



Observatorio

Visualización de información

Carlos Olmeda-Gómez

Artículos

How we draw texts. A review of approaches to text visualization and exploration

J. Nualart-Vilaplana, M. Pérez-Montoro y M. Whitelaw

Análisis visual en e-learning

D. A. Gómez-Aguilar, F. J. García-Peñalvo y R. Therón

Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales

M. Del-Fresno-García

Tendencias en mapas de la ciencia: co-uso de la información científica como reflejo de los intereses de los investigadores

D. Torres-Salinas, E. Jiménez-Contreras y N. Robinson

Estudio evolutivo de la investigación española con células madre. Visualización y líneas de investigación

G. Cantos-Mateos, M. A. Zulueta, B. Vargas-Quesada y Z. Chinchilla-Rodríguez

Graphical interface of the *SCLImago Journal and Country Rank*: an interactive approach to accessing bibliometric information

Y. Hassan-Montero, V. P. Guerrero-Bote y F. De-Moya

Adwords, imágenes y ceguera a los banners. Un estudio con eye tracking

L. Ortiz-Chaves, E. Martínez-de-Pisón, G. Cancela-López-Carrión, J. Gonçalves-de-Vasconcelos, C. Rovira y M. C. Marcos

Alfabetización mediática y co-innovación en la microempresa: primeras evidencias para España

J. F. Martínez-Cerdá y J. Torrent-Sellens

Credibilidad de los medios de comunicación: análisis de la prensa diaria desde el comportamiento del consumidor

C. Calvo-Porrá, V. A. Martínez-Fernández y Ó. Juanatey-Boga

Análisis

Visualización de la información en la democratización de los datos

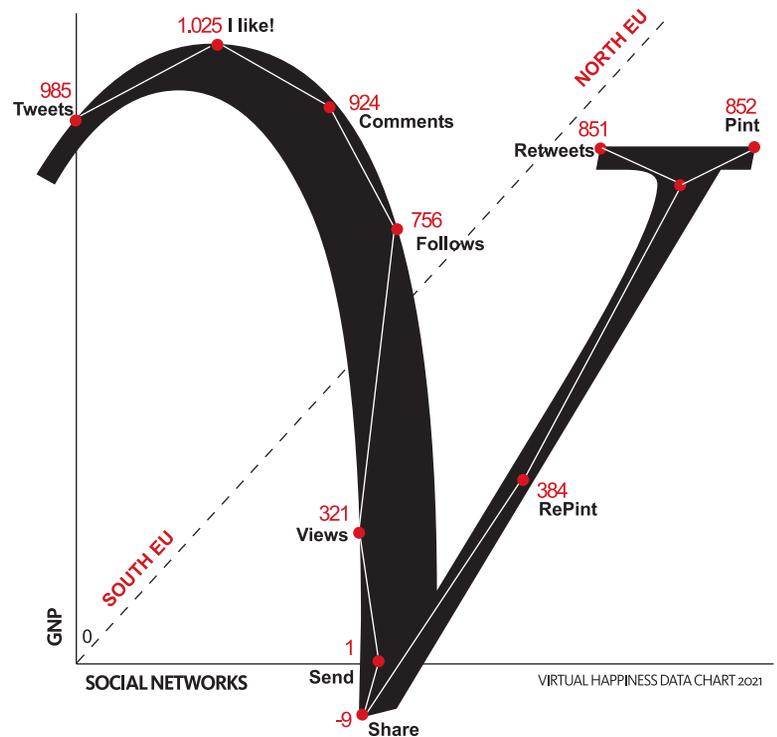
C. Sánchez-Bonvehí y M. Ribera

Visualización y recuperación en la web semántica

J. Morato, S. Sánchez-Cuadrado, A. Ruiz-Robles y J. A. Moreiro-González

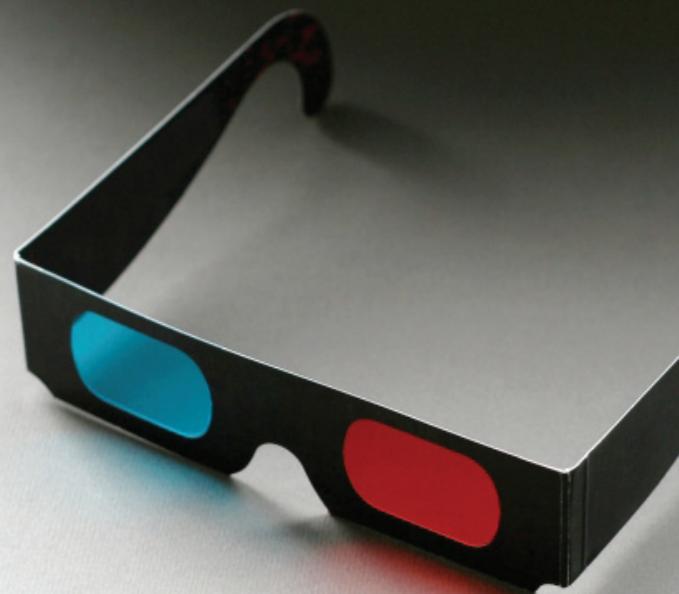
Búsqueda de expedientes de forma visual

I. Medrano-Corrales, D. Ruiz-Macías y M. J. Escalona-Cuesta



Visualización DE LA INFORMACIÓN

11:24 AM - 3D MINDMAP



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 NIF: J63664544

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

Redacción

El profesional de la información
Apartado 32.280
08080 Barcelona
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.
C/ Perú s/n, nave 1
08754 El Papiol (Barcelona)
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 / Univ. Oberta de Catalunya
<http://sites.google.com/site/sitiodeljavierguallar>

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 Universidad de León
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

CONSEJO ASESOR

Ernest Abadal *Universitat de Barcelona*, Barcelona.
Isidro F. Aguillo *Centro de CC Humanas y Sociales, CSIC*, Madrid.
Adela d'Alòs-Moner *Consultora*, Barcelona.
Carlos B. Amat *Inst. Agroquím. y Tecn. Alimentos, CSIC*, Valencia.
Jesús Bustamante *Biblioteca, Cedefop*, Salónica, Grecia.
Carlota Bustelo-Ruesta *Consultora*, Madrid.
Atilio Bustos *Pontificia Univ. Católica de Valparaíso*, Chile.
Lluís Codina *Universitat Pompeu Fabra*, Barcelona.
Emilio Delgado-López-Cózar *Universidad de Granada*, Granada.
Javier Díaz-Noci *Universitat Pompeu Fabra*, Barcelona.
Assumpció Estivill *Universitat de Barcelona*, Barcelona.
Antonia Ferrer-Sapena *Universidad Politécnica de Valencia*, Valencia.
António Fidalgo *Universidade da Beira Interior*, Portugal.
Francisco-Javier García-Marco *Universidad de Zaragoza*, Zaragoza.
Johannes Keizer *Food and Agriculture Org. (FAO)*, Roma, Italia.
Jesús Lau *Universidad Veracruzana*, Veracruz, México.
Mari-Carmen Marcos *Universitat Pompeu Fabra*, Barcelona.
Pere Masip *Blanquerna, Univ. Ramon Llull*, Barcelona.
Charles McCathieNevile *Yandex*, Moscow, Rusia.
José-Antonio Moreiro-González *Universidad Carlos III de Madrid*.
Enrique Orduña-Malea *Universidad Politécnica de Valencia*, Valencia.
Juan-Antonio Pastor *Universidad de Murcia*, Murcia.
Fernanda Peset *Universidad Politécnica de Valencia*, Valencia.
Joan Roca *Minnesota State University*, Mankato, USA.
Ramón Salaverría *Universidad de Navarra*, Pamplona.
Tomás Saorín *Universidad de Murcia*, Murcia.
Robert Seal *Loyola Univ. Chicago*, Evanston, Illinois, USA.
Ernesto Spinak *Consultor*, Montevideo, Uruguay.
Emir-José Suaidén *IBICT*, Brasilia, Brasil.
Daniel Torres-Salinas *Universidad de Navarra*, Pamplona.
Jesús Tramullas *Universidad de Zaragoza*, Zaragoza..

