¹¹ para una descripción más acotada de los mismos. Posibles ejemplos de aplicación son: la codificación de caracteres (UTF8, ascii, etc.), forma de creación (manual o automática), nombre del proyecto que ha financiado su construcción, etc. Se identifican con el prefijo “ms”.

Vocabularios

Permiten asignar valores a los metadatos seleccionados para la descripción. El detalle de los que se utilizan en el catálogo se muestra en la tabla 3.

« A través de la navegación entre los diferentes elementos interconectados del catálogo el usuario puede establecer un recorrido de consulta que puede ser meramente informativo o dar lugar directamente a la experimentación »

4. Ejemplo de consulta del catálogo

La gestión de la información en LOD se realiza con el programa *TopBraid*, gestor de ontologías, que permite la integración de vocabularios y esquemas de metadatos. Todos los datos son almacenados en un punto de acceso *Sparql*. Para facilitar el primer contacto con el catálogo piloto se crea una página introductoria con información contextual sobre sus objetivos, contenidos y características. La consulta del catálogo se implementa en un buscador local.

<http://www.topquadrant.com/tools/ide-topbraid-composer-maestro-edition>
<http://www.w3.org/TR/rdf-sparql-query>

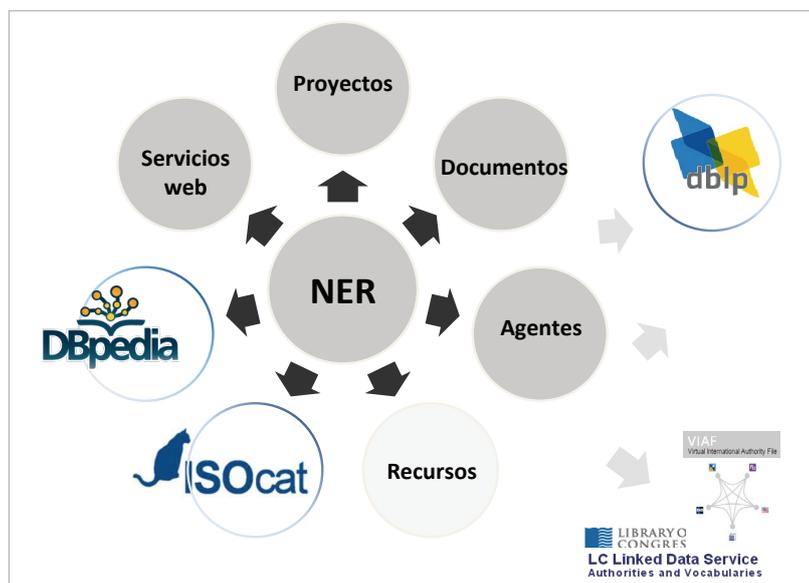


Figura 1. Modelo de vinculación LOD del catálogo

http://ws02.iula.upf.edu/corpus_data/webTest/home.html

El catálogo relaciona información sobre:

- Proyectos (de investigación que utilizan o han utilizado tecnologías lingüísticas);
- Herramientas (servicios web de tecnologías lingüísticas);
- Documentación (artículos, comunicaciones, etc., sobre las herramientas y cómo se han utilizado)¹².

Para acercar la consulta de los materiales a sus potenciales usuarios, facilitar el razonamiento analógico y la utilización de la información, se complementan los elementos anteriores con tres items más:

- Áreas (disciplinas y temas de investigación);
- Agentes (personas y organizaciones);
- Tareas (que realizan las herramientas).

En la versión actual del catálogo la tarea de reconocimiento de entidades nombradas actúa como hilo conductor de la interrelación de elementos (figura 1).

Por esta razón la relación entre los elementos se puede describir de la siguiente forma:

- los documentos explican el uso de *NER (named entity recognition)* en investigaciones que realizan análisis de textos;
- los proyectos hacen referencia a la utilización de *NER* en investigaciones de análisis de texto;
- los servicios web ejecutan *NER* sobre los textos que propongan los usuarios;
- los agentes (personas) son autores de documentos, crean servicios, o participan en proyectos;
- los agentes (instituciones) participan en proyectos o representan la filiación de las personas;
- los recursos son utilizados como corpus para detectar y extraer nombres de entidades (relación no implementada).

NER está presente como tema (si se trata de documentos o proyectos) y como tarea si está relacionada con los servicios.

¿Cómo se inicia el itinerario del usuario potencial?

Ya desde la página inicial de la web del centro de competencias se invita al usuario a realizar un itinerario de navegación, marcado por la analogía, al plantear como invitación *¡Mira qué hacen los demás!* como punto de acceso a la entrada al catálogo (figura 2).

http://iula02v.upf.edu/corpus_data/webTest/index.html

Como se ve en la figura 2, el usuario se encuentra con una nube de temas cuya finalidad es que éste se sienta identificado con alguno de ellos debido a su ejercicio profesional, estudios o investigación.

Esta nube de términos también guía su entrada en el catálogo ya que la primera pantalla que se visualiza es una lista de áreas

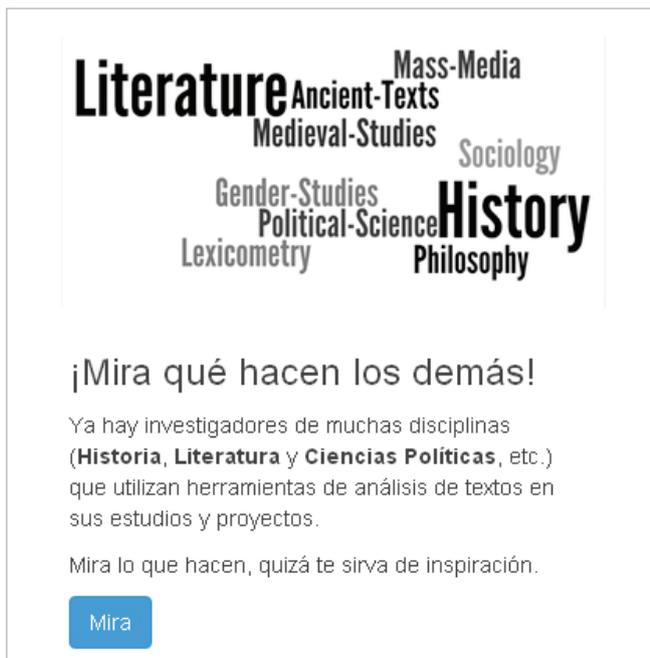


Figura 2. Web centro *Clarín IULA-UPF*

(en sentido amplio: se incluyen disciplinas pero también temas de investigación) a través de las cuales puede iniciar su itinerario de navegación (figura 3).

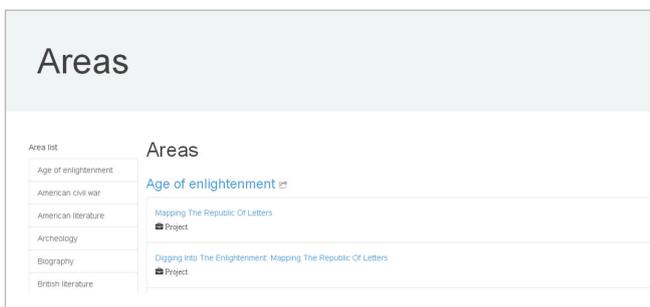


Figura 3. Catálogo: página de acceso por áreas

Una vez en esta página, el usuario puede recorrer la lista de áreas para seleccionar alguna de su interés, como no tener ninguna preferencia y visitar los primeros enlaces que se ofrecen. Ejemplo: selección del área de *American literature* (figura 4).



Figura 4. Catálogo: selección de área específica

Bajo cada área se listan documentos y proyectos relacionados. Aquí nuevamente el usuario decide su ruta de navegación entre los resultados que se ofrecen. Ejemplo: se escoge el primer ítem de la lista *The geographic imagination of Civil war-era American fiction* (figura 5).



Figura 5. Catálogo: ejemplo de publicación

A partir de este momento el usuario puede hacer clic en los elementos con hipervínculos (destacados en azul): autor (*creator*), materia (*subject*), identificador (*identifier*), o área (*area*). Hay dos tipos de enlaces:

- Externos: relacionan datos obtenidos en fuentes ajenas y están marcados con un icono en forma de flecha. Si el usuario visita estos enlaces saldrá del catálogo para consultar las fuentes originales de los datos externos (por ejemplo *DBpedia*).
- Internos: aquellos que vinculan información contenida en el propio catálogo (por ejemplo *named entity recognition*). Si el usuario decide visitar este enlace se traslada a otra página del catálogo donde encuentra todos los elementos que fueron etiquetados con dicha tarea (figura 6).

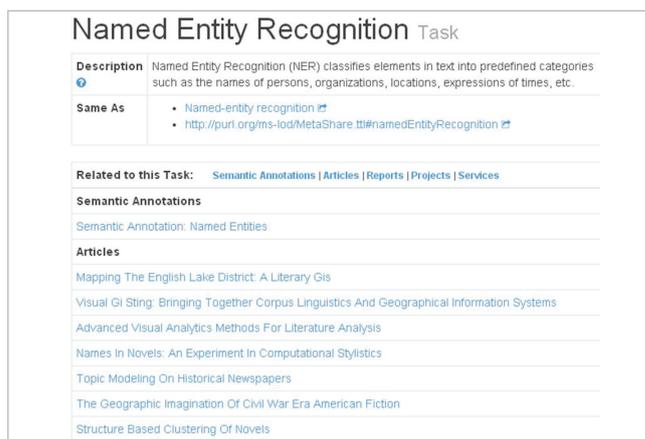


Figura 6. Catálogo: Tarea *named entity recognition*

Llegados a este punto es posible visualizar el potencial de las relaciones establecidas en las informaciones contenidas en el catálogo. Los usuarios tienen acceso a ampliar información sobre la tarea si no la conocen, con información interna del catálogo (definición (*description*) y tipificación como tipo de anotación semántica (*semantic annotation: named entities*), como también pueden profundizar dicha informa-

ción visitando fuentes externas enlazadas como la *DBpedia*. Además pueden ver que esta tarea está relacionada con artículos, informes, proyectos y herramientas para conocer de primera mano otras experiencias que la utilizan en su metodología (en pestañas *Articles, Reports & Projects*), y experimentar con herramientas similares (en pestaña *Services*).

Mediante la navegación entre los diferentes elementos interconectados del catálogo el usuario puede establecer por sí mismo un recorrido de consulta que puede ser meramente informativo o dar lugar directamente a la experimentación.

5. Conclusiones

El uso de tecnología en la práctica de las humanidades y ciencias sociales (HCS) no es una novedad (Roberts, 1997; Popping, 2000). Sin embargo, la utilización de términos como *humanidades digitales* (Burdick et al., 2012; Liu, 2012), *culturomics* (Michel et al., 2011; Bohannon, 2011) o *big data en humanidades, artes y ciencias sociales* (Parry, 2010; Leetaru, 2012) pone de manifiesto el interés por el uso de la tecnología en la praxis de la investigación en las HSC. Es necesario capacitar a los investigadores para que descubran trabajos existentes que sean referentes para utilizar los programas disponibles en su objeto de estudio.

Disponer de un repositorio de información especializada que incluya material de referencia sobre la utilización de aplicaciones informáticas en la investigación en HSC y DH, capacita a los investigadores para explorar prácticas alternativas a las habituales. Por otra parte, a nivel más técnico, la integración de recursos dispares en una única fuente de información beneficia la sostenibilidad, interoperabilidad e integridad de los datos, lo cual repercute en la optimización de los costos y la funcionalidad. La propuesta de un catálogo basado en LOD permite una explotación de datos simple y eficiente. Por ejemplo, a partir del ejemplo diseñado sobre *NER (named entity recognition)*, el usuario puede con una simple consulta acceder a publicaciones, proyectos, servicios web, y expertos relacionados con ese tema. Asimismo, el uso de LOD posibilita que el usuario sea el protagonista pudiendo escoger el itinerario de descubrimiento que más le convenga.

Notas

1. *Common Language Resources and Technologies*
<http://www.clarin.eu>
2. Este trabajo ha sido financiado por el *Fondo Europeo de Desarrollo Regional (Feder)*, Programa operativo *Feder de Cataluña 2007-2013*, Objetivo 1.
3. El centro *Clarin IULA-UPF* se ubica en la sede del *Instituto de Lingüística Aplicada (IULA)* de la *Universitat Pompeu Fabra (UPF)*.
4. *Glossary of Archival and Records Terminology*
<http://www2.archivists.org/glossary/terms/a/actionable-information>
5. Si se desea formar parte de los grupos de testeo del catálogo envíe un correo electrónico a: iulatrl@upf.edu
6. Se está implementando la inclusión de recursos lingüísticos,

como corpus y léxicos, aunque todavía no se encuentran relacionados con el resto de la información.

7. La descripción de las herramientas se incluye en un anexo al final del artículo.

8. *Topic models*: en procesamiento de lenguaje natural se denomina así a un modelo estadístico para descubrir los temas implícitos en una colección de documentos a partir del análisis de las palabras que aparecen más frecuentemente en el texto.

9. Aunque la aplicación está citada en diversas publicaciones no se ha podido identificar una dirección web para su consulta.

10. Cabe destacar que en el artículo la herramienta no estaba identificada claramente salvo a través de los investigadores que la desarrollaron, información a partir de la cual se pudo rastrear el nombre de la misma. Dado el interés del tema y la escasez de resultados obtenidos se ha decidido igualmente incluirla en el catálogo. Es una clara muestra de las carencias en la descripción de la metodología en las publicaciones tal y como se remarcaba en el apartado de metodología del presente trabajo.

11. El esquema de metadatos completo está disponible en el manual *Documentation and user manual of the META-Share metadata model*.

<http://www.meta-net.eu/meta-share/META-SHARE%20%20documentationUserManual.pdf>

12. Se recuerda que los recursos lingüísticos se contemplan como futura tarea en la vinculación de datos, por lo tanto también aparece en el modelado de la consulta.

6. Bibliografía

- Bohannon, John (2011). "Google books, Wikipedia, and the future of culturomics". *Science*, v. 331, n. 6014, p. 135.
<http://dx.doi.org/10.1126/science.331.6014.135>
- Brier, Alan; Hopp, Bruno (2011). "Computer assisted text analysis in the social sciences". *Quality & quantity*, v. 45, n. 1, pp. 103-128.
<http://dx.doi.org/10.1007/s11135-010-9350-8>
- Burdick, Anne; Drucker, Johanna; Lunenfeld, Peter; Presner, Todd; Schnapp, Jeffrey (2012). *Digital humanities*. Cambridge: MIT Press. ISBN: 978 0262018470
http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf
- Cohen, Dan; Gibbs, Frederick; Hitchcock, Tim et al. (2011). *Data mining with criminal intent*. Final white paper.
<http://criminalintent.org/wp-content/uploads/2011/09/Data-Mining-with-Criminal-Intent-Final1.pdf>
- Coll-Ardanuy, Mariona; Sporleder, Caroline (2014). "Structure-based clustering of novels". En: *Procs of the 3rd Workshop on computational linguistics for literature (CLfL)*, pp. 31-39.
<http://www.aclweb.org/anthology/W/W14/W14-0905.pdf>
- Cooper, David; Gregory, Ian (2011). "Mapping the English Lake District: a literary GIS". *Transactions of the Institute of British Geographers*, v. 36, n. 1, pp. 89-108.

<http://dx.doi.org/10.1111/j.1475-5661.2010.00405.x>

Finsterwald, Jean-Marc; Grefenstette, Gregory; Law-To, Julien; Bouchard, Hugues; Djali-Mezaour, Amar (2012). "The movie mashup application MoMa: geolocating and finding movies". En: *Procs of the ACM multimedia 2012 workshop on geotagging and its applications in multimedia, GeoMM'12*, pp. 15-18.

Gregory, Ian; Hardie, Andrew (2011). "Visual GISTing: bringing together corpus linguistics and geographical information systems". *Literary and linguistic computing*, v. 26, n. 3, pp. 297-314. <http://dx.doi.org/10.1093/lc/fqr022>

Iosif, Elias; Mishra, Taniya (2014). "From speaker identification to affective analysis: a multi-step system for analyzing children's stories". En: *Procs of the 3rd Workshop on computational linguistics for literature (CLfL)*, pp. 40-49. <http://aclweb.org/anthology/W14-0906>

Juola, Patrick (2008). "Killer applications in digital humanities". *Literary and linguistic computing*, v. 23, n. 1, pp. 73-83. <http://www.mathcs.duq.edu/~juola/papers.d/killer.pdf> <http://dx.doi.org/10.1093/lc/fqm042>

Leetaru, Kalev H. (2012). "A big data approach to the humanities, arts, and social sciences: *Wikipedia's* view of the world through supercomputing". *Research trends*, n. 30, September. <http://www.researchtrends.com/issue-30-september-2012/a-big-data-approach-to-the-humanities-arts-and-social-sciences-summary>

Liu, Alan (2012). "The state of the digital humanities: a report and a critique". *Arts and humanities in higher education*, v. 11, n. 1-2, pp. 8-41. <http://dx.doi.org/10.1177/1474022211427364>

Michel, Jean-Baptiste; Kui-Shen, Yuan; Presser-Aiden, Aviva; Veres, Adrian; Gray, Matthew K.; The Google Books Team; Pickett, Joseph P.; Hoiberg, Dale; Clancy, Dan; Norvig, Peter; Orwant, Jon; Pinker, Steven; Nowak, Martin A.; Liebermann-Pinker, Steven (2011). "Quantitative analysis of culture using millions of digitized books". *Science*, v. 331, n. 6014, pp. 176-182. http://scholar.harvard.edu/files/pinker/files/michel_et_al_quantitative_analysis_of_culture_science_2011.pdf <http://dx.doi.org/10.1126/science.1199644>

Oelke, Daniela; Kokkinakis, Dimitrios; Malm, Mats (2012). "Advanced visual analytics methods for literature analysis". En: *Procs of the 6th Workshop on language technology for cultural heritage, social sciences, and humanities*, pp. 35-44. <http://dl.acm.org/citation.cfm?id=2390357.2390364>

Parry, Marc (2010). "The humanities go Google". *The chronicle of higher education*, May 28. <http://chronicle.com/article/The-Humanities-Go-Google/65713>

Popping, Robert (2000). *Computer-assisted text analysis*. London: Sage. ISBN: 0761953795

Roberts, Carl W. (ed.) (1997). *Text analysis for the social sciences: methods for drawing statistical inferences from texts and transcripts*. Mahwah: Lawrence Erlbaum. ISBN: 0805817352

Schreibman, Susan; Hanlon, Anne (2010). "Determining value for digital humanities tools: report on a survey of tool developers". *Digital humanities quarterly*, v. 4, n. 2. <http://digitalhumanities.org/dhq/vol/4/2/000083/000083.html>

Shor, Eran; Van-de-Rijt, Arnout; Ward, Charles; Askar, Saoussan; Skiena, Steven (2014). "Is there a political bias? A computational analysis of female subjects' coverage in liberal and conservative newspapers". *Social science quarterly*. <http://dx.doi.org/10.1111/ssqu.12091>

Target, Andrew J.; Mihalcea, Rada; Christensen, Jon; McGehee, Geoff (2011). *Mapping texts: combining text-mining and geo-visualization to unlock the research potential of historical newspapers. A white paper for the National Endowment for the Humanities*. http://mappingtexts.stanford.edu/whitepaper/MappingTexts_WhitePaper.pdf

Van-Dalen-Oskam, Karina (2013). "Names in novels: an experiment in computational stylistics". *Literary and linguistic computing*, v. 28, n. 2, pp. 359-370. <http://dx.doi.org/10.1093/lc/fqs007>

Villegas, Marta; Melero, Maite; Bel, Nùria (2014). "Metadata as linked open data: mapping disparate XML metadata registries into one RDF/OWL registry [forthcoming, accepted paper]". En: *Ninth intl conf on language resources and evaluation (LREC-2014)*, 26-31 May 2014, Reykjavik, Iceland. http://www.lrec-conf.org/proceedings/lrec2014/pdf/664_Paper.pdf

Waila, Pranav; Singh, V. K.; Singh M. K. (2013). "Blog text analysis using topic modeling, named entity recognition and sentiment classifier combine" En: *Procs advances in computing, communications and informatics*, pp. 1166-1171. <http://dx.doi.org/10.1109/ICACCI.2013.6637342>

Wiedemann, Gregor (2013). "Opening up to big data: computer-assisted analysis of textual data in social sciences". *Forum qualitative sozialforschung/Forum: qualitative social research*, v. 14, n. 2. <http://www.qualitative-research.net/index.php/fqs/article/view/1949>

Wilkens, Matthew (2013). "The geographic imagination of Civil war-era American fiction". *American literary history*, v. 25, n. 4, pp. 803-840. <http://dx.doi.org/10.1093/alh/ajt045>

Yang, Tze-I; Target, Andrew J.; Mihalcea, Rada (2011). "Topic modeling on historical newspapers". En: *Procs of the 5th ACL-HLT Workshop on language technology for cultural heritage, social sciences, and humanities*, pp. 96-104. <http://dl.acm.org/citation.cfm?id=2107636.2107649>

7. Anexo. Descripción de herramientas

Se muestran breves descripciones de las aplicaciones y/o plataformas utilizadas tanto en los proyectos como en las publicaciones seleccionadas para la presente versión del catálogo. Éstas no se incluyen de momento en el catálogo. Todas permiten ejecutar la tarea de reconocimiento de entidades.

Alchemy API named entity extraction: interfaz de programación de aplicaciones (API, *application programming interface*) para la identificación de nombres de personas, empresas, organizaciones, ciudades, accidentes geográficos, etc., en textos digitales o webs. Emplea algoritmos estadísticos y tecnología de procesamiento del lenguaje natural para analizar y extraer la información. Es una aplicación comercial y dispone de una versión de prueba.

Brat: aplicación web para crear anotaciones para reconocimiento de entidades y relaciones binarias útiles para tareas de extracción de información, entre otras.

Lydia text analysis: sistema de procesamiento de lenguaje natural para el análisis y la extracción de información especializado en medios de comunicación online (periódicos, blogs, etc.). Permite relacionar personas, lugares y temas a partir del procesamiento del lenguaje natural y el análisis estadístico de frecuencias nombres de entidades y colocaciones. Actualmente está integrado en el buscador especia-

lizado en entidades *Textmap*.
<http://www.textmap.com/index.htm>

Mallet: paquete de aplicaciones *java* para procesamiento estadístico del lenguaje natural, clasificación de documentos, agrupamiento (*clustering*), modelado de temas (*topic models*), extracción de información, y otras aplicaciones de aprendizaje automático para textos.

Swedish FS: aplicación para el reconocimiento de entidades específica para la lengua sueca. Está desarrollada sobre la base de gramáticas de estado finito sensitivas al contexto (*context-sensitive finite-state grammars*). No se ha encontrado web específica.

Stanford NER: aplicación *java* para el reconocimiento de nombres de entidad.

U-Compare: sistema integrado de minería de texto y procesamiento del lenguaje natural desarrollado en entorno *UIMA* (*Unstructured information management applications*, entorno especializado en análisis de grandes volúmenes de información no estructurada). Se compone de un conjunto de módulos interoperables y flujos de trabajo a través de una interfaz única. Los módulos que incluye son: lectores de corpus, visualizadores, editores de texto, y herramientas sintácticas, semánticas (donde se incluye el reconocimiento de entidades), y estadísticas.



5ª Conferencia internacional sobre calidad de revistas de ciencias sociales y humanidades

CRECS 2015
7-8 de mayo

Universidad de Murcia. Facultad de Comunicación y Documentación

<http://www.thinkepi.net/crecs2015>





REALIDAD AUMENTADA EN TELEVISIÓN Y PROPUESTA DE APLICACIÓN EN LOS SISTEMAS DE GESTIÓN DOCUMENTAL



Jorge Caldera-Serrano



Jorge Caldera-Serrano es doctor por la *Universidad de Salamanca*, y profesor titular de la *Universidad de Extremadura* en el *Departamento de Información y Comunicación*. Durante el curso 2014/15 está vinculado a la *Universidad Central del Ecuador* como investigador *Prometeo* financiado por el *Gobierno de Ecuador*. Autor de publicaciones sobre archivos de televisión y documentación audiovisual. Cuenta con estancias de investigación en Argentina, Brasil, Cuba, Ecuador, Uruguay y varias universidades españolas. Es evaluador de proyectos para la *Senacyt* en Panamá y colaborador en el programa *IBERmemoria audiovisual y sonora*.
<http://orcid.org/0000-0002-4049-1057>

*Universidad de Extremadura,
Facultad de Ciencias de la Documentación y la Comunicación
Plazuela de Ibn Marwan, s/n. 06001 Badajoz, España
jcalser@alcazaba.unex.es*

Resumen

Entendemos como realidad aumentada (RA) la tecnología que por medio de diferentes dispositivos incorpora información digital (virtual o real) a información obtenida de un entorno real, aportando por tanto información adicional. Se analiza cómo la RA puede llegar a aplicarse a la difusión y captación de contenidos por parte de las empresas audiovisuales y televisivas. Se trataría de buscar soluciones en la gestión de la documentación audiovisual en las unidades de información. En primer lugar se realiza un estado de la cuestión sobre la RA, desde diferentes puntos de vista. Se muestran las aplicaciones que están surgiendo para aplicar la RA a los medios televisivos. Finalmente se exponen los cambios que deberían implementarse en la gestión de la información tanto audiovisual, sonora como textual (o hipertextual) para aprovechar la RA.

Palabras clave

Realidad aumentada, Televisión, Información digital, Servicios de gestión documental, Documentación audiovisual, Documentación en televisión.

Title: Augmented reality in television and proposed application to document management systems

Abstract

Augmented reality (AR) is a technology that uses different devices to incorporate digital information (virtual or real) into information obtained from a real-life environment, thereby providing additional information. We examine how AR can contribute to the dissemination and uptake of content from audiovisual and television companies. We try to find solutions in the management of audiovisual documentation in information units. First, a state of the art description of AR is provided from different points of view. Applications that are emerging to apply RA to television media are described. Finally the changes that should be implemented in the management of visual, audio and text (or hypertext) information to benefit from AR are discussed.

Keywords

Augmented reality, Television, Digital information, Managed document services, Audiovisual documentation, TV documentation.

Caldera-Serrano, Jorge (2014). "Realidad aumentada en televisión y propuesta de aplicación en los sistemas de gestión documental". *El profesional de la información*, v. 23, n. 6, noviembre-diciembre, pp. 643-650.

<http://dx.doi.org/10.3145/epi.2014.nov.12>

1. La televisión y la documentación: mutación constante

La televisión como invento no cuenta ni con ochenta años de historia, un período insignificante si lo comparamos con la historia de la humanidad, pero una eternidad des-

de la perspectiva del cambiante entorno tecnológico de las últimas décadas. No sólo se puede hablar de evolución constante en las tecnologías de la comunicación y la información, sino también de revolución tecnológica a la vista del cambio de paradigma en la manera de hacer y entender la información.

Artículo recibido el 07-08-2014
Aceptación definitiva: 22-10-2014

La televisión de la década de los años cincuenta del siglo pasado no tiene nada que ver con la actual. Sus formas de emisión y producción han variado de forma drástica, pero se ha producido además un hecho fundamental que ha dado lugar a un cambio de modelo: la digitalización.