BASES DE DATOS

Academic search premier (Ebsco)

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

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 science citation index, Social SCI, WoS (Thomson Reuters)

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

Impact Factor 2012 = 0,439

Inspec, Information services in physics, electronics and computing (IET, The Institution of Engineering and Technology)

<http://www.theiet.org/publishing/inspec/>

ISOC, Índice español de ciencias sociales y humanidades (Iedcyt)

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

ISTA, Information science and technology abstracts (Ebsco)

<http://www.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 2012 = 0,279

<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://ececubo.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: Visualización de información

OBSERVATORIO

213 Visualización de información

Carlos Olmeda-Gómez

ARTÍCULOS

221 How we draw texts. A review of approaches to text visualization and exploration

Jaume Nualart-Vilaplana, Mario Pérez-Montoro y Mitchel Whitelaw

236 Analítica visual en e-learning

Diego-Alonso Gómez-Aguilar, Francisco-José García-Peñalvo y Roberto Therón

246 Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales

Miguel Del-Fresno-García

253 Tendencias en mapas de la ciencia: co-uso de la información científica como reflejo de los intereses de los investigadores

Daniel Torres-Salinas, Evaristo Jiménez-Contreras y Nicolás Robinson-García

259 Estudio evolutivo de la investigación española con células madre. Visualización e identificación de las principales líneas de investigación

Gisela Cantos-Mateo, María-Ángeles Zulueta, Benjamín Vargas-Quesada y Zaida Chinchilla-Rodríguez

272 Graphical interface of the SCImago Journal and Country Rank: an interactive approach to accessing bibliometric information

Yusef Hassan-Montero, Vicente P. Guerrero-Bote y Félix De-Moya-Anegón

279 AdWords, imágenes y ceguera a los banners. Un estudio con eye tracking

Laura Ortiz-Chaves, Eduardo Martínez-de-Pisón, Germán Cancela-López-Carrión, Juliano Gonçalves-de-Vasconcellos, Cristófol Rovira y Mari-Carmen Marcos

288 Alfabetización mediática y co-innovación en la microempresa: primeras evidencias para España

Juan-Francisco Martínez-Cerdá y Joan Torrent-Sellens

300 Credibilidad de los medios de comunicación: análisis de la prensa diaria desde el comportamiento del consumidor

Cristina Calvo-Porrá, Valentín-Alejandro Martínez-Fernández y Óscar Juanatey-Boga

ANÁLISIS

311 Visualización de la información en la democratización de los datos: propuestas desde el periodismo y la narratividad

Clàudia Sánchez-Bonvehí y Mireia Ribera

319 Visualización y recuperación de información en la web semántica

Jorge Morato, Sonia Sánchez-Cuadrado, Alejandro Ruiz-Robles y José-Antonio Moreira-González

330 Búsqueda de expedientes de forma visual

Isabel Medrano-Corrales, Diego Ruiz-Macías y María-José Escalona-Cuaresma

339 AGENDA

344 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 y 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 2.0.

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 mappings a EDM 5.2.5 (Europeana Data Model)**
- **Bibliotecas digitales que permiten la creación, recuperación y recolección de metadatos (MARCXML, DCMI y RDF y RDFs)**
- **Archivos Web que facilitan la creación, recuperación y recolección de metadatos (EAD 2.0 y EAC 2010)**
- **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 (Analyzed Layout and Text Object)**
- **Generación e integración de registros SKOS mediante MARC 21(Up.18)/RDA**
- **Creación de eBooks o libros digitales en formatos: ePub y Mobipocket.**
- **Adaptación de DIGIBIB a Linked Open Data**



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

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



Nº ES042816-1



www.digibis.com



Carlos Olmeda-Gómez



Carlos Olmeda-Gómez es licenciado en historia moderna y contemporánea por la *Universidad Autónoma de Madrid*, master en documentación en ciencias de la información y doctor en filología hispánica por la *Universidad Complutense de Madrid*. Profesor titular del *Departamento de Biblioteconomía y Documentación* de la *Universidad Carlos III de Madrid*. Realiza estudios de contenido bibliométrico, colaboración científica y análisis de dominios. Participa en diversas iniciativas y proyectos en el *SCImago Research Group*.
<http://orcid.org/0000-0001-5955-6423>

Universidad Carlos III de Madrid
Departamento de Biblioteconomía y Documentación
Madrid, 128. 28903 Getafe (Madrid), España
carlos.olmeda@uc3m.es

Resumen

Visión general y sintética que introduce algunos conceptos básicos y ejemplos de visualizaciones de información. Se muestran algunas visualizaciones con datos geoespaciales, relacionales y textuales de una disciplina exigente. Se exponen conceptos y los ejemplos ilustran algunas prácticas y criterios.

Palabras clave

Visualización de información, Visualización de textos, Visualización de datos, Tipos de visualización, Gráficos, Visualización.

Title: Information visualization

Abstract

This overview introduces and synthesizes some basic concepts of the broad spectrum of information visualization, providing examples with geospatial, relational and textual data from a demanding discipline. The concepts are discussed and the examples illustrate some practices and criteria.

Keywords

Information visualization, Text visualization, Data visualization, Visualization types, Graphics, Visualization.

Olmeda-Gómez, Carlos (2014). "Visualización de información". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 213-219.

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

1. Introducción

La visualización de información de acuerdo con **Card, Mackinlay** y **Shneiderman** (1999), se define como "el uso de representaciones visuales interactivas de datos abstractos, no físicos, para ampliar el conocimiento". En otras palabras, se centra en la extracción de estructuras de grandes conjuntos de datos, que puedan ser representadas de forma visual a los usuarios. En consecuencia, en las visualizaciones de información intervienen distintos procesos. Hay una transformación de datos brutos en abstracciones analíticas que, a continuación, se transforman en un modelo espacial-visual abstracto, para que finalmente, mediante procesos de diseño visual, el modelo visual se presente al usuario de forma gráfica y visible.

Cada proyecto de visualización difiere en sus fases o etapas, dependiendo de:

- si los datos están estructurados o no;
- las técnicas a emplear para la limpieza, selección y almacenamiento de los datos;
- los algoritmos empleados para simplificar las vistas;
- la definición de los parámetros del escenario donde se presentan; y
- el propósito de uso.

Se supone que la obtención de visualizaciones de información puede ayudar al reconocimiento de patrones, de estructuras, tendencias y relaciones que ayuden a la comprensión de los datos y del fenómeno que representen, que de otro modo serían incomprensibles por su complejidad.

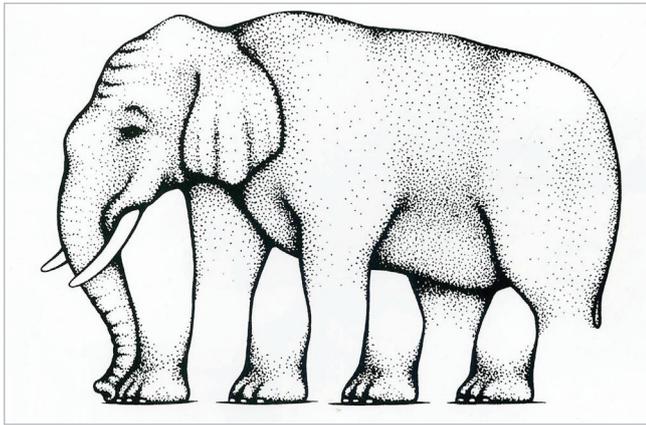


Figura 1. <http://en.wikipedia.org>

Por comprensión se puede entender la construcción de un concepto, de un modelo mental a partir de los datos o del fenómeno que se analiza. Una visualización de información captura las características esenciales de los datos, no presenta los items de forma individual y por eso el modelo es necesariamente abstracto, pero de forma simultánea, las herramientas de visualización deben facilitar la percepción de subconjuntos de datos, items o de unidades.

Las herramientas de visualización deben facilitar la percepción de subconjuntos de datos, items o de unidades

El observador de la visualización puede querer emplear o usar la visualización para explorar, para confirmar una hipótesis o presentar los resultados a una audiencia; para detectar anomalías, es decir datos que no se comportan de forma consistente con las expectativas; descubrir grupos o *clusters*, datos que comparten características similares de algún tipo, que de ese modo pueden indicar la presencia de un fenómeno en particular; o tendencias, es decir, datos que cambian de tal modo que puede ser característico y, por consiguiente, ser usado en modelos predictivos. El observador no es sino un *lector* que debe interpretar la visualización a partir de la forma física externa de la estructura y de los atributos de los datos que se usan para definirlos de forma visual como objetos. Fundamentalmente hay cuatro clases de atributos:

- valores visuales (posición, longitud, ángulo, dirección, forma, área, volumen, saturación y color);
- sistema de coordenadas (cartesiano, polar y geográfico);
- escala (numérica, categórica, temporal); y
- contexto (la información que proporciona el mayor entendimiento de quién, qué, cuándo, dónde y el porqué de los datos), que puede aclarar los datos a los lectores y guiarlos en la dirección correcta.

Otro componente característico de los procesos de visualización de información se refiere al modo en que se proporcionan al usuario controles para la manipulación interactiva o la selección de variables, atributos o parámetros. Las visualizaciones de información son procesos dinámicos,

donde se permite al usuario en su grado máximo, el control virtual de todas las fases de los procedimientos, desde la selección y control de los datos que se desean ver, hasta la posible selección del color o el modo de afinar las vistas. No existe una fórmula que garantice la efectividad de una visualización. Los usuarios son diferentes en su formación y grado de familiaridad, en sus habilidades de percepción y en sus preferencias, y se diferencian por las opiniones que les susciten las distintas visualizaciones. Las tareas de los usuarios también afectan a la utilidad de las visualizaciones. Un cambio en los datos puede tener consecuencias en el resultado. Por eso es obligado permitir que el usuario pueda adaptar, modificar o refinar de forma interactiva la visualización, hasta que crea que ha logrado sus objetivos, por ejemplo porque ha extraído una descripción completa, rigurosa, exacta de lo que contienen los datos o porque ha obtenido una representación clara de los patrones que los datos transmiten.

Un aspecto crítico está relacionado con las capacidades cognitivas del usuario y con las habilidades y las limitaciones propias de nuestro sistema visual. Si el objetivo de la visualización es ayudar a comprender fenómenos complejos de información mediante su transformación en objetos visuales abstractos, es esencial tener en consideración las habilidades de percepción (**Ware**, 2012) que nuestro sistema nos facilita o con el que, en ocasiones, nos traicionamos. Todo el mundo conoce la expresión “una imagen vale más que mil palabras” o el modo en que un buen dibujo puede ser estimulante para la comprensión de un fenómeno, pero si se presentan resultados de visualización de información, nadie desea ambigüedades como cuando observamos el elefante de múltiples patas de **Shepard** (1990) (figura 1).