Dicha evolución ha generado transformaciones en la forma de producción en las industrias televisivas, al igual que en los métodos de captación y difusión de la señal audiovisual. Los procesos de digitalización han modificado muchos aspectos que estaban consolidados en las televisiones, posibilitando la producción audiovisual digital, la gestión de información online, el acceso a los contenidos desde móviles, la interconexión con la red generando lo que se denomina televisión social, etc.

El modelo digital audiovisual ha cambiado el medio televisivo hacia una televisión social, accesible, multiformato y multisoporte, móvil y en continua evolución

El modelo digital audiovisual ha cambiado el medio televisivo hacia una televisión social, accesible, multiformato y multisoporte, móvil y en continua evolución hacia una personalización e individualización en la difusión de los contenidos. La televisión es ahora mucho más que un canal de difusión de información, se ha convertido en un lugar para el intercambio.

Los servicios de gestión documental de las empresas audiovisuales se han adaptado a una realidad cambiante en el marco de un modelo diferente de producción (que tiene en cuenta las distintas plataformas comunicativas que la televisión utiliza), y de interacción con sus usuarios (más allá de los periodistas). En este escenario los servicios de información tienen la obligación de adaptar sus hábitos a las nuevas formas de hacer televisión, y estar alerta ante los sistemas que hoy se vislumbran como potenciales y que pueden estar a disposición de los usuarios en los próximos años.

Este trabajo se centra en el análisis de la realidad aumentada (RA) en las televisiones, describiendo la situación actual y sus potencialidades, e identificando cómo puede hacer cambiar los modelos de gestión documental en las empresas televisivas.

Los servicios de documentación de tales empresas han ido habitualmente a remolque de los requerimientos institucionales, algo que se explica tanto por el trepidante cambio tecnológico, como por ciertos hábitos de trabajo. Por otra parte la *Federación Internacional de Archivos de Televisión (FIAT/IFTA)*, la organización de carácter supranacional más importante del sector, facilita recomendaciones pero en la práctica no suelen ser utilizadas por la mayoría de empresas televisivas. Además, esa institución va igualmente por detrás de las realidades.

La RA no se traducirá probablemente en un cambio en el modelo televisivo, pero sí representará una ampliación en los canales y métodos de difusión de los contenidos que afectará a la gestión de la información.

2. Objetivos y metodología

Los objetivos planteados en este trabajo se enmarcan en el uso de la tecnología denominada realidad aumentada por parte de las empresas audiovisuales televisivas, en los diferentes formatos de difusión de contenidos.

Los objetivos están concatenados, siendo necesaria la consecución del primero para alcanzar el siguiente:

1. Identificar tecnología relacionada con RA que pueda ser utilizada en las televisiones.
2. Identificar experiencias de uso de tecnología de RA en empresas audiovisuales televisivas.
3. Identificar los cambios que genera la incorporación de RA en los servicios de información de las empresas televisivas.

La metodología en las distintas fases es:

- Fase 1. Revisión bibliográfica.
- Fase 2. Revisión del uso de RA en la televisión.
- Fase 3. Análisis y valoración de los cambios que conlleva el uso de RA en las empresas televisivas, concretamente en los servicios de documentación audiovisual.

Para la revisión bibliográfica se han consultado bases de datos, recursos web recuperados desde páginas relacionadas con el desarrollo tecnológico y medios de comunicación que informan sobre la utilización de la RA.

Se ha complementado la fase anterior con el contacto con profesionales de los medios españoles, con el fin de determinar hasta qué punto se está utilizando actualmente la RA en los medios audiovisuales.

Por último se han extrapolado los requerimientos a contemplar por los departamentos de documentación de las empresas audiovisuales con el fin de conseguir controlar el contenido que será difundido mediante RA. Se ha podido detectar que no existen experiencias en los servicios documentales de identificación de dicho contenido, consecuencia del escaso uso que en nuestro país se realiza de esta tecnología asociada a la televisión.

3. Estado de la cuestión

La evolución tecnológica ofrece cada vez más posibilidades. La RA, a diferencia de otras tecnologías, no suplanta la realidad sino que la complementa y hasta cierto punto la mejora, implementando en tiempo real la combinación de objetos virtuales y reales.

Historia y descripción del concepto

Como primera experiencia de RA debemos retroceder al año 1966, cuando **Ivan Sutherland** (1968) crea el primer sistema de RA. Dicha experiencia consistía en usar un casco con una pantalla, en la que se podía ver información simple.

El término RA no se acuña hasta 1992, cuando es utilizado para identificar la superposición de materiales generados por computadora sobre imágenes del mundo real por **Tom Caudell** y **David Mizell** (1992).

Dos años más tarde encontramos unos los principales trabajos sobre RA, de la mano de **Milgram** y **Kishino** (1994) quienes analizan el concepto relacionado con la tecnología HMD (*head-mounted display*)

Un año más tarde, los mismos autores (**Milgram et al.**, 1994) consideran RA como un término medio entre los entornos virtuales y la realidad. Otros autores, con posterioridad, han definido RA como la mezcla y/o unión entre el mundo real como parte principal con elementos virtuales que lo acompañan, complementan o aclaran.

Destacamos también la clasificación que realizan **Parhizkar et al.** (2011) de las aplicaciones de la RA: de escritorio, para móviles y web.

Azuma (1997) describe las peculiaridades de la RA desde el punto de vista tecnológico, centrándose en el “aumento” y en “los visores”. Define el primer elemento como la posibilidad de los dispositivos de incluir mayor cantidad de información sobre los objetos del mundo real (visual, sonora, etc.). El autor aborda la necesidad de contar con visores para poder llevar a cabo la combinación del mundo real con la información virtual, utilizando para ello tecnología óptica o de vídeo.

Fermín-Lobo (2013) por su parte identifica elementos comunes, y en cierta manera irrenunciables, de la RA: la ubicuidad y la información 3D.

La mayor parte de los usos vienen por la incorporación de una cámara que aporta información a la realidad y que cada vez cuenta con mayor capacidad de interacción, tanto en dispositivos fijos como móviles

La RA está en expansión y se aplica en ámbitos diversos como la educación, medicina, negocios, industria del videojuego, y también el mundo televisivo. La industria del videojuego es la que ha apostado con mayor énfasis por el uso de RA, aunque en cada entorno de los señalados se han ido analizando sus posibles vías de utilización.

La mayor parte de los usos vienen por la incorporación de una cámara que aporta información a la realidad y que cada vez cuenta con mayor capacidad de interacción, tanto en dispositivos fijos como móviles (**Henrysson; Billingham; Ollila**, 2005).

La RA sigue siendo una tecnología de futuro que aún necesita ser investigada para poder establecer qué oportunidades ofrece. En la bibliografía científica consultada se ha hallado un conjunto de ventajas e inconvenientes a nivel tecnológico. **Fermín-Lobo** (2013) muestra como su principal ventaja la obtención de información con valor añadido sobre un aspecto determinado de forma ubicua. En definitiva, información que complementa, suplanta y/o mejora la obtención de conocimiento asimilada de la realidad.

Dispositivos

Flórez-Aristizábal (2013) indica que la RA cuenta con los dispositivos de pantalla, ordenador, dispositivo de entrada y rastreo. **Olmedo y Augusto** (2012) destacan tres tecnologías:

- Visión por computadora: potencialidad de aprovechar las características visuales de lo capturado naturalmente por la cámara.
- Rastreo y geolocalización: el vídeo es etiquetado por la capacidad del dispositivo de detectar su orientación y posición, utilizando para ello sensores. Ejemplos de dispositivos son los *smartphones* y *tablets*.
- Visualización: tecnología que se utiliza para el despliegue de una escena aumentada. Las más empleadas son proyectores de vídeo, dispositivos que se ubican en la cabeza como cascos o equipos a la altura de los ojos, sistema basado en navegadores, etc.

Hardware

Algunos de los dispositivos más destacados son:

- *Head-mounted displays*: dispositivo colocado en la cabeza o en un casco. Obtiene la información por medio de cámaras. Son utilizados tanto para realidad virtual como para realidad mixta.
- Computadores personales y *laptops*: desplazados actualmente por los móviles, fueron los primeros utilizados.
- Pantallas de mano (*smartphones* y *tablets*): dispositivos más usados para RA en estos momentos, debido a que sus sensores aportan información sobre geolocalización del usuario.
- Pantallas espaciales: hologramas, dispositivos ópticos, vídeo proyectores, etc.
- Guantes de gestos: el usuario puede manipular objetos virtuales mediante guantes con sensores.
- Dispositivos para videojuegos: originalmente los dispositivos de RA fueron creados para videojuegos, implementados por *Wii* de *Nintendo*, *Move* de *Sony PS3* o *Kinect* de *Microsoft*.

Software

Los programas de RA tienen facilidad de uso, sin altos requerimientos de conocimientos tecnológicos. Hay programas en el mercado tanto gratuitos como de pago, de costes muy variados, y sobre plataformas diversas como PC, web, móvil, o varias de ellas (**Meneses-Fernández; Martín-Gutiérrez**, 2013).

4. Realidad aumentada en televisión

Científicos, ingenieros, humanistas, etc., han utilizado como herramienta la síntesis de imágenes por ordenador para representar sus ideas de forma visual, así como para representar ambientes aún por construir en la vida real. El continuo intento por controlar y utilizar las imágenes va mucho más allá del uso de computadoras, se remonta a la fotografía y al cine.

La tendencia al uso de metáforas visuales para representar el mundo ha hecho posible, y en cierta manera necesaria, la investigación para la representación en imágenes, desarrollando para ello software y hardware cada vez más sofisticado -imagen tridimensional, por ejemplo- (**Heras-Lara; Villarreal-Benítez**, 2004).

El realismo alcanzado con el trabajo de imágenes por computadora es impresionante. Evolución que puede verse de forma clara en la evolución y calidad de imágenes para vi-



<http://www.neoteo.com/visual-synca-realidad-aumentada-sincronizada-con>

deojuegos, así como en la realidad virtual y en otras experiencias como el teatro-domo.

No sólo se copia la realidad, sino que se implementa para aportar nuevos detalles informativos, formativos o de entretenimiento.

La RA ya ha sido ensayada en artes escénicas, por medio de la incorporación de elementos audiovisuales implementados con realidad virtual expuesta entre los actores (Boj; Díaz, 2007). Igualmente en el ámbito de la prensa existen ejemplos de medios que utilizan RA en su edición electrónica, ya sea como mecanismo de marketing o para ofrecer información. Son ejemplos la revista estadounidense *Esquire*, el diario británico *Guardian.co.uk*, el magazine semanal italiano dedicado a la moda *Grazia*; la alemana *SZ Magazin* o la revista española *Fotogramas* (Meneses-Fernández; Martín-Gutiérrez, 2013). Su uso no es prolongado en el tiempo, utilizan RA de forma aislada y normalmente con fines publicitarios.

Se debe aclarar que RA no es un elemento extraño para la televisión, que ya se incorpora en muchas partes escenográficas, publicitarias, etc., de las empresas audiovisuales. La utilización de RA en televisión se ha traducido en algunos cambios en producción y postproducción, al incluir elementos generados por ordenador (Galán-Cubillo, 2008).

Como señala Galán-Cubillo (2010), en televisión se utiliza la escenografía virtual en aplicaciones de realidad virtual desde la década de 1990, cuando se realiza en algunos casos la versión inversa de la RA: en televisión se superpone imagen real sobre el entorno virtual, mientras que en la RA por definición, en su formato *video-see-through* se superpone imagen virtual sobre imagen real. Se debe entender por tanto la realidad virtual como antecedente de la RA en televisión, cuyo más claro exponente es la superposición de imágenes en *Croma-key*.

Podemos hablar de tres estadios en el uso de RA en televisión:

- 1) Primera fase: utilización de *Croma-key*, especialmente para escenografía, aunque también se ha empleado con otros fines.
- 2) Segunda fase: utilización de RA superponiendo imagen generada por ordenador a la imagen real. Se incorpora información sobre la imagen real de diferente naturaleza (RA propiamente dicha). Esta utilización es más habitual de lo

estimado. Un ejemplo es la publicidad sobre el terreno de juego en competiciones deportivas, escudos de los equipos, líneas que delimitan el terreno de juego, etc. Otros ejemplos son: uso de grafismos de forma que aparecen ante el presentador como si fueran realmente un holograma; estadísticas en tiempo real; utilización de *Google Maps* para acceder al lugar exacto de las imágenes...

3) Tercera fase: Utilización de dispositivos externos para la visualización de RA. Esta fase, aún en estado embrionario y de investigación, puede cambiar realmente la forma de ver, utilizar e interactuar con la televisión. Consiste en ofrecer información textual, hipertextual, sonora, visual y/o audiovisual por medio de dispositivos externos a la pantalla de televisión, interactuando con ella, y obteniendo información que complementa, mejora y aumenta la información facilitada por el método tradicional.

Los objetos visibles por medio de RA no podrán ser perceptibles en la televisión si no es mediante la utilización de dispositivos externos.

Los dispositivos móviles como *tablets* y *smartphones* son el mejor sistema para obtener la nueva información que facilita la RA, aunque no se deben descartar los cascos tradicionales de RA, ni tampoco el desarrollo de gafas por medio de las cuales se podría acceder a contenidos complementarios.

El material facilitado puede ser de cualquier tipo de naturaleza documental o con posibilidades de interacción con la Red. Una opción facilitada por la RA para televisión podría ser una fotografía, un texto, material audiovisual que complementa las imágenes, e incluso un hipertexto. Las posibilidades son infinitas, aunque se estima que el desarrollo más espectacular de la RA será la incorporación de material audiovisual, por lo que el "juego televisivo" se realizará en dos niveles de representación.

Se están llevando a cabo grandes logros por parte de empresas japonesas, como *NHK*, que está trabajando en la interacción por medio de una *tablet* o *smartphone* con la televisión. *Hybridcast* (o emisión híbrida) se basa en la interconexión de la emisión con información presente en la Red. Existen algunos ejemplos visuales del funcionamiento de *Hybridcast* que une el acceso a la Red con información, en algunos casos, relacionada con la RA, como la inclusión de datos estadísticos, información e identificación de los jugadores en tiempo real en competiciones deportivas, etc. Aunque aún está en modo de prueba en Japón, esta tecnología está siendo implementada por grandes empresas tecnológicas como *Sony*, *Panasonic*, *Mitsubishi*, *Toshiba* o *Sharp*. La carrera por la interconexión con la Red en tiempo real está en marcha.

<http://larepublica.pe/blogs/realidad-aumentada/tag/nhk-hybridcast>

Otra empresa japonesa, *NTT*, está elaborando un software que utiliza RA que permite ver contenidos audiovisuales a los que sólo se puede acceder por medio de una segunda pantalla, y que se tendrá que visualizar de forma sincrónica con la señal tradicional televisiva. En definitiva, dos pantallas que se tienen que ver a la vez para entender el mensaje televisivo. Las pantallas necesarias para ver esta segunda imagen serán elementos móviles como *tablets* o *smartpho-*

nes. *Visual SyncAR* también mide distancia y ángulo del telespectador a la pantalla para determinar volúmenes, perspectivas y ángulos, generando así sensaciones e ilusiones ópticas.

<http://www.neoteo.com/visual-syncar-realidad-aumentada-sincronizada-con>

Aún están por determinar, descubrir y testar todas las posibilidades que nos puede facilitar esta tecnología. Lo que parece claro es que producirá un enriquecimiento informativo y/o formativo al completar y mejorar los contenidos audiovisuales.

5. Implicaciones en los sistemas de información en televisión

Caldera-Serrano y Arranz-Escacha (2011), entre otros autores (**López-de-Quintana, 2007**; **Hidalgo-Goyanes, 2007**), han mostrado los cambios producidos en los departamentos de gestión de la información audiovisual de las empresas televisivas tras la digitalización de las redacciones y de la documentación contenida en los archivos. Señalan que los sistemas digitales fragmentan y jerarquizan los documentos audiovisuales convirtiendo las secuencias en pequeñas unidades de análisis, selección y tratamiento, y que dichas secuencias aunque heredan metadatos de la matriz cuentan con metadatos propios que los individualizan.

La evolución de los hábitos documentales también ha modificado la descripción de imágenes, ya que aunque ésta sigue siendo un conjunto de elementos referenciales que nos llevan al documento original, ahora no es necesario reseñar tantos elementos referenciales como antes, al poder acceder al documento completo de forma rápida. Para qué llevar a cabo una descripción tan exhaustiva si se cuenta con la imagen (en baja resolución) a un toque de clic.

Las descripciones se suavizan, la gestión de la información se acelera, la disponibilidad de la información para el usuario-periodista es mucho más rápida, e incluso, dicha información está accesible sin la necesaria colaboración del gestor de información.

Por medio de la RA puede integrarse información textual que puede ir más allá de los meros rótulos descriptivos o informativos

Los sistemas digitales modifican funciones como la selección del material que integra el archivo (**Caldera-Serrano; Arranz-Escacha, 2013; Giménez-Rayó, 2007**), la ingesta, la gestión de las imágenes, o el análisis documental de los elementos de acceso tradicionales onomásticos, geográficos, cronológicos y temáticos (**Caldera-Serrano; Arranz-Escacha, 2012**).

No obstante, existen aspectos aún por estudiar que podrían agilizar la descripción documental de la información audiovisual televisiva, como:

a) identificación por patrones, realizando labores de reconocimiento por patrones de formas, colores, siluetas, etc.,



<http://www.losmanyas.com/actualidad/578/noticia-televisión-realidad-aumentada>

automatizando por tanto la indicación de diferentes elementos visuales (*Google* ya cuenta con experiencias de búsquedas por imágenes) (**Pérez-Álvarez, 2006**);

b) identificación biométrica, por medio del mapeado y asignación de puntos nodales especialmente de rostros (métodos utilizados hace años por agencias gubernamentales de seguridad);

c) utilización de audio como información, mediante la transcripción a texto, pudiéndose llevar a cabo posteriormente indización automática o reconocimiento por patrones textuales (interrelacionando índices y tesauros con la descripción);

d) reconocimiento de personajes por el audio;

e) identificación de elementos de postproducción, como rotulación y otros textos por medio de lectores OCR o tecnología similar y adaptada;

f) análisis de la información textual aportada durante todo el proceso de generación del producto audiovisual para identificación y extracción de contenidos relevantes (partes de grabación, escaletas, guiones, etc.)

En este contexto de cambios constantes y evolución tecnológica, se ponen en marcha los procesos de producción audiovisual con tecnología de RA. Ello implica un análisis documental a dos niveles descriptivos que supondrá una pequeña transformación en las bases de datos que no repercutirá en los métodos de descripción documental.

Tipos documentales difundidos por RA

Con RA se puede integrar información textual que va más allá de los meros rótulos descriptivos o informativos. En caso de que dicha información sea de ampliación o de carácter complementario, se incluirá de forma íntegra la información textual facilitada. Por lo tanto, todo el texto incluido podrá ser indizado y recuperado por medio de texto libre. Este campo será utilizado igualmente para la descripción de otros materiales. Además, en un campo resumen se describirá de forma concreta el contenido de esos textos, destacando e incorporando palabras clave para identificar el contenido y potenciar su recuperación.

Otra posibilidad es la inclusión de documentación sonora. Esta opción será válida para cierto tipo de documentos, tales como documentales, algunas secciones de programas de entretenimiento o incluso películas, donde por medio de la audición de material se complementará lo aportado por las



<http://www.panoramaaudiovisual.com/2010/11/29/realidad-aumentada-en-la-noche-electoral-de-tv3>

imágenes. Desde el punto de vista descriptivo se unirá al registro documental televisivo el clip de audio así como su transcripción textual en el campo que anteriormente se ha utilizado. Igualmente se realizará un breve resumen en el mismo campo anterior. Para la transcripción de la información sonora a formato textual existen tecnologías que hacen esa labor de forma automática. En definitiva, se está transcribiendo la información para posteriormente poder recuperarla mediante dichos términos.

Se plantea como mejor opción la descripción a dos niveles: en la realidad televisiva y en la RA

Se podría contar igualmente con información en forma de enlace hipertextual. Esta opción está potenciada por la conectividad de las televisiones con la Red. Se incluirá el enlace así como un breve resumen de los contenidos.

En caso de que en un mismo documento exista más de un tipo de información de RA, se transcribirán en el mismo documento, incorporándolos al ítem documental, pudiéndose recuperar posteriormente la información por todos los elementos que la integran: RA e información televisiva.

Información audiovisual facilitada por RA: cambios en los gestores de bases de datos

La gestión de la información audiovisual se ampliará a dos niveles: a) nivel televisivo, b) nivel RA. A priori parece oportuno realizar el análisis documental de estos dos niveles por separado. No obstante, parece también imprescindible el análisis de una nueva realidad: la fusión de los elementos audiovisuales de las dos realidades. ¿Por qué? El material audiovisual de la RA puede tener una relación sincrónica o asíncrona respecto al contenido. Por ello, en algunos casos son necesarias las dos realidades para captar correctamente el contenido del mensaje, mientras que en otros momentos la relación puede ser inexistente. El primer caso (en el que sí existe relación) es el que se habrá de analizar de manera conjunta no sólo desde el punto de vista de la información audiovisual sino desde su contenido.

El análisis documental de material audiovisual cuenta con abundante literatura científica, que trata cómo analizar con-

juntos de secuencias, información, hechos, acontecimientos, o incluso emisiones. La división de la unidad documental viene así marcada por diversos parámetros ya tratados. Centrémonos por tanto en los dos niveles de análisis que conlleva la utilización de RA: la imagen propuesta por la RA y el mensaje enviado por el método tradicional (y excepcionalmente el nuevo mensaje generado por la fusión de las dos realidades).

Se plantea como mejor opción la descripción a dos niveles, donde quede reflejado en la base de datos de gestión documental cómo una noticia y/o emisión es tratada por un lado en la realidad televisiva y por otro en la RA. La descripción del material RA será idéntica al material televisivo, pero en un segundo nivel de análisis, pudiendo así analizar y describir con el mismo sistema las imágenes de ambas realidades. Ahora bien, al contar con una información en dos niveles siempre se debe determinar en qué realidad ha sido emitida, ya que se debe tener presente que la visualización de RA no se habrá instalado en los hogares en un primer momento sino probablemente tras diferentes oleadas de integración tecnológica.

Las posibilidades de acceso a la realidad virtual por medio de dispositivos móviles (teléfonos y *tablets*) así como la gran potencialidad de las gafas de RA, con la combinación de la integración de la televisión con la navegación sincronizada por la Red (potenciando la televisión a la carta), hace imparable la evolución tecnológica y el cambio de paradigma de la televisión.

Por lo tanto, consideramos dos realidades de descripción de la imagen, unificando los puntos de acceso onomásticos, cronológicos, temáticos y geográficos. Dicha descripción será idéntica, con las mismas normas y requerimientos.

La RA no supondrá un cambio drástico para los servicios de documentación de las cadenas televisivas; únicamente se tendrán que adaptar los gestores de bases de datos para incluirla

Por último, el análisis de RA más realidad televisiva, es decir, del mensaje que se desea transmitir de la unión de ambas realidades, se llevará a cabo en un campo de resumen en texto libre con dos alcances:

- descripción de los contenidos sonoros, textuales, hipertextuales y audiovisuales;
- descripción del contenido del documento televisivo-RA.

El análisis propuesto soluciona los problemas derivados de la emisión de contenidos en varias realidades, sin tener que llevar a cabo grandes cambios en los sistemas de gestión documental de las empresas audiovisuales. Y no sólo eso, sino que se incorporan nuevas posibilidades para la recuperación desde el momento en que se integra mayor información en texto libre de diferente naturaleza (textual, fotográfica, sonora, hipertextual, etc.) que podrá servir de base para la recuperación posterior. No obstante, se deberá tener

cuidado ya que dichas descripciones podrían traducirse en un aumento del ruido documental en la recuperación de información.

6. Conclusiones

La RA puede hacer cambiar algunos paradigmas de la transmisión y acceso a la información televisiva. Puede provocar grandes mutaciones e infinitas posibilidades en el acceso a nuevos contenidos, sobre todo en aparatos externos a la propia televisión. Gafas de RA, *tablets*, *smartphones*, tecnologías móviles de *Apple*, nos permiten contar con información a dos niveles accesible por la pantalla televisiva y por una pantalla auxiliar.

La forma que proponemos para gestionar dicha información es mediante la integración de estos contenidos en los sistemas de gestión de bases de datos, realizando una descripción a dos niveles de representación, en la que se podrá integrar información audiovisual, textual, sonora, fotográfica, hipertextual, o de cualquier naturaleza documental.

En definitiva, la RA no va a suponer un cambio drástico para los servicios de documentación de las cadenas televisivas; únicamente se tendrán que adaptar los gestores de bases de datos para incluir la información emitida a través de ella, que atendiendo a su naturaleza no tendrá un tratamiento especial ni diferente al realizado ya para los documentos.

Sí se traducirá en un cambio en los sistemas de difusión y, en cierta manera, de producción, ya que habrá que pensar en varios niveles audiovisuales a la hora de ofrecer productos televisivos. Todo ello constituye un nuevo campo que se abre especialmente en el terreno de la publicidad y muy especialmente como elemento integrado dentro de la filosofía de la televisión social, ya que por medio de lectores periféricos (vislumbramos las gafas y celulares como herramientas más adaptadas), no sólo se podrá acceder a contenidos sino a interconectar la televisión con los cibertelespectadores.

7. Agradecimientos

Este trabajo ha sido financiado por la *Junta de Extremadura (Consejería de Educación, Ciencia y Tecnología)* y el *Fondo Social Europeo* dentro del plan de apoyo a las actuaciones de los grupos de investigación inscritos en el catálogo de la *Junta de Extremadura*. GR10019.

8. Bibliografía

Azuma, Ronald T. (1997). "A survey of augmented reality". *Presence-teleoperators and virtual environments*, v. 6, n. 4, pp. 355-385. <http://www.cs.unc.edu/~azuma/ARpresence.pdf>

Boj, Clara; Díaz, Diego (2007). "La hibridación a escena: realidad aumentada y teatro". *Revista digital universitaria*, v. 8, n. 6, pp. 1-16. http://www.revista.unam.mx/vol.8/num6/art44/jun_art44.pdf

Caldera-Serrano, Jorge; Arranz-Escacha, Pilar (2011). "Descripción de secuencias en los archivos televisivos digitales". En: Pulgarín, Antonio; Vivas, Agustín. *Límites, fronteras y espacios comunes: encuentros y desencuentros en las ciencias de la información*. Badajoz: Abecedario, pp. 551-560. ISBN: 978 84 9978 021 4

Caldera-Serrano, Jorge; Arranz-Escacha, Pilar (2012). *Documentación audiovisual en televisión*. Barcelona: Editorial UOC, colección El profesional de la información n. 13. ISBN: 978 84 9029 982 1

Caldera-Serrano, Jorge; Arranz-Escacha, Pilar (2013). "Cambios en los métodos de selección documental en los sistemas de información digital en las televisiones". *Investigación bibliotecológica*, v. 27, n. 60, pp. 15-26. http://www.scielo.org.mx/scielo.php?pid=S0187-358X2013000200002&script=sci_arttext

Caudell, Thomas P.; Mizell, David W. (1992). "Augmented reality: An application of heads-up display technology to manual manufacturing processes". En: *Procs of the 25th Hawaii intl conf on system sciences*, v. 2, pp. 659-669. <http://dx.doi.org/10.1109/HICSS.1992.183317>

Fermín-Lobo, Marcos (2013). *Integrando realidad aumentada basada en móvil en entornos e-learning*. Oviedo: Universidad, Escuela de Ingeniería Informática. Trabajo fin de máster. <http://digibuo.uniovi.es/dspace/handle/10651/18014>

Flórez-Aristizábal, Leandro (2013). *Realidad aumentada y realidad mixta*. Cali: Universidad Autónoma de Occidente. Maestría en Ingeniería, Seminario de Investigación.