Los usuarios interactúan con las visualizaciones basándose en lo que se ve y en lo que se interpreta

Los usuarios interactúan con las visualizaciones basándose en lo que se ve y en lo que se interpreta. Comprender cómo vemos, debería ayudarnos a producir mejores dispositivos de presentación de visualizaciones de información o al menos, no construirlas con deficiencias. En un mundo moldeado de forma creciente por algoritmos, la visualización de información juega un papel relevante para dotarle de sentido. Los autores relevantes en este aspecto de la percepción proceden de las ciencias cognitivas, la psicología, la psicofísica, la educación y el diseño gráfico. La comprensión de los procesos que están implicados en la percepción visual y de los aspectos cognitivos conexos es obligada para generar visualizaciones efectivas y eficaces (**Costa**, 1998).

La representación gráfica de información cuantitativa tiene también raíces profundas que se basan en las historias que contienen la cartografía temática, la estadística gráfica y la visualización de datos. Se pueden reconocer antecedentes en la generación de diagramas geométricos, de los mapas y las cartografías que ayudaron a la navegación y a la exploración en los siglos de los descubrimientos o de la *Cosmo-*

graphia de Ptolomeo. También en las técnicas e instrumentos precisos para la observación y la medición de cantidades, así como de las notaciones y fórmulas necesarias para su expresión. Otros antecedentes se sitúan en la invención de la geometría analítica, la estadística demográfica y la aritmética política, todos ellos orientados para tratar de comunicar con mayor facilidad o facilitar la inspección visual (Tufte, 1990).

2. Visualización de datos geoespaciales

Los datos geoespaciales son diferentes de cualquier otro tipo de dato, ya que describen objetos o fenómenos con una localización específica en el mundo real. Los objetos de información susceptibles de ser ubicados en un espacio, pueden proceder de múltiples aplicaciones, como las transacciones financieras, solicitudes de créditos, comunicaciones en *Twitter*, registros medioambientales, censos demográficos o votaciones políticas. Son datos que pueden ser potencialmente geolocalizados por las denominaciones de los lugares, direcciones o códigos postales y ser vinculados a una ubicación geoespacial en la pantalla. Con mucha frecuencia los datos geolocalizables se ubican en mapas temáticos. Los mapas constituyen la forma más familiar de representar objetos de información geolocalizable, que una vez ubicados en el espacio pueden revelar patrones y relaciones difíciles de descubrir o de ser apreciados de otro modo.

Los mapas pueden subdividirse basándose en las propiedades de los datos: cualitativos frente a cuantitativos; dis-

cretos o continuos, y por las propiedades de las variables gráficas: puntos para localizar, líneas para interconectar, tamaños, volúmenes. Ello da como resultado la generación de mapas simbólicos, de mapas de puntos, mapas de área y cartogramas, mapas coropléticos (con colores según densidad) e isopléticos (con líneas de igual valor), mapas de línea o de superficie (figura 2).

En la visualización de objetos de información geolocalizable las proyecciones elegidas juegan un papel crítico. Las proyecciones en los mapas tienen que ver con la transposición de la localización de los objetos de información en un globo, que es esférico, a una superficie plana como es la pantalla que es interactiva. Las proyecciones de los mapas se pueden clasificar dependiendo del tipo de superficie sobre el que la esfera se proyecta. Las superficies de proyección en la pantalla son cilíndricas, planas y cónicas y en otros contextos, cartesianas, polares o hiperbólicas (figura 3).

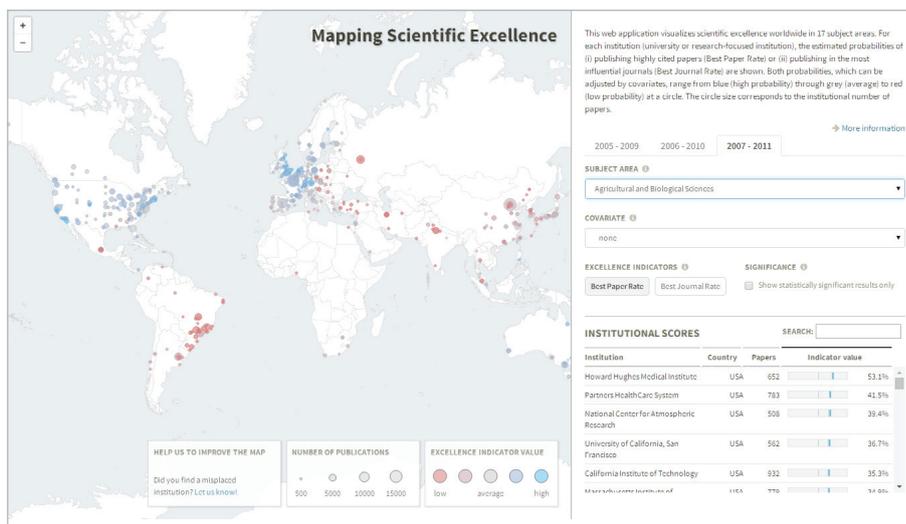
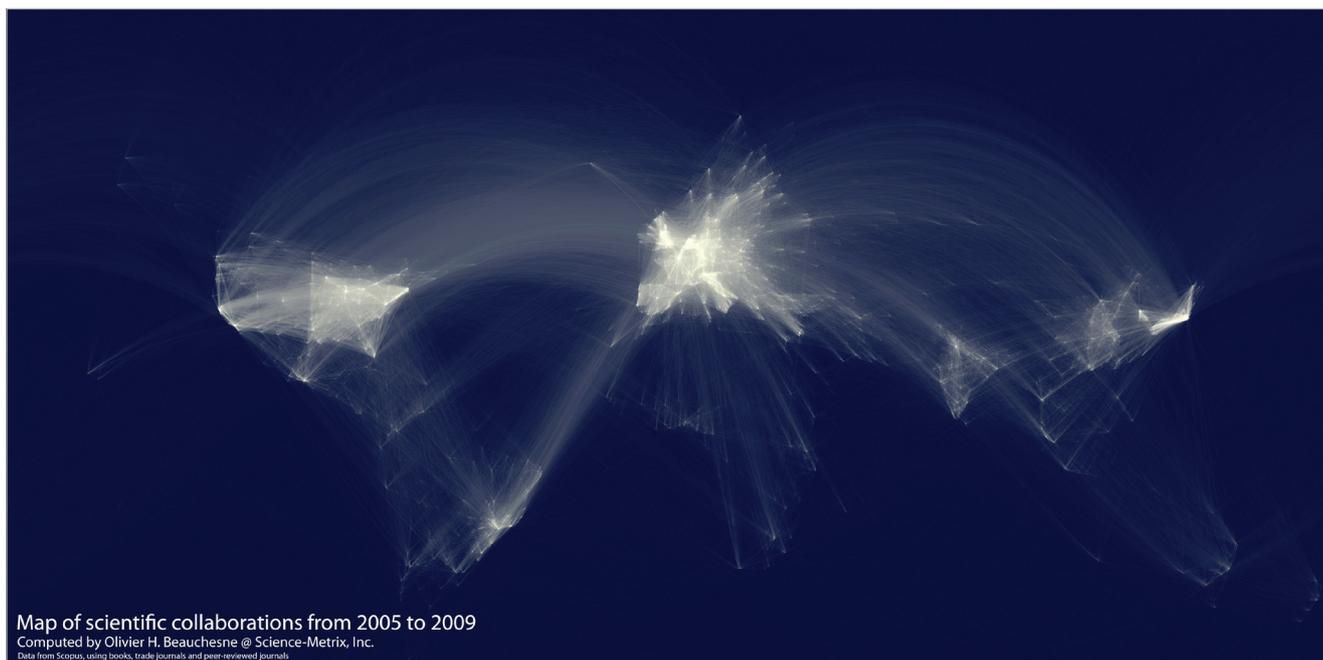


Figura 2. <http://www.excellencemapping.net>



Map of scientific collaborations from 2005 to 2009
 Computed by Olivier H. Beauchesne @ Science-Metrix, Inc.
 Data from Scopus, using books, trade journals, and peer-reviewed journals

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

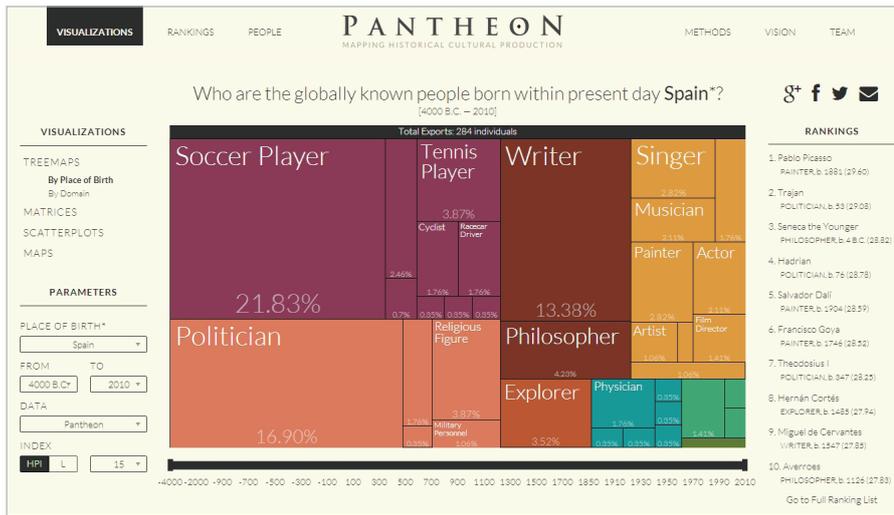


Figura 4. <http://pantheon.media.mit.edu>

Los mapas se usan de múltiples formas. Por ejemplo para proporcionar información específica sobre posiciones concretas, para generar información general sobre patrones espaciales o para comparar patrones en múltiples mapas. Por ello es obligado que las variables visuales reflejen las propiedades de los datos espaciales. Las variables visuales para datos espaciales son el tamaño de los símbolos utilizados, la anchura de las líneas o los tamaños de las áreas; la forma de los símbolos individuales o los patrones de las líneas y las áreas. Otras variables son el brillo, el color, la textura, la orientación, el tipo de perspectiva y la disposición de los patrones en el seno de los símbolos individuales (por ejemplo, distribuciones regulares o aleatorias de símbolos).