Galán-Cubillo, Esteban (2008). "Escenografía virtual en TV. Análisis del uso de escenografía virtual en la realización de un programa de televisión". *Revista latina de comunicación social*, n. 63, pp. 31-42. http://www.ull.es/publicaciones/latina/_2008/04/Galan_Cubillo.html

Galán-Cubillo, Esteban (2010). "La realidad virtual en televisión: el paso del átomo al bit". *Ámbitos*, n. 19, pp. 9-24. <http://grupo.us.es/grehcco/ambitos19/01galan.pdf>

Giménez-Rayó, Mabel (2007). *Documentación audiovisual en televisión: la selección del material*. Gijón: Trea. ISBN: 978 84 9704 320 5

Henrysson, Anders; Billinghamurst, Mark; Ollila, Mark (2005). "Face to face collaborative AR on mobile phones". En: *4th IEEE and ACM intl symposium on mixed and augmented reality*. Viena: IEEE Computer Society, pp. 35-40. ISBN: 0 7695 2459 1. <http://goo.gl/fWMWD5> <http://dx.doi.org/10.1109/ISMAR.2005.32>

Heras-Lara, Lizbeth; Villarreal-Benítez, José-Luis (2004). "La realidad aumentada: una tecnología en espera de usuarios". *Revista digital universitaria*, v. 5, n. 7, pp. 1-9. http://www.revista.unam.mx/vol.8/num6/art48/jun_art48.pdf

Hidalgo-Goyanes, Paloma (2007). "La documentación audiovisual en televisión. Los retos de la digitalización". *Boletín de la Anabad*, v. 57, n. 2, pp. 125-136.

López-de-Quintana, Eugenio (2007). "Transición y tendencias de la documentación en televisión: digitalización y nuevo mercado audiovisual". *El profesional de la información*, v. 16, n. 5, pp. 397-408. <http://www.elprofesionaldelainformacion.com/contenidos/2007/septiembre/01.pdf> <http://dx.doi.org/10.3145/epi.2007.sep.01>

Meneses-Fernández, María-Dolores; Martín-Gutiérrez, Jorge (2013). "Realidad aumentada e innovación tecnológica en prensa. La experiencia de ver y escuchar un periódico impreso". *Estudios sobre el mensaje periodístico*, v. 19, n. 1, pp. 207-211.
<http://revistas.ucm.es/index.php/ESMP/article/view/42517>

Milgram, Paul; Kishino, Fumio (1994). "A taxonomy of mixed reality visual displays". *Ieice Transactions on information and systems E77-D*, v. 12, n. 20, pp. 1321-1329.
http://etclub.mie.utoronto.ca/people/paul_dir/IEICE94/ieice.html

Milgram, Paul; Takemura, Hauro; Utsumi, Akira; Kishino, Fumio (1994). "Augmented reality: a class of displays on the reality-virtuality continuum". *Spie*, v. 2351, pp. 282-292.
http://web.cs.wpi.edu/~gogo/hive/papers/Milgram_Takemura_SPIE_1994.pdf

Olmedo, Héctor; Augusto, Jorge (2012). "From augmented reality to mixed reality: technological options". En: *Procs of the 13th Intl conf on interacción persona-ordenador (Interac-*

cion'12). Elche, España. New York: Assoc. Computing Machinery. ISBN: 978 1 4503 1314 8
<http://dx.doi.org/10.1145/2379636.2379671>

Parhizkar, Behrang; Al-Modwahi, Ashraf-Abbas M.; Lashkari, Arash-Habibi; Bartaripou, Mohammad-Mehdi; Babae, Hossein-Reza (2011). "A survey on web-based AR applications". *International journal of computer science issues (IJCS-I)*, v. 8, n. 4.
<http://arxiv.org/pdf/1111.2993.pdf>

Pérez-Álvarez, Sara (2006). "Aproximación al estudio de los sistemas de recuperación de imágenes 'CBIR' desde el ámbito de la documentación". *Documentación de las ciencias de la información*, v. 29, pp. 301-315.
<http://revistas.ucm.es/index.php/DCIN/article/view/DCIN0606110301A>

Sutherland, Ivan E. (1968). "A head-mounted three dimensional display". En: *Procs of the Afips*. New York: ACM, pp. 757-764,
<http://dx.doi.org/10.1145/1476589.1476686>

Revistas de comunicación, documentación y biblioteconomía indexadas en el Social Sciences Citation Index, de Thomson Reuters, que aceptan artículos en español

Impact factor 2013

Revista española de DOCUMENTACIÓN CIENTÍFICA



0,717

Revista española de documentación científica (España)

<http://redc.revistas.csic.es/index.php/redc>

INFORMATION RESEARCH
an international electronic journal
ISSN 1368-1613

0,660

Information research (Reino Unido)

<http://informationr.net/ir>

El profesional de la **información**

0,402

El profesional de la información (España)

<http://www.elprofesionaldelainformacion.com>

Comunicar
Revista de Comunicación y Documentación
ISSN 1134-3473

0,350

Comunicar (España)

<http://www.revistacomunicar.com>

Transinformação

0,083

Transinformação (Brasil)

http://www.scielo.br/scielo.php?script=sci_serial&pid=0103-3786

I&S

0,080

Informação & sociedade: estudos (Brasil)

<http://www.ies.ufpb.br/ojs2/index.php/ies>

Investigación Bibliotecológica

0,062

Investigación bibliotecológica (México)

<http://www.revistas.unam.mx/index.php/ibi>

Indexadas pero sin factor de impacto en 2013

Perspectivas em ciência da informação (Brasil)
<http://portaldeperiodicos.eci.ufmg.br/index.php/pci>

PERSPECTIVAS

Comunicación y sociedad (España)
<http://www.unav.es/lfcom/comunicacionsociedad/les>

COMMUNICATION & SOCIETY

Estudios sobre el mensaje periodístico (España)
<http://www.ucm.es/info/emp>

Estudios sobre el Mensaje Periodístico

noviembre 2014-noviembre 2015

12-14 de noviembre de 2014

VII CONGRESO NACIONAL DE BIBLIOTECAS PÚBLICAS. Bibliotecas públicas, conectados contigo
Badajoz
Subdirección General de Coordinación Bibliotecaria, Ministerio de Educación, Cultura y Deporte
<http://www.mecd.gob.es/congresobp-vii-14>

12-14 de noviembre de 2014

IV REUNIÓN LATINOAMERICANA DE ANÁLISIS DE REDES SOCIALES
La Plata, Argentina
Universidad Nacional de La Plata
<http://encuentroredes.wordpress.com>

18-19 de noviembre de 2014

2ND INTL SYMPOSIUM DIGITAL FRONTIERS
Fez, Marruecos
Paragraphe, Geriico, Arts des images & art contemporain, LabEx Arts-H2M, Higher School of Engineering in Applied Sciences, Idefi Créatic, Esisa.
<http://numerique.paragraphe.info>

19-20 de noviembre de 2014

LIVRE POST-NUMÉRIQUE: historique, mutations et perspectives
Fez, Marruecos
CiTu Paragraphe, Esisa, Europia, GERiico, IDEFI, LEDEN MSH.
<http://cide17.europia.org>

20-21 de noviembre de 2014

2014 INTL CONF ON JOURNALISM AND MASS COMMUNICATIONS (ICJMC) Milán
International Economics Development and Research Center (IEDRC)
<http://www.icjmc.org>



21 de noviembre de 2014

VI JORNADA TÉCNICA. Documentación, política y activismo
Salamanca
Asociación de Archiveros de Castilla y León (ACAL)
<http://www.acal.es/index.php/eventos-jornadas/item/1401>

24-26 de noviembre de 2014

10TH INTL CONF ON KNOWLEDGE MANAGEMENT (ICKM2014)
5TH INTL SYMPOSIUM ON INFORMATION MANAGEMENT IN A CHANGING WORLD Research data management and knowledge discovery
Antalya, Turkey
<http://imcw2014.bilgiyonetimi.net>

25-28 de noviembre de 2014

INTERFACES. Foro intl. Correlaciones, comunicación y redes
Colima, México
<http://eventos.ucol.mx/interfaces>

26-28 de noviembre de 2014

MTSR 2014. 8th Metadata and semantics research conf
Karlsruhe
Karlsruhe Institute of Technology (KIT)
<http://www.mtsr-conf.org>

1-3 de diciembre de 2014

SWIB CONF. Semantic web in libraries
Bonn
North Rhine-Westphalian Library Service Centre; German National Library of Economics / Leibniz Information Centre for Economics
<https://swib.org/swib14>

12-13 de enero de 2015

FORCE2015. Research communication and e-scholarship conf
Oxford
University of Oxford
<https://www.force11.org/meetings/force2015>



14-15 de enero de 2015

2ND INTL CONF ON INFORMATION, COMMUNICATION AND COMPUTER NETWORKS (ICI2CN)
Londres
City University
<http://www.ici2cn.com>

15-16 de enero de 2015

2015 INTL CONF ON KNOWLEDGE (ICK 2015)
Portsmouth, Reino Unido
American Society for Research
<http://www.ick.org>

19 de enero de 2015

ICSTI 2015. Annual members' meeting
Berlín
International Council for Scientific and Technical Information (ICSTI)
<http://www.icsti.org>

20-21 de enero de 2015

APE-ACADEMIC PUBLISHING IN EUROPE 10. Web25: The road ahead. Exploring the future of scholarly communication & academic publishing
Berlín
<http://www.ape2015.eu>

28-30 de enero de 2015

23TH BOBCATSSS
Brno, República Checa
Euclid (European Association for Library and Information Education and Research); Masaryk University (Rep. Checa); Tampere University (Finlandia).
<http://bobcatsss2015.com>

6-7 de febrero de 2015

1ª CONF INTL SOBRE INDUSTRIA Y MERCADO DE LA INFORMACIÓN
Madrid

El profesional de la información; Biblioteca de la Universidad Complutense de Madrid; Casa del Lector - Fundación Germán Sánchez Ruipérez
<http://www.confimi.info>



9-12 de febrero de 2015

10TH INTL DIGITAL CURATION CONF

Londres

Digital Curation Centre (DCC), University of Glasgow
<http://www.dcc.ac.uk/events/idcc15>

16-17 de febrero de 2015

ASSOCIATION OF SUBSCRIPTION AGENTS (ASA) ANNUAL CONF 2015. Research-to-researcher

Londres

Association of Subscription Agents Ltd.
<http://www.subscription-agents.org/conferences>

27 de febrero de 2015

1ST INTL CONF ON ACADEMIC COMMUNICATION JOURNALS (ICACJ-2015)

Barcelona

Department of Communication/Faculty of Communication - Universitat Pompeu Fabra

5-6 de marzo de 2015

6TH EDGE2015 Conf. Changing challenge to opportunity

Edinburgh

Edinburgh City Libraries
<http://edgeconference.co.uk>



11-13 de marzo de 2015

VI JORNADAS DE REPOSITORIOS

XIV WORKSHOP DE REBIUN SOBRE PROYECTOS DIGITALES

Córdoba

Universidad de Córdoba, GT Repositorios de Rebiun

Web aún no disponible

Reme Melero, rmelero@iata.csic.es

14-16 de marzo de 2015

13TH INTL CONF E-SOCIETY 2015

Madeira

International Association for Development of the Information Society

<http://www.esociety-conf.org>

19-20 de marzo de 2015

ICEIT 2015. 4th Intl conf on educational and information technology
Florencia

American Society for Research

<http://www.iceit.org>

30 de marzo-1 de abril de 2015

UKSG 38TH ANNUAL CONF AND EXHIBITION

Glasgow

UK Serials Group

<http://www.uksg.org/event/conference15>

8-10 de abril de 2015

LILAC 2015. Librarians' information literacy annual conf

Newcastle, Reino Unido

Cilip Information Literacy Group

<http://www.lilacconference.com/WP>

9-10 de abril de 2015

7TH EUROPEAN CONF ON INTELLECTUAL CAPITAL (ECIC 2015)

Cartagena, España

Universidad Politécnica de Cartagena

<http://academic-conferences.org/ecic/ecic2015/ecic15-home.htm>

19-24 de abril de 2015

KESA 2015. Intl workshop on knowledge extraction and semantic annotation

Barcelona

International Academy, Research, and Industry Association (IARIA)

<http://www.iaria.org/conferences2015/KESA.html>



20-22 de abril de 2015

EMERGING TECHNOLOGIES IN ACADEMIC LIBRARIES INTL CONF

Trondheim, Noruega

Norwegian University of Science and Technology (NTNU), Library

<http://www.emtacl.com>

27-30 de abril de 2015

ICEIS 2015. 17th Intl conf on enterprise information systems

Barcelona

Institute for Systems and Technologies of Information, Control and Communication

<http://www.iceis.org>



7-8 de mayo de 2015

5ª CONF INTL SOBRE CALIDAD DE REVISTAS DE CIENCIAS SOCIALES Y HUMANIDADES (CRECS 2015)

Murcia

Facultad de Comunicación y Documentación; Grupo ThinkEPI; El Profesional de la Información

<http://www.thinkepi.net/crecs2015>



11-14 de mayo de 2015

13TH ANNUAL INTL CONF ON COMMUNICATION AND MASS MEDIA

Atenas

Athens Institute for Education and Research

<http://www.atiner.gr/media.htm>

13-15 de mayo de 2015

RCIS 2015. 9th IEEE Intl conf on research challenges in information science

Atenas

<http://rcis-conf.com/rcis2015>

18-19 de mayo de 2015

ICIII 2015. 4th Intl conf on industrial and intelligent information

Roma

<http://www.iciii.org>

18-22 de mayo de 2015

WWW 2015. 24th Intl world wide web conf

Florenia

International World Wide Web Conferences Steering Committee (IW3C2)

<http://www2015.wwwconference.org>



19-21 de mayo de 2015

14TH INTL SYMPOSIUM OF INFORMATION SCIENCE (ISI)

Zadar, Croacia

German Academic Association for Information Science (Hochschulverband Informationswissenschaft e. V.); Department of Information Sciences, University of Zadar

<http://isi2015.de/?lang=en>

26-29 de mayo de 2015

QQML2015. 7th Intl conf on qualitative and quantitative methods in libraries

París

IUT Université Paris Descartes

<http://isast.org/qqml2015.html>

28-30 de mayo de 2015

XIV JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN – FESABID 2015

Gijón

<http://www.fesabid.org>

31 de mayo-4 de junio de 2015

ESWC. 12th European semantic web conf

Portoroz, Slovenia

<http://2015.eswc-conferences.org>

4-5 de junio de 2015

III SEMINARIO INTL DE DOCENCIA E INVESTIGACIÓN EN BIBLIOTECONOMÍA Y DOCUMENTACIÓN (LIS-ER)

Barcelona

Facultat de Biblioteconomia i Documentació, Universitat de Barcelona.

<http://bd.ub.edu/liser>



8-10 de junio de 2014

WEB ARCHIVES AS SCHOLARLY SOURCES: Issues, Practices and Perspectives

Aarhus, Dinamarca

Resaw (Research Infrastructure for the Study of Archived Web Materials)

<http://resaw.eu>

17-19 de junio de 2015

OAI9. CERN Workshop on innovations in scholarly communication

Ginebra

European Organization for Nuclear Research (CERN) Organizing Committee

<http://indico.cern.ch/event/332370/overview>

23-26 de junio de 2015

5TH INFORMATION: INTERACTIONS AND IMPACT (i³)

Aberdeen, Reino Unido

Robert Gordon University

<http://www.rgu.ac.uk/research/conferences/i-2015>



24-27 de junio de 2015

KM conf 2015

Katowice, Polonia

International Institute for Applied Knowledge Management

<http://www.iiakm.org/conference/index.php>

25-26 de junio de 2015

LIBER'S ANNUAL CONF

Londres

<http://libereurope.eu/news/libers-2015-annual-conference>

28 de junio-2 de julio de 2015

44TH ANNUAL INTL CONF & 17TH INTL FORUM ON RESEARCH IN SCHOOL LIBRARIANSHIP

Maastricht, Holanda

International Association of School Librarianship (IASL)

<http://iasl2015.org>

29 de junio-4 de julio de 2015

15TH ISSI2015

Estambul

Intl Society for Informetrics and Scientometrics (ISSI); Bogazici University

<http://www.issi2015.org/en>



13-15 de julio de 2015

WIMS 2015. 5th Intl conf on web intelligence, mining and semantics
Limassol, Chipre

University of Cyprus

<http://cyprusconferences.org/wims2015>

13-16 de julio de 2015

LIBEURO 2015. The European conf on literature and librarianship
Brighton, Reino Unido

The International Academic Forum

<http://iafor.org/iafor/conferences/libeuro2015>

20-22 de julio de 2015

11th NORTHUMBRIA INTL CONF ON PERFORMANCE MEASUREMENT IN LIBRARIES AND INFORMATION SERVICES
Edimburgo

University of York; National Library of Scotland

<http://www.york.ac.uk/about/departments/support-and-admin/information-directorate/northumbria-conference>

27-30 de julio de 2015

2ND ANNUAL INTL CONF ON LIBRARY AND INFORMATION SCIENCE
Atenas

Athens Institute for Education and Research

<http://www.atiner.gr/library.htm>

28-30 de julio de 2015

SCIENCE AND INFORMATION CONFERENCE (SAI)
Londres

<http://thesai.org/SAIConference2015>



15-21 de agosto de 2015

IFLA 2015. WORLD LIBRARY AND INFORMATION CONGRESS
Ciudad del Cabo

<http://conference.ifla.org/ifla81>



2-4 de septiembre de 2015

20TH SCIENCE AND TECHNOLOGY INDICATORS CONFERENCE. RESEARCH ORGANIZATIONS UNDER SCRUTINY. New indicators and analytical results

Lugano

Università della Svizzera Italiana

<http://www.sti2015.usi.ch>



9-11 de septiembre de 2015

ALPSP Intl Conf 2015

Londres

Association of Learned and Professional Society Publishers

<http://www.alpssp.org/Ebusiness/TrainingAndEvents/ALPSPInternationalConference.aspx?ID=394>

30 de septiembre-2 de octubre de 2015

33^a FERIA INTL DEL LIBRO

Madrid

Federación de Gremios de Editores de España; Ifema

<http://www.salonliber.es>

1-3 de octubre de 2015

14TH IFLA INTERLENDING AND DOCUMENT SUPPLY CONF

Estambul

IFLA Document Delivery and Resource Sharing Section

<http://www.ifla.org/docdel>

14-18 de octubre de 2015



FRANKFURTER BUCHMESSE. FERIA DEL LIBRO

Frankfurt

<http://www.book-fair.com/en>

16-17 de noviembre de 2015

VII ENCUESTRO IBÉRICO EDICIC

Madrid

Asociación de Educación e Investigación en Ciencia de la Información de Iberoamérica y del Caribe (EDICIC), Facultad de Ciencias de la Documentación de la Universidad Complutense de Madrid

<http://edicic2015.org.es/index.php/inicio/edicic2015>

19-20 de noviembre de 2015

XII CONGRESO DEL CAPÍTULO ESPAÑOL DE ISKO
II CONGRESO ISKO ESPAÑA-PORTUGAL

Murcia

Universidad de Murcia

<http://www.iskoiberico.org>



EVALUADORES DE LA REVISTA

El profesional de la información 2013-2014

- Abadal, Ernest.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Aceituno-Aceituno, Pedro.** Universidad a distancia de Madrid.
- Afuera, Ángeles.** Cadena SER, Madrid.
- Aguillo, Isidro F.** Centro de Ciencias Humanas y Sociales. Consejo Superior de Investigaciones Científicas, Madrid.
- Alberich-Pascual, Jordi.** Facultad de Comunicación y Documentación. Universidad de Granada.
- Albiñana, Ricardo.** Universitat Oberta de Catalunya.
- Aleixandre-Benavent, Rafael.** Instituto de Historia de la Ciencia y Documentación López Piñero, Valencia.
- Alvarez-Monzoncillo, José María.** Universidad Rey Juan Carlos, Madrid.
- Alvim, Luisa.** Casa Camilo- Museu Centro de Estudos. S. Miguel de Seide, Portugal.
- Amat, Carlos B.** Ingenio (CSIC-UPV), Institute of Innovation and Knowledge Management. Universidad Politécnica de Valencia.
- Arroyo-Vázquez, Natalia.** Fundación Germán Sánchez Rui-pérez. Peñaranda de Bracamonte, Salamanca.
- Astigarraga, Eneko.** CodeSyntax, Universidad de Deusto. Bilbao.
- Baiget, Tomàs.** El profesional de la información, Barcelona.
- Benavides, David.** Universidad de Sevilla.
- Bermejo-Berros, Jesús.** Universidad de Valladolid.
- Bezuntea-Valencia, Ofa.** Universidad del País Vasco.
- Borrego, Ángel.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Bueno-de-la-Fuente, Gema.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Bustamante-Díaz, Jesús.** Centro Europeo para el Desarrollo de la Formación Profesional (Cedefop), Tesalónica, Grecia.
- Cabezas-Clavijo, Álvaro.** Facultad de Comunicación y Documentación. Universidad de Granada.
- Campos-Freire, Francisco.** Universidad de Santiago de Compostela.
- Calderón, Andoni.** Biblioteca. Universidad Complutense de Madrid.
- Canals, Agustí.** Estudios CC de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.
- Carvalho, Andrea.** Universidade Federal do Rio Grande do Norte. Brasil.
- Casero, Andreu.** Departament de Ciències de la Comunicació. Universitat Jaume I, Castellón.
- Cerdá-Díaz, Julio.** Ayuntamiento de Arganda del Rey (Madrid), Universidad Carlos III de Madrid.
- Cobarsí-Morales, Josep.** Estudios CC de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.
- Codina, Lluís.** Departamento de Comunicación. Universitat Pompeu Fabra, Barcelona.
- Colorado Castellary, Arturo.** Universidad Complutense de Madrid.
- Córdoba, Saray.** Universidad de Costa Rica.
- Costas-Comesaña, Rodrigo.** Centre for Science and Technology Studies (CWTS), Leiden University, Países Bajos.
- Curto-Gordo, Víctor.** Universitat Ramon Llull, Barcelona.
- D'Alòs-Moner, Adela.** Consultora. Barcelona.
- De-Inclán-Sánchez, María.** Banco de España, Madrid.
- De-la-Cuadra, Elena.** Facultad de Ciencias de la Información. Universidad Complutense de Madrid.
- Del-Fresno-García, Miguel.** UNED, Madrid.
- Del-Valle-Gastaminza, Félix.** Facultad de Ciencias de la Información. Universidad Complutense de Madrid.
- Delgado-López-Cózar, Emilio.** Facultad de Comunicación y Documentación. Universidad de Granada.
- Díaz-Noci, Javier.** Departamento de Comunicación. Universitat Pompeu Fabra, Barcelona.
- Docampo, Javier.** Museo del Prado, Madrid.
- Echeverría, Mercedes.** Facultad de Medicina de la Universidad Autónoma, Madrid.
- Eito-Brun, Ricardo.** Grupo GMV, Madrid.
- Estivill-Rius, Assumpció.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Fernández-Molina, Juan Carlos.** Facultad de Comunicación y Documentación. Universidad de Granada.
- Fernández-y-Fernández-Cuesta, Paz.** Instituto Juan March de Estudios e Investigaciones, Madrid.
- Fernández-Travieso, Carlota.** Universidade da Coruña.
- Ferran-Ferrer, Núria.** Estudios CC de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.
- Ferraro, Gabriela.** College of Engineering and Computer Science, Australian National University. Canberra. Australia.
- Ferrer-Sapena, Antonia.** Depto de Comunic. Audiovisual, Docum. e Historia del Arte. Univ. Politécnica de Valencia.

Franganillo, Jorge. Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.

Fuentes-Pujol, Eulàlia. Facultat de Ciències de la Comunicació. Universitat Autònoma de Barcelona.

Galán-Cubillo, Esteban. Universitat Jaume I de Castelló.

Gallardo-Rodríguez, Xavier. Universitat Autònoma de Barcelona.

Gallego, Cristina. Biblioteca. Universidad Complutense de Madrid.

García-de-Torres, Elvira. Universidad CEU Cardenal Herrera, Valencia.

García-Jiménez, Antonio. Facultat de Ciències de la Comunicació. Universidad Rey Juan Carlos, Madrid.

García-Marco, Francisco-Javier. Departamento de Ciencias de la Documentación. Universidad de Zaragoza.

Giménez-Rayó, Mabel. SPW, Madrid.

Gómez-García, Juan-Antonio. Facultat de Derecho. UNED, Madrid.

Gómez-Hernández, José-Antonio. Facultat de Comunicació i Documentación. Universidad de Murcia.

González-Albo-Manglano, Borja. Centro de Ciencias Humanas y Sociales. Consejo Superior de Investigaciones Científicas. Madrid.

González-Caro, Cristina. Universidad Autónoma de Bucaramanga. Colombia.

González-Guitián, Carlos. Servizo Galego de Saúde. A Coruña.

González-Sánchez, José-Luis. Universitat de Lleida.

González-Fernández-Villavicencio, Nieves. Universidad Pablo de Olavide, Sevilla.

Granell-Canut, Carlos. European Commission. Joint Research Centre, Ispra, Italia.

Guallar, Javier. Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.

Hassan, Yusef. Grupo SCImago. Universidad de Granada.

Hernández-Pérez, Tony. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.

Hernández-Carrascal, Francisca. Digibis, Madrid.

Hernández-Sánchez, Hilario. Fundación Germán Sánchez Ruipérez, Salamanca.

Herrero-Pascual, Cristina. Universidad de Murcia.

Hidalgo, Paloma. RTVE, Madrid.

Jiménez-Morales, Mònika. Universitat Pompeu Fabra, Barcelona

Jiménez-Zafra, Purificación. RTVE, Madrid.

Juanatey-Boga, Óscar. Universidad de La Coruña.

Keefer-Riva, Alice. Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.

Labastida, Ignasi. Oficina de Difusión del Conocimiento. Universidad de Barcelona.

Lalueza, Ferran. Estudios CC de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.

Lascurain, María-Luisa. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid

Leiva-Aguilera, Javier. Catorze.com, Barcelona.

Llorens-Maluquer, Carles. Facultat de Comunicació de la Universitat Autònoma de Barcelona.

López-Cepeda, Ana-María. Universidad de Castilla La Mancha.

López-de-Quintana-Sáenz, Eugenio. Antena 3 TV, Madrid.

López-de-Solís, Iris. RTVE, Madrid.

López-del-Ramo, Joaquín. Universidad Rey Juan Carlos.

Magán-Wals, José-Antonio. Biblioteca, Universidad Complutense de Madrid.

Mañá-Terré, Teresa. Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.

Mañana-Rodríguez, Jorge. Centro de Ciencias Humanas y Sociales. Consejo Superior de Investigaciones Científicas. Madrid.