En sistemas de geovisualización exploratorios la interacción con los mapas es crucial. En contraste con la cartografía tradicional, la clasificación y el mapeo de datos pueden adaptarse para que sea el usuario quien interactúe con los mapas, al mismo tiempo que puede filtrar, ojear y formular preguntas desde la interfaz o manipular la presentación en la pantalla, por ejemplo mediante zoom. De este modo se

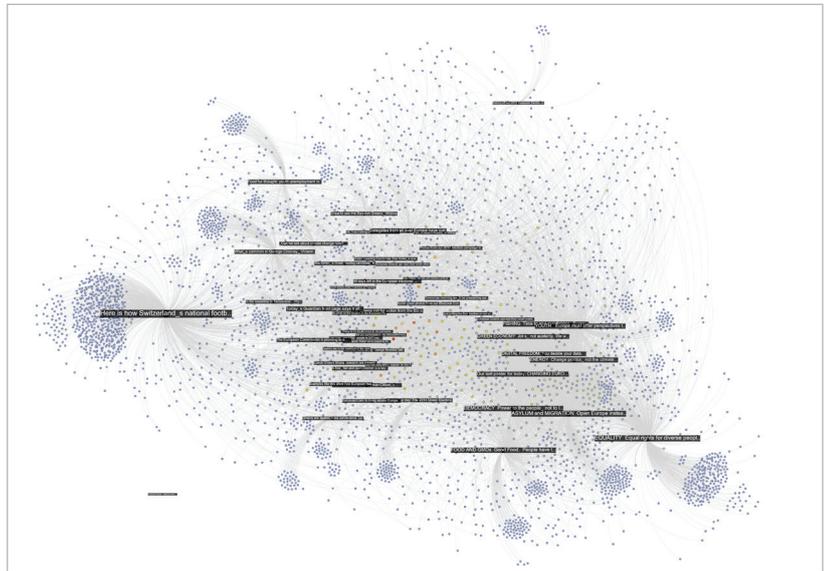


Figura 5. <http://www.theguardian.com/uk>

pueden lograr mapas interactivos de objetos de información. Los diseñadores de mapas interactivos parten de un mapa base que constituye el soporte, sobre el que se superponen capas de datos, que llevan incorporados los valores de coordenadas y de otro tipo.

3. Visualización de estructuras jerárquicas y de redes

La visualización de grafos, simples, elegantes, directos, pueden informarnos, sin conocer nada sobre los datos subyacentes y, sin ni siquiera intentarlo, se descubren relaciones y patrones. Los nodos y los enlaces pueden ser conexiones

entre miembros de *Facebook*, *Twitter*, redes de distribución de datos, llamadas a centralitas telefónicas, mensajes de correo electrónico o interconexiones en gramáticas. Todas son redes, entendidas como entidades interconectadas, y todas ellas se pueden analizar mediante generación de visualizaciones de grafos, pudiéndose disponer como interfaz visual de acceso a bibliotecas digitales, ya que la diferencia entre visualización e interfaz, se está borrando.

La visualización de redes se centra en las interrelaciones de la información relacional. Las interrelaciones pueden adoptar distintas formas:

- jerárquicas (padre/hijo; partes/sub-partes);
- conectividad (interconexiones entre lugares; redes de ordenadores);
- derivaciones, a partir de una secuencia de pasos o de fases (redes de citación);
- clasificaciones compartidas;
- semejanza entre valores o en atributos (por ejemplo, espaciales, temporales).

Las relaciones pueden ser simples o complejas, directas o bidireccionales, ponderadas o sin ponderar, ciertas o con grados de incertidumbre. Las relaciones pueden proporcionar más información que la que contienen los registros de las bases de datos tomados de uno en uno. Hay información en las conexiones. A simple vista se identifican nodos que atraen un número superior de enlaces, los nodos se pueden arracimar y constituir distintos subgrupos o se pueden detectar nodos aislados que carecen de conexiones.

Los árboles o jerarquías son unas de las estructuras más comunes que contienen información relacional. La visualización de jerarquías es una de las ramas más maduras y con mayor actividad de la visualización de información. Los conjuntos de datos de este tipo, proceden de estruc-

turas de directorios, organizaciones jerárquicas, procesos ramificados, genealogías, clasificaciones. Los mapas arborescentes (*tree maps*), desarrollados originalmente en la *University of Maryland* (**Shneiderman; Plaisant, 2006**), son una de las formas de visualización más populares. En su forma básica, un rectángulo se subdivide de forma recurrente, en rectángulos de menor tamaño, alternando el tamaño de las líneas horizontales y verticales y por consiguiente de las áreas, basándose en las poblaciones de las sub ramas en un determinado nivel. Obliga a adoptar un buen sistema de jerarquía de codificación visual (figura 4).

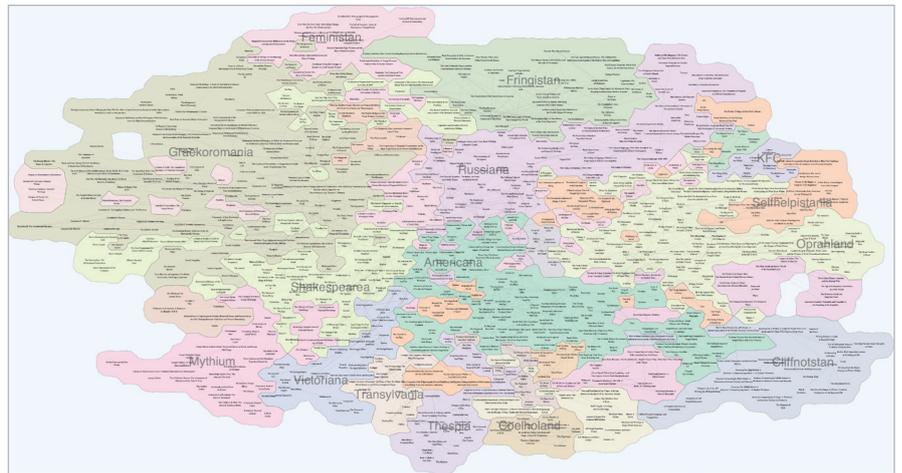


Figura 6. <http://www.amazon.com>



Figura 7. <http://www.wordle.net>

Técnicamente los árboles son gráficos conectados, no ponderados y acíclicos. No son sino uno de los tipos de una familia de representación de relaciones denominada grafos o redes. El estudio de las redes tiene una gran tradición en teoría de grafos y matemáticas discretas, en la investigación en sociología de la comunicación, pero también en bibliometría, informetría, webmetría y cibermetría. Las visualizaciones de redes están por todas partes y, de algún modo, deben estar, porque si no, no habría modo alguno de dotar de sentido a grandes conjuntos de datos, como los que proceden, por ejemplo de medios sociales (blogs, micro blogs, wikis, marcadores sociales, multimedia compartidos) (figura 5).

Los grafos son sencillos de realizar gracias a la existencia de múltiples programas automáticos que toman la informa-

ción que describe un nodo y sus conexiones y los representan mostrando la topología resultante. La estética es importante, no sólo por la apariencia de la red, sino por su legibilidad por parte del analista o espectador. Los enlaces que se entrecruzan y los nodos que se superponen unos sobre los otros producen una pobre legibilidad. Los programas de visualización de grafos tratan de minimizar el número de intersecciones, de dar suficiente espacio en blanco alrededor de cada nodo para separarlos de sus nodos vecinos y de maximizar la simetría. Para ello implementan diferentes algoritmos de distribución espacial de nodos y de enlaces que muestran por ejemplo, las redes de forma aleatoria, circulares o con diseños bipartitos.

En una visualización de redes aparecen conjuntos de nodos muy relacionados, donde muchos comparten los mismos enlaces. Visualizar estos subconjuntos ayuda a comprender las relaciones. Pero los conjuntos de nodos no se detectan siempre de forma sencilla en la red, incluso si los nodos comunes tienen el mismo color. Un modo de destacar conjuntos similares es marcándolos con colores diferentes a modo de regiones en los mapas convencionales y etiquetando las regiones para permitir al observador comprender lo que se muestra. De este modo se pueden mapear, por ejemplo las preferencias de compra seguidas por los consumidores y ayudar al consumidor para que si está interesado, explore ese nuevo paisaje de información (figura 6).