Marcos, Mari-Carmen. Departamento de Comunicación. Universitat Pompeu Fabra, Barcelona.

Martínez-Fernández, Valentín-Alejandro. Facultat de Comunicació. Universidad de La Coruña.

Martínez-Martínez, Silvia. Universitat Oberta de Catalunya, Barcelona.

Marzal, Miguel-Ángel. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.

Masip, Pere. Facultat de Comunicació Blanquerna. Universitat Ramon Llull, Barcelona.

Melero, Remedios. Inst. de Agroquímica y Tecnología de Alimentos. CSIC, Valencia.

Mendo-Carmona, Concepción. Facultat de Ciències de la Documentación. Universidad Complutense de Madrid.

Mendoza-García, Isabel. Biblioteca de la Escuela Técnica Superior de Ingenieros Industriales de la Universidad Politécnica de Madrid.

Micó-Sanz, Josep-Lluís. Facultat de Comunicació Blanquerna. Universitat Ramon Llull, Barcelona.

Mínguez, Norberto. Facultat de Ciències de la Informació. Universidad Complutense de Madrid.

Morales-García, Ana-María. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.

Morato, Jorge. Departamento de Informática. Universidad Carlos III de Madrid.

Moreiro-González, José-Antonio. Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.

Muñoz-Cañavate, Antonio. Facultat de Biblioteconomía y Documentación. Universidad de Extremadura.

Moya, Eva. S21sec, Instituto Ciencias Forenses - Autónoma de Madrid, Universitat Oberta de Catalunya.

- Muñoz-de-Solano-Palacios, Bárbara.** Museo Nacional Centro de Arte Reina Sofía, Madrid.
- Muñoz-Tinoco, Concha.** Biblioteca. Hospital Ramón y Cajal, Madrid.
- Muriel-Torrado, Enrique.** Universidade Federal de Santa Catarina, Brasil.
- Noguera, José-Manuel.** Facultad de CC Sociales y de la Comunicación. Universidad Católica San Antonio, Alicante.
- Oliveira-dos-Santos, Andrea.** Universitat de Girona.
- Orduña-Malea, Enrique.** Escuela Técnica Superior de Informática. Universidad Politécnica de Valencia.
- Ortiz-Repiso, Virginia.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Ortiz-Sobrino, Miguel-Ángel.** Facultad de Ciencias de la Información. Universidad Complutense de Madrid.
- Ortoll, Eva.** Estudios de Ciencias de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.
- Otero, María-Luisa.** RTVE, Madrid.
- Pacios-Lozano, Ana-Reyes.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Palacios-Madrid, Vicente.** Universidad Carlos III de Madrid.
- Palafox-Parejo, Manuela.** Biblioteca. Universidad Complutense de Madrid.
- Pascual-Miguel, Félix-José.** E. T. S. Ingenieros de Telecomunicación. Universidad Politécnica de Madrid.
- Pascual, Víctor.** Consultor en Ciencias de la Computación y Comunicación Digital
- Pastor-Sánchez, Juan-Antonio.** Facultad de Comunicación y Documentación. Universidad de Murcia.
- Pedraza-Gracia, Manuel-José.** Departamento de Ciencias de la Documentación. Universidad de Zaragoza.
- Peón, Jaime.** Ministerio de Fomento. Universidad Complutense de Madrid.
- Pérez-Lorenzo, Belén.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Pérez-Montoro, Mario.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Peset, Fernanda.** Depto de Comunic. Audiovisual, Docum. e Historia del Arte. Univ. Politécnica de Valencia.
- Pinto, Adilson-Luiz.** Universidade Federal de Santa Catarina, Brasil.
- Robinson-García, Nicolás.** Facultad de Comunicación y Documentación. Universidad de Granada
- Robledano-Arillo, Jesús.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Rodríguez-Gairín, Josep-Manuel.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Rodríguez-Yunta, Luis.** Centro de Ciencias Humanas y Sociales. Consejo Superior de Investigaciones Científicas. Madrid.
- Roig-Telo, Antoni.** Universitat Oberta de Catalunya, Barcelona.
- Rovira, Cristòfol.** Departamento de Comunicación. Universitat Pompeu Fabra, Barcelona.
- Rueda-Ramírez, Pedro.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Ruiz-Caballero, Carlos.** Facultat de Comunicació Blanquerna. Universitat Ramon Llull, Barcelona.
- Salaberria Lizarazu, Ramón.** Consultor independiente. México.
- Salaverría, Ramón.** Facultad de Comunicación. Universidad de Navarra.
- San-José-Montano, Blanca.** Hospital Universitario de Móstoles, Madrid
- Sánchez-Jiménez, Rodrigo.** Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid.
- Sánchez-Navarro, Jordi.** Estudios CC de la Información y de la Comunicación. Universitat Oberta de Catalunya, Barcelona.
- Sanz-Casado, Elías.** Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- Scolari, Carlos.** Departamento de Comunicación. Universitat Pompeu Fabra, Barcelona.
- Sevillano-Pintado, Olga.** Museo Nacional Centro de Arte Reina Sofía, Madrid.
- Spinak, Ernesto.** Consultor. Uruguay.
- Tardón, Eugenio.** Biblioteca. Universidad Complutense de Madrid.
- Tejedor, Santiago.** Departamento de Periodismo y Ciencias de la Comunicación. Universitat Autònoma de Barcelona.
- Tena-Millán, Joaquín.** Universitat Pompeu Fabra, Barcelona.
- Térmens-Graells, Miquel.** Facultat de Biblioteconomia i Documentació. Universitat de Barcelona.
- Tejada-Artigas, Carlos.** Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid.
- Torres-Salinas, Daniel.** Centro de Investigación Médica Aplicada, Universidad de Navarra.
- Torres-Padrosa, Víctor.** Institute of Informatics and Applications. Universitat de Girona.
- Tramullas, Jesús.** Departamento de Ciencias de la Documentación. Universidad de Zaragoza.
- Trujillo, Javier.** Consultoría informática, Madrid.
- Urueña-López, Alberto.** Universidad Politécnica de Madrid.
- Valls-Pasola, Anna.** Institut del Teatre, Barcelona.
- Villaseñor, Isabel.** Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid.
- Vivas-Moreno, Agustín.** Facultad de Biblioteconomía y Documentación. Universidad de Extremadura.
- Voces-Merayo, Ramón.** Facultat de Ciències de la Comunicació. Universitat Autònoma de Barcelona.
- Xifra-Triadú, Jordi.** Universitat Pompeu Fabra, Barcelona.
- Yegros, Alfredo.** Centre for Science and Technology Studies (CWTS), Leiden University, Países Bajos.
- Zozaya-Montes, Leonor.** Fundação para a Ciência e a Tecnologia, Coimbra, Portugal.

PUBLICACIONES EL PROFESIONAL DE LA INFORMACIÓN

REVISTA EL PROFESIONAL DE LA INFORMACIÓN

A partir de 2015 sólo online

- Suscripción anual institucional 130 €
- Suscripción anual individual 79 €

Números anteriores en papel

2014 v. 23

- n. 1: Documentación audiovisual 30 €
- n. 2: Políticas de información 30 €
- n. 3: Visualización de información 30 €
- n. 4: Altmétricas 30 €
- n. 5: Humanidades digitales 30 €
- n. 6: Big data y analítica digital 30 €

2013 v. 22

- n. 6: Formación y aprendizaje 30 €
- n. 5: Gestión de contenidos 30 €
- n. 4: Economía de la información 30 €
- n. 3: Bibliotecas y documentación de museos 30 €
- n. 2: Educación y biblioteca 30 €
- n. 1: Soportes digitales 30 €

ANUARIO THINKEPI (versión digital)

Tarifas institucionales

- Anuario 2015 75 €
- Anuario 2014 68 €

Tarifas individuales (particulares)

- Anuario 2015 45 €
- Anuario 2014 35 €

<http://recyt.fecyt.es/index.php/ThinkEPI/index>

EPI + ANUARIO THINKEPI (suscripción conjunta)

- Suscripción anual institucional 190 €
- Suscripción anual individual 110 €

Formulario de compra EPI y Anuario ThinkEPI

<http://www.elprofesionaldelainformacion.com/suscripciones.php>

El profesional de la
información

ANUARIO
Think
EPI

En el caso de las publicaciones en papel hay que añadir los gastos de envío

LIBROS EL PROFESIONAL DE LA INFORMACIÓN

(Editorial UOC)

- 1. Tecnologías de la web semántica 12,00 €
Juan-Antonio Pastor
- 2. La revolución del libro electrónico 11,50 €
José-Antonio Cordón
- 3. Sistemas de información en la empresa 11,50 €
Josep Cobarsí-Morales
- 4. Información en el móvil 12,00 €
Natalia Arroyo-Vázquez
- 5. Acceso abierto a la ciencia 12,00 €
Ernest Abadal
- 6. Fuentes de información médica 12,00 €
Pablo Medina-Aguerrebera
- 7. Gestión de la reputación online 11,50 €
Javier Leiva-Aguilera
- 8. *Wikipedia* de la A a la W 12,50 €
Tomás Saorín
- 9. Etiquetar en la web social 12,00 €
Raquel Gómez-Díaz
- 10. Mejorar las búsquedas de información 12,00 €
Silvia Argudo y Amadeu Pons
- 11. Clubes de lectura 11,50 €
Óscar Carreño
- 12. Plan social media y community manager 13,50 €
Julián Marquina-Arenas
- 13. Documentación audiovisual en televisión 12,00 €
Jorge Caldera y Pilar Arranz
- 14. Gestión de documentos en la e-administración 12,00 €
Elisa García-Morales
- 15. *El film researcher* 12,00 €
Iris López-de-Solís
- 16. Preservación digital 12,00 €
Miquel Térmens
- 17. Gestión de contenidos 13,00 €
Ricardo Eito-Brun
- 18. Documentación fotográfica 14,00 €
Juan-Miguel Sánchez-Vigil y Antonia Salvador-Benítez
- 19. Documentación cinematográfica 12,00 €
Elena De la Cuadra
- 20. Archivos 14,00 €
Ramón Alberch-Figuera
- 21. Inteligencia en redes sociales 11,00 €
Eva Moya
- 22. Bibliotecas escolares 10,00 €
Concepción M^a Jiménez-Fernández y Raúl Cremades-García
- 23. Marca y comunicación empresarial 11,00 €
Pablo Medina-Aguerrebera
- 24. El content curator 14,00 €
Javier Guallar y Javier Leiva-Aguilera
- 25. Gestión de la calidad en la biblioteca 13,50 €
Nuria Balagué y Jarmo Saarti
- 26. Innovación en bibliotecas 13,50 €
Isabel Riaza
- 27. La intranet social 13,50 €
Ana Carrillo Pozas
- 28. Los impresos antiguos 12,50 €
Jon Zabala

Formulario de compra libros EPI-UOC

<http://www.elprofesionaldelainformacion.com/suscripciones.php>

Información y pedidos: Isabel Olea / epi.iolea@gmail.com / Tel.: +34-608 491 521

<http://www.elprofesionaldelainformacion.com>

<http://www.thinkepi.net>



EL PROFESIONAL DE LA INFORMACIÓN (EPI)

es una revista bimestral, internacional, científica y profesional en español, inglés, francés y portugués, líder en el área de información, comunicación, bibliotecas y nuevas tecnologías de la información



La revista

EPI fue la primera revista sobre Bibliotecas, Documentación, Información y Comunicación en español indizada por las bases de datos *WoS Social Sciences Citation Index (ISI, de Thomson-Reuters)* y *Scopus (de Elsevier)*, entre otras tales como *Inspec, Lista, Lisa, Isoc, Pascal, Francis, Academic Search...* Tiene *SJR (SCImago journal rank)* desde 2007 y Factor de impacto desde 2008. Desde 2014 está reconocida por la *Aeres (Francia)*, y en ese mismo año ha conseguido el sello de calidad de revistas científicas españolas de la *Fecyt*.

La suscripción anual incluye 6 números en papel y el acceso online a toda la colección desde el año 2000. Los números retrospectivos hasta 1992 pueden consultarse en acceso abierto en la web:

<http://www.elprofesionaldeinformacion.com>

Materias tratadas

Archivos administrativos

Arquitectura de la información

Bases y bancos de datos

Bibliometría y cibermetría

Bibliotecas especializadas

Bibliotecas públicas

Bibliotecas universitarias

Buscadores

Comunicación científica

Comunicación social

Edición electrónica

Gestión de contenidos

Gestión del conocimiento

Indicadores

Información científica

Información en la empresa

Inteligencia competitiva

Legislación y copyright

Libros-e y revistas-e

Marketing y promoción

Mercado de la información

Multimedia, audiovisual

Ontologías, tesauros, metadatos

Open access

Open software

Patentes

Periodismo

Redes sociales

Sistemas de información

Tecnologías de la información

Usabilidad

Vigilancia tecnológica

Web

Historia de *EPI*

Fundada en 1992 por su actual director, la revista empezó a publicarse por la editorial *Learned Information*, Oxford, con el nombre *Information World en Español (IWE)*. Tenía el formato de newsletter o boletín de noticias y reportajes.

En 1997 fue comprada por la compañía *Swets & Zeitlinger* y se publicó desde Lisse, Holanda.

Un año más tarde, por razones legales, cambió el nombre a *El profesional de la información*. Al mismo tiempo se transformó en revista científica con artículos originales evaluados por pares por el sistema doble anonimato (autores y evaluadores no se conocen).

En 2004, S&Z vendió toda su cartera de revistas a la editorial inglesa *Taylor & Francis*.

Finalmente en 2006 los derechos de publicación de *EPI* fueron adquiridos por la editorial española *EPI SCP*.

En 2015 deja de publicarse en papel, manteniendo la edición electrónica.