Figura 8. <http://www.manyeyes.com>

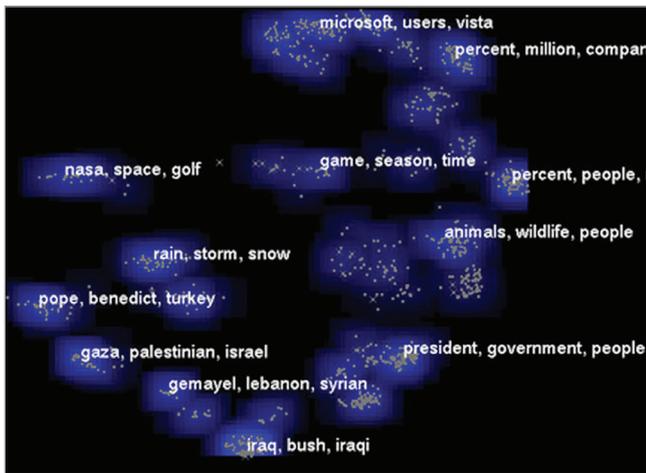


Figura 9. <http://in-spire.pnnl.gov>

4. Visualización de textos

La visualización de textos es una clase de técnica de visualización de información que transforma objetos contenidos en los *corpora* o colecciones de documentos, en un formato visual. Estos objetos pueden ser palabras, sentencias, párrafos, atributos y metadatos, a nivel individual o en colecciones. Si se identifican entidades importantes en el seno de los textos, se puede analizar su contenido en tres niveles de representación: en un nivel léxico, sintáctico o semántico. Por ejemplo mediante la obtención de frecuencias o distribuciones de palabras; el contenido semántico y repeticiones; el tema o los temas que definen conjuntos de documentos o los temas centrales de los documentos. También pueden detectarse semejanzas mediante conexiones, cambios o fenómenos de difusión de la información.

La visualización de un solo documento permite conocer los temas centrales del contenido y las características del texto de forma rápida. La forma más sencilla de visualización se conoce como nubes de textos o palabras, y no son sino representaciones de la frecuencia en la que aparecen términos, en los que se usan vectores de los términos y algún algoritmo de diseño para que el usuario modifique la fuente, el color o la configuración (figura 7).

En los análisis de la literatura es un lugar común analizar un texto o una colección de textos, extrayendo una concordancia: un índice de todas las palabras que contiene un texto, en el que se muestran las palabras en el contexto en que aparecen. Es un modo similar a presentar palabras claves en contexto (KWIC, *keyword in context*). Permite comprender la frecuencia de términos y su distribución, o la creación de estructuras de conocimiento en bases de datos de naturaleza léxica o de otro tipo. Otro tipo de visualización puede reflejar el contenido semántico, es decir la representación semántica asociada al lenguaje. Normalmente se extraen tripletas sujeto-verbo-atributos del objeto, después de pre procesar el texto, extraer entidades, resolver anáforas (elementos que dependen del contexto) y normalizar de forma semántica. Por últimos se les dota de apariencia visual, en forma de redes semánticas (figura 8).

Buscamos texturas de datos con valor, en contextos comparables, que ayuden a comprender lo que la complejidad revela, con economía de medios

En la mayoría de los casos, el objetivo de la visualización de documentos es ubicar documentos semejantes cerca unos de los otros, después de computar su semejanza entre pares de documentos y definir un diseño de la vista. Se trata de mapear matrices de términos muy complejas y traducirlas en representaciones de menores dimensiones que retengan tanto significado como sea posible. Normalmente se emplean métodos de descomposición y de factorización, como el de descomposición de valores singulares (*singular value decomposition, SVD*), análisis de componentes principales o mapas autoorganizativos (Vargas-Quesada; De-Moya-Aneón, 2007).

La visualización de los temas que se contienen en grandes colecciones persigue descubrir temas concretos o las relaciones entre diversos temas. Se emplea en el descubrimiento de disciplinas emergentes, evoluciones, detección de ten-

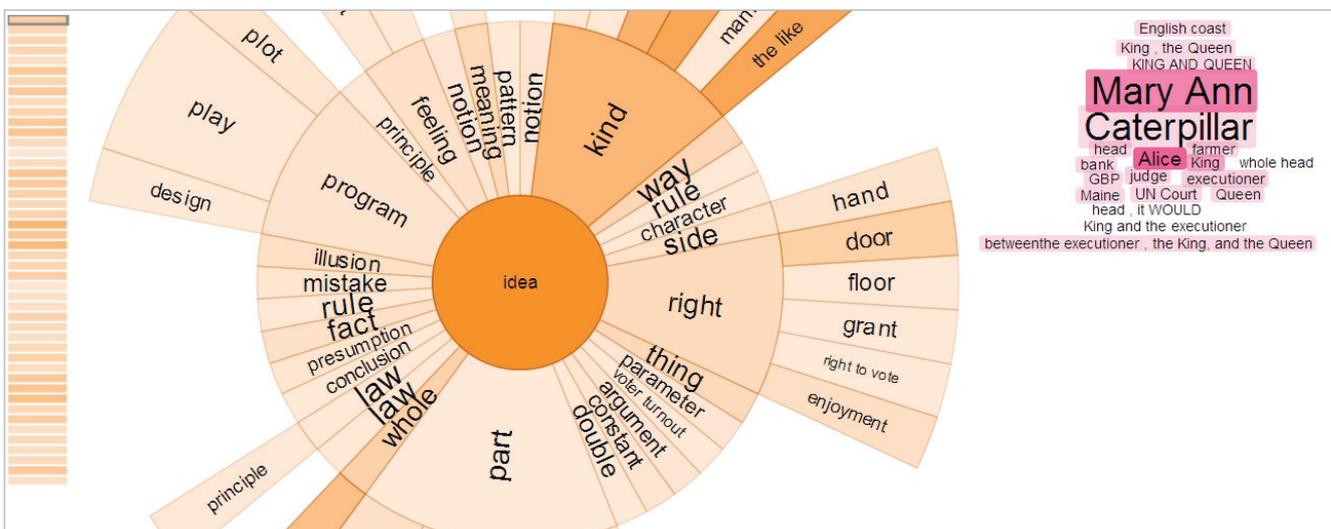


Figura 10. <http://vialab.science.uoit.ca>

dencias o rupturas de paradigmas (Chen, 2006). Son resúmenes de *corpora* que emplean paisajes en 2D o 3D, en el que el peso y el color se usan para representar densidades de documentos semejantes. Otros sistemas emplean redes, tablas o mapas geoespaciales (figura 9).

Los textos electrónicos cambian en sus distintas versiones, y visualizar las diferencias entre las múltiples adaptaciones puede ser objeto de un tipo de visualización específica. Las visualizaciones deben asumir que los datos de entrada progresan a lo largo del tiempo. Los diseños visuales de datos textuales temporales muestran las diferencias entre las múltiples versiones, las correcciones efectuadas, los patrones de cooperación en el caso de textos electrónicos con multiautoría (*Wikipedia*), el predominio o no del anonimato frente a las autorías reconocidas, las formas de negociación de contenidos, las zonas de texto estables o inalteradas, las procedencias geográficas de los lectores... (figura 10).

La visualización de literaturas científicas y las relaciones entre citas de los autores que escriben artículos científicos, es un campo sumamente amplio de análisis. La visualización de las formas en las que los autores, las revistas y los temas citan y son citados da origen a la generación de grafos de coautoría, mapas de la ciencia o sistemas interactivos con interfaces visuales para acceder a información bibliométrica (figura 11).

En un mundo moldeado de forma creciente por algoritmos, la visualización de información juega un papel relevante para dotarle de sentido

5. Final

La representación de estructuras de información multidimensional mediante sistemas de información no es un asunto trivial. Los procesos de diseño requieren tanto métodos de razonamiento analítico, como espaciales y visuales, además de múltiples habilidades en codificación. La creación obliga a codificar, el uso a decodificar. Las visualizaciones de información generadas por ordenador, dinámicas e interactivas ya son una realidad tal y como señalaba Edward Tufte, y sin duda, nos acompañarán en el porvenir, porque “capturan una parte valiosa de la sutileza del itinerario humano”.

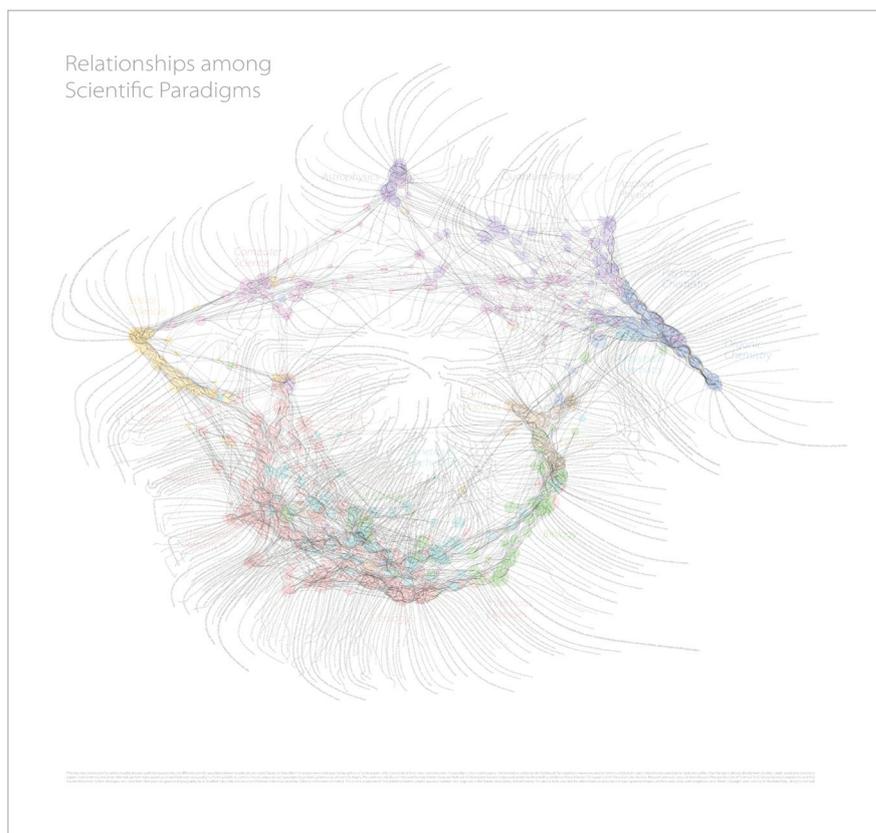


Figura 11. <http://wbpaley.com>

6. Bibliografía

Card, Stuart K. (1999). *Readings in information visualization, using vision to think*. Stuart K. Card, Jack D. Mackinlay, Ben Shneiderman (Eds.). San Francisco: Morgan Kaufmann. ISBN: 15 586 0533 9

Chen, Chaomei (2004). *Information visualization: Beyond the horizon*. London: Springer. ISBN: 18 523 3789 3

Costa, Juan (1998). *La esquemática. Visualizar la información*. Barcelona: Paidós. ISBN: 84 493 0611 6

Shepard, Roger N. (1990). *Mind sights: Original visual illusions, ambiguities, and other anomalies. With a commentary on the play of mind in perception and art*. New York: WH Freeman & Co. ISBN: 9780 716 72133 8

Shneiderman, Ben; Plaisant, Catherine (2006). *Diseño de interfaces de usuario: estrategias para una interacción persona-computadora efectiva*. Madrid: Pearson Addison-Wesley. ISBN: 9788 420 54803 6

Tufte, Edward R. (1990). *Envisioning information*. Cheshire: Graphic Press.

Vargas-Quesada, Benjamín; De-Moya-Anegón, Félix (2007). *Visualizing the structure of science*. Berlin: Springer. ISBN: 9783 540 69727 5

Ware, Colin (2012). *Information visualization: perception for design. 3rd ed.* Waltham (MA): Morgan Kauffmann. ISBN: 15 586 0511 8

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

HOW WE DRAW TEXTS: A REVIEW OF APPROACHES TO TEXT VISUALIZATION AND EXPLORATION

Jaume Nualart-Vilaplana, Mario Pérez-Montoro y Mitchell Whitelaw

Nota: Este artículo puede leerse traducido al español en:
http://www.elprofesionaldelainformacion.com/contenidos/2014/may/02_esp.pdf



Jaume Nualart-Vilaplana is a PhD candidate in the *Faculty of Arts and Design, University of Canberra* (Australia), research engineer at *Nicta* (Australia), and a PhD candidate in the *Faculty of Information Science, University of Barcelona*. MAS and MSc (Licenciatura) at *Autonomous University of Barcelona*

<http://orcid.org/0000-0003-4954-5303>

*Machine Learning Research Group at NICTA, Canberra Research Laboratory
Tower A, 7 London Circuit, Canberra City ACT 2601, Canberra, Australia
jaume.nualart@canberra.edu.au*



Mario Pérez-Montoro holds a PhD in Philosophy and Education from the *University of Barcelona* and a Master in Information Management and Systems from the *Polytechnic University of Catalonia*. He studied at the *Istituto di Discipline della Comunicazione* at the *Università di Bologna* (Italy) and has been a visiting scholar at the *Center for the Study of Language and Information (CSLI)* at *Stanford University* (California, USA) and at the *School of Information at UC Berkeley* (California, USA). He is a professor in the *Department of Information Science* at the *University of Barcelona*. His work has focused on information architecture and visualization. He is author of the book *Arquitectura de la información en entornos web* (Trea, 2010).

<http://orcid.org/0000-0003-2426-8119>

*Facultat de Biblioteconomia i Documentació, Universitat de Barcelona
Melcior de Palau, 140. 08014 Barcelona, España
perez-montoro@ub.edu*



Mitchell Whitelaw is an academic, writer and practitioner with interests in new media art and culture, especially generative systems and data-aesthetics. His work has appeared in journals including *Leonardo*, *Digital creativity*, *FibreCulture*, and *Senses and society*. In 2004 his work on a-life art was published in the book *Metacreation: art and artificial life* (MIT Press, 2004). His current work spans generative art and design, digital materiality, and data visualisation. He is currently an associate professor in the *Faculty of Arts and Design* at the *University of Canberra*, where he leads the *Master of Digital Design*. He blogs at *The Teeming Void*.

<http://orcid.org/0000-0001-9013-9732>

*Faculty of Arts and Design, University of Canberra
Bldg, Floor & Room: 9, C12. ACT 2617, Canberra, Australia
mitchell.whitelaw@canberra.edu.au*

Abstract

This paper presents a review of approaches to text visualization and exploration. Text visualization and exploration, we argue, constitute a subfield of data visualization, and are fuelled by the advances being made in text analysis research and by the growing amount of accessible data in text format. We propose an original classification for a total of 49 cases based on the visual features of the approaches adopted, identified using an inductive process of analysis. We group the cases (published between 1994 and 2013) in two categories: single-text visualizations and text-collection visualizations, both of which can be explored and compared online.

Keywords

Review, Text visualization, Data visualization, Data exploration, Data display, Information visualization, Text analysis.

Título: Cómo dibujamos textos. Revisión de propuestas de visualización y exploración textual

Article received on 19-01-2014

Approved on 09-03-2014

Resumen

En este trabajo se presenta una revisión de estrategias para la visualización y exploración de textos. Se argumenta que la visualización y exploración de textos constituye un subcampo de la visualización de datos que se nutre de los avances en el análisis de textos y de la creciente cantidad de datos accesibles en formato texto. Proponemos una clasificación original para un total de cuarenta y nueve casos revisados. La clasificación está basada en las características visuales de cada caso, identificadas mediante un proceso inductivo de análisis. Agrupamos los casos (publicados entre 1994 y 2013) en dos categorías: las visualizaciones de texto individuales y la visualizaciones de colecciones de textos. Los casos revisados pueden ser explorados y comparados en línea.

Palabras clave

Visualización de texto, Visualización de datos, Exploración de datos, Visualización de información, Análisis de textos.

Nualart-Vilaplana, Jaume; Pérez-Montoro, Mario; Whitelaw, Mitchell (2014). "How we draw texts: a review of approaches to text visualization and exploration". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 221-235.

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

1. Introduction

The aim of this review is to propose a classification of text visualization and exploration tools, while describing the broader context in which they operate. To do so, we list, classify and discuss the most important contributions made in the field of text visualization and exploration between 1994 and 2013. This field is undergoing rapid growth –fuelled by open data initiatives and web scraping– and has become highly diversified, developing in parallel in a range of disciplines. Some of the most important visualization methods invented between 1765 and 1999 were the timeline, bar chart, pie chart, flow map, Venn diagram, histogram, Gantt chart, flowchart, tag cloud, social networks, boxplot, star plot, treemap, headmap, and sparkline. Figure 1 presents a word cloud (using *Wordle*) of the professions practiced by their respective inventors. Given this diversity, our search for cases has been conducted in many different contexts and has involved the examination of many different sources, ranging from the sciences to the humanities, from academic journals to blog sites, from universities to freelance studios, and from open data institutions to open data communities. Clearly this proliferation of disciplines has meant the adoption of a variety of different philosophies and points of view.

This review aims to help those that work with data, and especially with texts (but by no means limited to academics), to use visualization techniques that can identify patterns or behaviours present in the textual reality. Moreover, these techniques can help users improve –in terms of both the

speed and the clarity of the process– the way in which they visualize and discover the facts that lie within the data.

Drawing a clear conceptual line between approaches to text visualization and exploration is no straightforward task, but here we have opted to review cases dedicated to both processes, be they described separately or together. Note that on occasions, for the sake of simplicity, we use the term text visualization in reference to both approaches.

The two types of text visualization considered here are:

1) Single-text representation, that is, ways of extracting meaning from texts based on writing style, document structure and language register as opposed to pure statistics. Our interest lies in representing the meaning and salient features of texts because their convenient visualization can speed up and/or improve our ability to select texts and manage the time required to tackle them. The research output of fields such as natural language processing, linguistic computing and machine learning provides techniques for producing high quality data representing complex texts. It is our belief that by combining these techniques with a suitable text visualization method we can improve the way in which we examine and understand texts.