Equipo editorial

La redacción de *EPI* está compuesta por un grupo de profesionales que cuentan con años de experiencia en el campo de la información, en toda la gama desde la archivística y las bibliotecas hasta la web social y los medios de comunicación, pasando por la gestión de la información en empresas y organizaciones.

Director: Tomàs Baiget, *EPI*, Barcelona

Subdirector: Javier Guallar, *Univ. de Barcelona, Univ. Ramon Llull*

Coordinadores: Carlos Tejada-Artigas, *Universidad Complutense de Madrid*
Luis Rodríguez-Yunta, *CCHS-CSIC*, Madrid

Redactora jefe: Isabel Olea, *EPI*, León

Redactores: Natalia Arroyo-Vázquez, *FGSR*, Salamanca
Ricardo Eito-Brun, *Grupo GMV*, Madrid
Javier Leiva-Aguilera, *Catorze.com*, Barcelona
Toon Lowette, *Grid*, Bruselas
Roser Lozano, *Universitat Rovira i Virgili*, Tarragona
Jorge Serrano-Cobos, *MASmedios*, Valencia

Revisión de lengua inglesa: Elaine Lilly, *Writer's First Aid*, Minnesota

Gestión de la web: María Moreno, *Florida Universitaria*, Valencia

Diseño: Moisés Mañas, *MASmedios*, Valencia

Maquetación: Isabel Olea, *EPI*, León

Suscripciones: Isabel Olea, *EPI*, León

Publicidad: Isabel Olea, *EPI*, León

Comité científico

Además del citado equipo editorial la evaluación de la calidad de los artículos la llevan a cabo:

Ernest Abadal (UB), Barcelona; Isidro F. Aguillo (CCHS-CSIC), Madrid; Rafael Aleixandre-Benavent (Instituto de Historia de la Medicina y de la Ciencia López Piñero), Valencia; Lluís Anglada (CBUC), Barcelona; Jesús Bustamante (Cedefop), Grecia; Atilio Bustos (Pontificia Univ. Católica de Valparaíso), Chile; Lluís Codina (UPF), Barcelona; Emilio Delgado-López-Cózar (UGR), Granada; Javier Díaz-Noci (UPF), Barcelona; Assumpció Estivill (UB), Barcelona; Antonia Ferrer-Sapena (UPV), Valencia; António Fidalgo (UBI), Portugal; Fco.-Javier García-Marco (Unizar), Zaragoza; Roma; Johannes Keizer (FAO), Roma; Jesús Lau (Universidad Veracruzana), México; Mari-Carmen Marcos (UPF), Barcelona; Pere Masip (URL), Barcelona; Charles McCathieNevile (Yandex), Moscow; José-Antonio Moreira (UC3M), Madrid; Enrique Orduña-Malea (UPV), Valencia; Juan-Antonio Pastor (UM), Murcia; Fernanda Peset (UPV), Valencia; Joan Roca (MNSU), Minnesota; Ramón Salaverría (UNAV), Pamplona; Tomás Saorín (UM), Murcia; Robert Seal (LUC), Chicago; Ernesto Spinak (consultor), Uruguay; Emir-José Suaidén (IBICT), Brasilia; Daniel Torres-Salinas (UNAV), Pamplona; Jesús Tramullas (Unizar), Zaragoza; y otros evaluadores externos ocasionales.

Acceso a los textos completos

Recyt (2000-)

<http://recyt.fecyt.es/index.php/EPI>

Ebscohost Electronic Journals Service (2000 - embargo 1 año)

<http://ejournals.ebsco.com>

LISTA with full text (2000 - embargo 1 año)

<http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=584>

Digitalia Publishing (2011 -)

<http://www.digitaliapublishing.com>

El profesional de la información (1992 - embargo 2 años)

<http://elprofesionaldelainformacion.com/contenidos.html>

ALPSP Learned Journals Collection (2011 -)

<http://ajc.swets.com>

e-Libro (2011 -)

<http://www.e-libro.net>

EPI en las redes sociales



EPI en Facebook

<http://facebook.com/elprofesionaldelainformacion>



EPI en Pinterest

<http://www.pinterest.com/revistaepi>



EPI en Twitter

http://twitter.com/revista_EPI



EPI en LinkedIn

<http://www.linkedin.com/company/el-profesional-de-la-informacion-epi>

Información para los autores

Los profesionales que lo deseen pueden remitir a la redacción de la revista *El profesional de la información* sus colaboraciones, en forma de:

- Notas descriptivas, experiencias, servicios, productos... para la sección "Análisis",
- Trabajos de investigación, más amplios, sobre temas de fondo para la sección "Artículos".

Deben seguirse estrictamente las normas que figuran en esta página:

<http://www.elprofesionaldelainformacion.com/autores.html>

Todas las contribuciones deben tramitarse a través del programa OJS de la web del *Repositorio Español de Ciencia y Tecnología (Recyt)*, sección de EPI:

<http://recyt.fecyt.es/index.php/EPI/index>

Los autores que envíen un trabajo por primera vez deberán registrarse en:

<http://recyt.fecyt.es/index.php/EPI/user/registerJournal>

Todos los trabajos son aprobados según el sistema "peer review": al menos dos expertos en el tema, del consejo científico 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 EPI 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.

Precios 2015

Institucional (online)

107,44 € + 21% IVA = 130 €

Individual (online)

65,29 € + 21% IVA = 79 €

En este caso la suscripción se hace exclusivamente a particulares

Número suelto papel (anteriores a 2015)

28,85 € + 4% IVA = 30 €

(gastos de envío: Europa: 9 €; Américas: 15 €)

Fuera de Europa no se cobra IVA

Puede suscribirse online en:

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



Autores

Para hacer frente a los costes de edición, a los autores se les solicita abonar 200 euros por artículo, efectivos una vez el artículo se haya aprobado en el proceso de *peer review* y esté listo para su publicación.

Open choice

Los autores pueden liberar su artículo en *open access* en las webs de *EPI* mediante pago de 200 € adicionales a los anteriores, es decir, en total 400 €

Publicidad

Con suscriptores repartidos a partes iguales entre empresas, universidades y administraciones públicas, probablemente *EPI* es el medio más eficaz para anunciar sus productos y servicios de información en España y Latinoamérica.

Póngase en contacto por correo-e:
publici@elprofesionaldelainformacion.com
o telefono: +34 609 352 954

Boletín de suscripción

<input type="checkbox"/> Institucional	<input type="checkbox"/> Personal
Nombre <input type="text"/>	Institución <input type="text"/>
<small>(Los suscriptores individuales no han de escribir ningún nombre de institución, sólo indicar la dirección particular)</small>	
Departamento <input type="text"/>	NIF institucional <input type="text"/>
Dirección <input type="text"/>	Código postal <input type="text"/> Ciudad <input type="text"/> País <input type="text"/>
Teléfono <input type="text"/>	Fax <input type="text"/> Correo-e <input type="text"/>
Ejemplares solicitados	<input type="text"/>
Método de pago	
Tarjeta de crédito:	<input type="checkbox"/> VISA <input type="checkbox"/> Master Card <input type="checkbox"/> American Express
Titular de la tarjeta <input type="text"/>	Código de seguridad CVC2 <input type="text"/>
Número de tarjeta <input type="text"/>	Caducidad (mm/aaaa) <input type="text"/>
<input type="checkbox"/> Cheque nominativo en euros a nombre de El profesional de la información	
<input type="checkbox"/> Transferencia bancaria a la cuenta de La Caixa 2100 0818 93 0200745544 <small>Enviar, fotocopiado o escaneado, el resguardo de la transferencia.</small>	
<small>Las transferencias desde fuera de España deben hacerse a IBAN ES95 2100 0818 9302 0074 5544 - BIC/Código Swift CAIXESBBXXX</small>	
<input type="checkbox"/> Giro postal al apartado de correos 32.280 de Barcelona <small>Enviar, fotocopiado o escaneado, el resguardo del giro.</small>	
<input type="checkbox"/> Domiciliación en cuenta bancaria	
Entidad <input type="text"/>	Oficina <input type="text"/> DC <input type="text"/> Núm <input type="text"/>
Titular de la cuenta <input type="text"/>	
<input type="checkbox"/> PayPal a la cuenta EPISCP@gmail.com	

Enviar el boletín relleno por correo electrónico, o postal al APARTADO 32.280 - 08080 BARCELONA - ESPAÑA.

Consultas: suscripciones@elprofesionaldelainformacion.com o +34 609 352 954



Te entendemos”

Sistemas de gestión de bibliotecas Open Source
Interfaces interactivas y OPACs
Repositorios OAI
Gestión documental y de archivos
Digitalización
Outsourcing de servicios documentales
Desarrollo de sitios web / multimedia / e-learning
Comunicación y e-marketing de servicios de información



Oficinas Centrales:
C/Garcilaso 15-B
46003 Valencia

Tel.: 96 369 41 23
Fax: 96 369 34 39
info@masmedios.com

ISO 9001
BUREAU VERITAS
Certification



ISO 27001
BUREAU VERITAS
Certification



Deseo suscribirme a la revista EPI a partir del mes de enero del año 20 Las suscripciones van por años naturales, de enero a diciembre

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

Titular de la tarjeta Código de seguridad CVC2

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 **ES95 2100 0818 93 0200745544** Enviar, fotocopiado o escaneado, el resguardo de la transferencia.

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

PayPal a la cuenta EPISCP@gmail.com



Enviar el boletín cumplimentado por correo electrónico, o postal al APARTADO 32.280 - 08080 BARCELONA - ESPAÑA.

Consultas: suscripciones@elprofesionaldelainformacion.com o +34 609 352 954

Open choice. Los autores pueden liberar su artículo en open access en las webs de EPI mediante el pago de 400 €

Precios 2015

Importante: A partir de 2015 EPI sólo se publicará online

Suscripción institucional EPI: 107,44 € + 21% IVA = 130 €

Suscripción institucional conjunta EPI + Anuario ThinkEPI: 157,03 € + 21% IVA = 190 €

Suscripción personal: 65,29 € + 21% IVA = 79 €

Suscripción personal conjunta EPI + Anuario ThinkEPI: 90,91 € + 21% IVA = 110 €

Número suelto: 35,57 € + IVA = 37 €

Gastos de envío fuera de España: Europa: 9 € Américas: 15 €

Fuera de Europa no se cobra el IVA

NORMAS PARA LOS AUTORES

El profesional de la información tiene dos secciones principales:

ARTÍCULOS: Trabajos de investigación y temas analizados en profundidad.

ANÁLISIS: Experiencias, estudios de casos, análisis de productos, reseñas, etc.

Las contribuciones han de ser originales e inéditas, no pueden haberse publicado previamente en soporte papel o electrónico. El tamaño ideal es de 3.500 palabras, aunque en algunos casos la Redacción puede autorizar una mayor extensión.

El texto ha de enviarse en Word, rtf u odt. Las tablas deberán ir pegadas en el mismo documento. Todos los materiales gráficos (diagramas, fotografías, capturas de pantalla, etc.) deben pegarse en el Word y además enviarse en ficheros independientes (en formatos xls, jpg, pdf, etc.). Las imágenes jpg deben tener una resolución de al menos 300 pp (unos 200 KB cada una).

El texto debe presentarse completamente plano, sin autoformatos ni automatismos de Word (subsecciones, viñetas, citas enlazadas, pies de página, sangrías, tabulaciones, colores, etc.), pero debe seguir el estilo de EPI en cuanto a **negritas** (nombres de los autores citados), *cursivas* (instituciones, títulos de revista, marcas) y mayúsculas. Los urls deben estar sin hipervínculo.

Las citas bibliográficas en el texto se realizarán de la forma: (**Apellido**, año) o (**ApellidoAutor1**; **ApellidoAutor2**, año).

La redacción debe ser concisa y precisa, evitando la retórica.

Los trabajos deben incluir: a) título, b) resumen de 100-150 palabras, c) 5-10 palabras clave, d) title, e) abstract de 100-150 palabras, f) 5-10 keywords.

Aparte se incluirá el nombre de los autores, su lugar de trabajo y dirección (postal y electrónica), su foto tipo carnet superior a 100 KB en jpg, y un currículum de unas 70 palabras.

EVALUACIÓN

Los trabajos son revisados según el sistema tradicional "peer review" en doble ciego por al menos dos expertos en el tema, del Consejo Asesor de la revista y/o externos. La revista se compromete a informar del resultado a los autores.

REFERENCIAS BIBLIOGRÁFICAS

Ordenadas alfabéticamente por autor, se limitarán a las obras citadas en el artículo. No se acepta bibliografía de relleno.

Artículos de una publicación periódica:

Apellido, Nombre; Apellido2, Nombre2. "Título del artículo".

Título de la publicación periódica, año, v., n., mes, pp. xx-yy.

Dirección url iniciada en nueva línea sin barra y sin punto finales

<http://dx.doi.org/10.xxxx/doi>

Ponencia presentada en un congreso:

Apellido, Nombre; Apellido2, Nombre2. "Título de ponencia".

En: *Nombre del congreso*, año, pp. xx-yy.

Dirección url iniciada en nueva línea sin barra y sin punto finales

<http://dx.doi.org/10.xxxx/doi>

Monografías e informes:

Apellido, Nombre; Apellido2, Nombre2. *Título del trabajo.*

Lugar de publicación: editor, fecha, ISBN: espacios, sin guiones

Dirección url iniciada en nueva línea sin barra y sin punto finales

<http://dx.doi.org/10.xxxx/doi>

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. xx-yy. ISBN: espacios, sin guiones

Dirección url iniciada en nueva línea sin barra y sin punto finales

Recurso en línea:

Apellido, Nombre; Apellido2, Nombre2. *Título del recurso (sin fecha de la consulta).*

Dirección url iniciada en nueva línea sin barra y sin punto finales

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>