2) Representation and exploration of collections of texts. Exploring and selecting individual texts and navigating and analyzing collections of texts are daily tasks for many of those who work with computers and datasets, and there is clearly plenty of room for new ideas and tools to facilitate

their work. Information retrieval is a critical factor in an environment characterized by an excess of information (Baeza-Yates *et al.*, 1999). When a user conducts a search, the information retrieval systems normally respond with a list of results. More often than not, the presentation of these results plays an important role in satisfying the user's information needs, so a poor or inad-

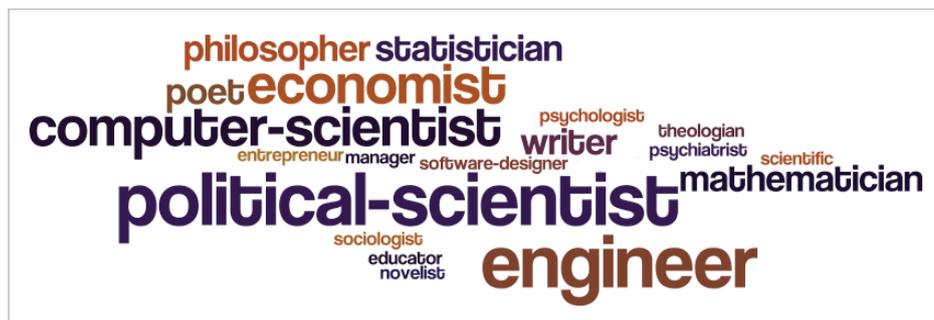


Figure 1. Word cloud of the professions practiced by inventors of visualization methods

Table 1. Leading universities and their data visualization departments

Institution	Rank in 2012	Department/Course	URL
Harvard University	1	Broad Institute of Harvard and MIT	http://www.broadinstitute.org/vis
Massachusetts Institute of Technology	2	Broad Institute of Harvard and MIT	http://www.broadinstitute.org/vis
University of Cambridge	3	--	--
Stanford University	4	Stanford Vis Group	http://vis.stanford.edu
University of California, Berkeley	5	VisualizationLab	http://vis.berkeley.edu
University of Oxford	6	Visual Informatics Lab at Oxford	http://oxvii.wordpress.com
Princeton University	7	PrincetonVisLab	http://www.princeton.edu/researchcomputing/vis-lab
University of Tokyo	8	--	--
University of California, Los Angeles	9	IDRE GIS and visualization	https://idre.ucla.edu/visualization
Yale University	10	--	--

Table 2. Conferences dedicated primarily to data visualization ordered by number of participants (Stefaner, 2013)

Conference	Location	Topic	No. participants	URL
Nicar 2013	USA	Data journalism	149	http://ire.org/conferences/nicar-2013
Dd4d 2009	France	Information visualization	52	http://www.dd4d.net
FutureEverything 2013	UK	Technology/society/art	52	http://futureeverything.org
Resonate 2013	UK	Creative code	44	http://www.thisisresonate.co.uk/resonate-13
Graphical web 2012	Switzerland	Open web/datavis	38	http://www.graphicalweb.org/2012
IEEE Vis - VisWeek 2012	USA	Information visualization	-	http://ieevis.org
EuroVis 2013	Germany	Computational aesthetics	-	http://www.eurovis2013.de
Siggraph 2013	USA	Computer graphics and interactive techniques	-	http://s2013.siggraph.org
OzViz 2012	Australia & NZ	Workshops for visualisation practitioners, academics and researchers	-	http://www.ozviz2012.org

equate presentation can thwart the user (Baeza-Yates *et al.*, 2011). Typically, information retrieval systems present the results of a query in a flat, one-dimensional list. Such lists tend to be opaque in terms of the order they give to the information, i.e., the users are unaware as to why the list is presented in a particular order. To refine their search, users have to interact again, normally by filtering the first output of results. It is our belief that new techniques for representing collections of texts –including search results– can help improve navigation, exploration and retrieval.

As we show below, data visualization can today be considered a consolidated academic field (Strecker; IDRC, 2012). Thus:

- Seven of the top 10 universities according to the *Times Higher Education ranking* (2012) have departments or research groups working in the field of data visualization. The discipline is incorporated in a wide variety of departments, ranging from computer science and statistics to linguistics and graphical design, and from chemistry and physics to genetics and history. Recently, data visualization has emerged as a distinct field, with specific departments dedicated to its study and master's programs being taught in the subject (table 1).

- Over the last five years a number of conferences have been dedicated primarily to data visualization. These are listed in table 2.

- A number of journals are now specifically dedicated to studies in data visualization, and important contributions can be found also in conference proceedings (table 3).

Finally, a number of leading websites –including *Infosthetics*, *Visualcomplexity* and *Visualizingdata.com*– play a key role in the dissemination of the subject.

1.1. Text visualization

Shneiderman (1996) classifies regular texts as one-dimensional data, that is, data organized in a sequential manner, running right-to-left (or left-to-right), line-by-line, top-to-bottom. Yet, a text can have multiple internal structures, a morphology made up of paragraphs, sentences and words.

Table 3. Main journals dedicated to data visualization

Name	Url
<i>Parsons journal for information mapping</i>	http://pjim.newschool.edu/issues/index.php
<i>Journal of visualization</i>	http://springer.com/materials/mechanics/journal/12650
<i>IEEE Transactions on visualization and computer graphics</i> (TVCG)	http://www.computer.org/portal/web/tvcg
<i>Information visualization</i>	http://ivi.sagepub.com
<i>International journal of image processing and data visualization</i> (Ijipdv)	http://iartc.net/index.php/Visualization
<i>IEEE Vis</i> (former <i>Visweek</i>)	http://ieevis.org
<i>EuroVis</i>	http://www.eurovis2013.de
<i>ACM CHI</i>	http://chi2013.acm.org
<i>EG CGF</i>	http://www.eg.org
<i>IVS</i>	http://www.graphicslink.co.uk/IV2013

Depending on its information structure, a text may be ordered by chapters, parts, sections, subsections, etc. If a text is given in a specific format, such as html, then it may be organized into bodies, divs, paragraphs, etc. In these examples the text includes tree structures as well as a one-dimensional structure. Additionally, texts may have a subjective component and an abstract structure that is not readily analysed by a computer. All in all, these data types and structures constitute the specificities of a text.

The amount of data to which we have access grows on a daily basis. Most of these data are in text format, as **Fernanda Viégas** and **Martin Wattenberg** in an interview with **Jeff Heer** argue: “One of the things I think is really promising is visualizing text. That has been mostly ignored so far in terms of information visualization approaches, and yet a lot of the richest information we have is in text format” (**Heer**, 2010).

Seven of the top 10 universities have departments or research groups working in data visualization

Data analysis defines the boundaries of data visualization, i.e., it provides the fine line between multiple truths and lies. In the case of text visualization, this role has been taken on by text analysis: in the main, via computational linguistics, natural language processing, machine learning and statistics. The advances made in text analysis at a whole range of levels have provided computers with text understanding, enabling them to modify a text, the so-called unstructured data (see next subsection “Text analysis”).

There is some discussion as to whether text visualization might be considered a specific subfield of data visualization. Some authors tend to disagree: **Illinski** (2013) claims that text cannot be considered a data type; **Šilić** (2010) argues that “unstructured text is not suitable for visualization”. Yet, as discussed above, most text visualizations transform the initial “unstructured” textual data into a reduced, structured dataset. This new dataset is no longer one-dimensional, but rather it constitutes a categorical or a network dataset and it can be represented with a wide range of tools that are not specific to text representation (**Hearst**, 2009; **Grobelnik**; **Mladenić**, 2002).

As we show in the cases we review here, most text visualizations do not represent raw data: that is, the text as it is. Rather what they do is transform the text into smaller chunks of data, normally extracting a representative part of that text. This process is one of data transformation and it occurs, for example, when a text is reduced to a list of words based on their frequency of appearance. In that case, the method chosen to represent the data will belong to a family of methods best suited to the data type. In this review we consider the most frequently employed strategies to represent single texts or collections of texts, paying special attention to strategies for representing textual data as it is, as a regular text, with all its complexities, irregularities and rich abstractions.

Text analysis is a key field for text visualization. Below, we present a brief commentary on this matter and its relationship with text visualization.

1.2. Text analysis

Text analysis, roughly synonymous with text mining (**Feldman**; **Sanger**, 2006), is an interdisciplinary field that includes information retrieval, data mining, machine learning, statistics, linguistics and natural language processing. According to **Marti Hearst** (2003), the goal of text mining is to discover “heretofore unknown information, something that no one yet knows and so could not have yet written down”. Text mining is a subfield of data mining whose typical applications include the analysis or comparison of literary texts, the analysis of biological and genomic data sequences and, more recently, the identification of consumer behaviour patterns or the detection of the fraudulent use of credit cards. **Hearst** differentiates these applications from information extraction operations, such as the extraction of people’s names, addresses or job skills. This latter task can be done with >80% accuracy, but the former, the full interpretation of natural language by a computer program, looks like it will not be possible for “a very long time” (**Hearst**, 2003).

To study text visualization and exploration it is important to examine the literature dedicated to both data visualization and text analysis, given the significant interrelationships that exist. Thus, while the text analysis output may limit the possibilities of visual presentation and interaction with the text, there is strong empirical evidence indicating that people learn better with a combination of text and illustration (visualization) than with text alone (**Anglin et al.**, 2004; **Levie**; **Lentz**, 1982).

2. Review

In this section we propose a possible classification based on the visual features that characterize the approaches to textual visualization and exploration, as identified in 49 cases.

The methodology to collect the cases is a two-part process. First, a traditional literature search and review (including practical examples and visualisation studies); and second, a subset of these have been selected, based on a preliminary analysis of their features. The aim was to select cases that provided a representative overview of the range of work in the field.

The classification of the cases is the product of empirical observation following an inductive analysis. The classification is followed by an analysis of these cases.

There are alternatives to those used in this paper for the selection and categorization of primary source methodologies such as **Kitchenham** (2004) and **Benavides**; **Segura**; **Ruiz-Cortés** (2010).

2.1. Classification of approaches

The basic classification of text visualization approaches comprises two categories according to the type of data to which they are applied:

1) Textual documents: that is, representations of single texts, where text is understood as a sequence of words ordered according to the hierarchy: document > paragraphs

> sentences > other punctuation marks > words > syllables and phonemes or morphemes. Where a text is a book or another kind of structure, then, it may have more granularities, including: chapters > sections > sub sections > etc. We also include the metadata of the text and other attached texts, i.e., title, author(s), publisher, copyright notes, acknowledgement, dedication, preface, table of contents, forward, glossary, bibliography, index, etc.

2) Text collections: that is, a group of texts in which each item constitutes a clearly differentiable entity. Typically when speaking of collections of texts, we speak of texts that have elements in common, be it their register, length or structure. All the cases we review here are collections of the same text type. Heterogeneous collections of texts are also referenced in the literature (Meeks, 2011), especially in representative analyses of a field of knowledge, where the aim is to include the greatest possible variety of expressions and vocabulary. In such cases the dataset can be said to be heterogeneous in term of its structure and register.

To these two data types, we then add several subjective subdivisions to each category according to the visual features used to represent the textual features. The aim here is to be able to describe and explain the cases under review, as well as to identify the key features of the text visualization approaches.

Single texts

- Whole <-> Part
- Sequential <-> Non sequential
- Discourse structure <-> Syntactic structure
- Search
- Time

Text collections

- Items <-> Aggregations
- Landscape
- Search
- Time

2.1.1. Single texts

In the specific instance of single texts, we classify the cases according to the part of the text that is represented, whether the approach follows the same sequence as that of the text, and the text structure employed in each case.

Whole or part?

In some instances, one part of the text is considered the essence of the text and is used in the visualization process rather than the whole text. Yet, there are processes that use the whole text, at least implicitly. Examples include:

- chapters of a book but not the whole text.
- representation of all the sentences of the text as coloured lines.
- verbs of a text, providing an impression of the style of the text.
- characters of a novel and their appearance within the text.
- places or dates present in the text.
- etc.

The cases in which the whole text is explicitly represented are, for obvious reasons, cases involving relatively short texts, e.g., song lyrics, speeches, poems, etc.

In some instances, such as when using *Radial word connections* (see, case 1 below) only certain words from the text are represented; yet, we classify this case as a whole text representation because the whole novel, chapter by chapter, is implicitly represented in the circle.

In those instances in which the whole text is represented (even implicitly) as one central element in the visualization, we classify it as being a whole-text visualization.

Does the visualization follow the same sequence as that of the text?

If the visualization follows the same sequence, or order, as that of the text, then the case is considered sequential; if not, then it is considered non-sequential. For example, a typical case that does not follow the same sequence as that of the original text would be a word cloud (see figure 1).

“ Most text visualizations transform the initial ‘unstructured’ textual data into a reduced structured dataset ”

Does the visualization use elements from discourse structure or from syntactic structure?

A text may present one of two kinds of structure that we consider useful for our research. One is so-called discourse structure. Depending on the nature of the text, the discourse structure can be completely subjective to the author’s point of view—as in literature—, or restricted to a given structure—as in legal and scientific texts. In linguistics, discourse is a broad concept, but here we use it to refer to the parts of a text and the outline of a document: parts, chapters, sections, subsections, etc. The discourse structure is widely used when visualizing texts because it is a relatively straightforward way to represent the text sequence.

The second structure is the text’s syntactic structure, referred to text structure in sentences, phrases and word classes—including verbs and nouns. This is an objective structure and is dependent on the rules of linguistics. In text visualizations, the elements comprising this structure, such as sentences, are very common.

2.1.2. Text collections

In the specific instance of text collections we classify the cases according to pure items or aggregations, i.e., as pure data or data landscapes. Thus we determine whether the items making up the collection can be differentiated or represented as aggregations. The specific questions we address are: How is each item in the collection graphically represented? Is each text represented as a graphical entity, i.e., as a point, a word or short sentence? Can the items in the visualization be counted, i.e., are they visually differentiated?

There are cases in which each item is not represented by a graphically distinct entity, but rather, for example, as a coloured block. Alternatively, the items are accumulated and shown as frequency distributions. When the items of the collection are not graphically distinct (visually countable)



Figure 2. The 49 reviewed cases visualized with the *Area* software (screen shot).

then we speak in terms of the visualization of an aggregation rather than that of an item.

Pure data or data and landscape?

Are the items of the collection accompanied by any graphical content? Is another dataset, apart from that emanating from the text, also being represented? Some cases present the items embedded in a graphical environment, such as a map. This context might be an actual geographical map, a metaphor, or, for example, a conceptual landscape composed of words that form a second layer complementing that of the data collection, in which every distance plays a role: item-item (similarity between documents), word-item (importance of a word in a document), word-word (similarity between words in the collection).

Scales and axes are not considered as landscapes, nor are the elements of the interface in which the representation is embedded. This data layer, if not considered as the main dataset, would reduce substantially Tufte's data-ink ratio (Tufte; Graves-Morris, 1983) compared to the ratio of a pure data representation.

2.1.3. Both single texts and text collections

Properties that are equally applicable to single-text and text-collection visualizations include time, search results and dataset size.

Does time play a role?

Do the texts change over time? One set of visualization approaches highlights the changes undergone by a dataset over time. The most common approaches of this kind have been developed in computer science to represent code evolutions or in *Wikipedia* to indicate various aspects of article revisions.

This category also includes visualizations in which the dataset itself changes over time; for example, the visualization of the latest news will see the dataset grow over time.

Does the visualization result from a search query?

Visualizations of the output of information system retrieval is a well-defined kind of visualization characterized by the changing number of represented items depending on the number of search results obtained. This is a growing visualization subfield related to the disciplines of information systems and information retrieval (Mann, 2002; Hearst, 2009).

Validity for small or large datasets

It is rare that a visualization tool is independent of the size of the dataset that is to

be represented. Here, in those cases in which the tool has been clearly designed for a specific dataset size, the reader will be given the corresponding explanation.

2.2. Analysis of visualization approaches

We review a total of 49 cases applying the classification outlined above. In an attempt to incorporate the most crucial aspects of text visualization, our review concentrates on the specific ideas underpinning the text visualization, rather than the dataset and the contexts of each case.

Sixteen fields have been collected for each case: name, short name, author(s), year of publication, URL for further information, original dataset, discipline related to the work, description of the visualization method, description of the case, screen shot, thumbnail, classification (single or collection), classification (single-whole, single-part, collection-items, collection-aggregations), classification (time), classification (search), classification (dataset small, dataset large, N/A).

The cases are grouped into two sections and four subsections:

Single-text visualizations (23 cases)

- Whole-text visualizations (15 cases)
- Partial-text visualizations (8 cases)

Text collection visualizations (26 cases)

- Collection of items (16 cases)
- Collection of aggregations (10 cases)

For each subsection the cases are sorted by year of publication (descendant). To assist the reader, the collection of all reviewed cases can be viewed using the visualization and exploration software (also included in the review) known as *AREA* (Nualart, 2013).

2.2.1 Single-text visualization

We present single texts grouped as whole-text visualizations, partial-text visualizations and other subcategories.

The latter includes sequential and non-sequential visualizations, discourse-structures and syntactic-structures visualizations, search results and datasets dependent on time visualizations. Each subsection adheres to the following structure: list of cases, description of the group and discussion.

a) Whole-text visualizations

- 1) Literature. *Novel views: Les misérables, Radial word connections* by Jeff Clark (2013)
- 2) Literature. *Novel views: Les misérables, Character mentions* by Jeff Clark (2013)
- 3) Literature. *Poem viewer* by Katharine Coles et al. (2013)
- 4) Politics. *State of the Union 2011, Sentence bar diagrams* by Jeff Clark (2011)
- 5) Literature. *Visualizing lexical novelty in literature* by Matthew Hurst (2011)
- 6) Science/papers. *On the origin of species: The preservation of favoured traces* by Ben Fry (2009)
- 7) Science/papers. *Texty* by Jaume Nualart (2008)
- 8) Religion. *Bible cross-references* by Chris Harrison (2008)
- 9) Literature. *Literature fingerprint* by Daniel A. Keim and Daniela Oelke (2007)
- 10) Wikipedia. *History flow* by Fernanda Viégas and Martin Wattenberg (2003)
- 11) Literature. *Colour-coded chronological sequencing* by Joel Deshayé and Peter Stoicheff (2003)
- 12) Literature. *2-D display of time in the novel* by Joel Deshayé (2003)
- 13) Literature. *3-D display of time in the novel* by Joel Deshayé (2003)
- 14) Any. *Wattenberg's arc diagram* by Martin Wattenberg (2002)
- 15) Health. *TileBars* by Marti A. Hearst (1995)

Description

- Number of cases: We identify 15 cases that can be categorized as whole-text visualizations.
- Years: The cases were published over an 18-year period from 1995 to 2013.
- Authors: All the authors work in academic fields. The most prolific authors in this category are Jeff Clark and Joel Deshayé (with three cases each), followed by Martin Wattenberg (with two cases).
- Datasets: Most of the text corpora in this category are taken from literature (eight cases). Most authors draw on novels, especially well-known texts such as the classics, to demonstrate new visualization approaches.
- Methods: All the cases except case 14 (*arc diagram*) use colour as part of the visualization method. Five cases use

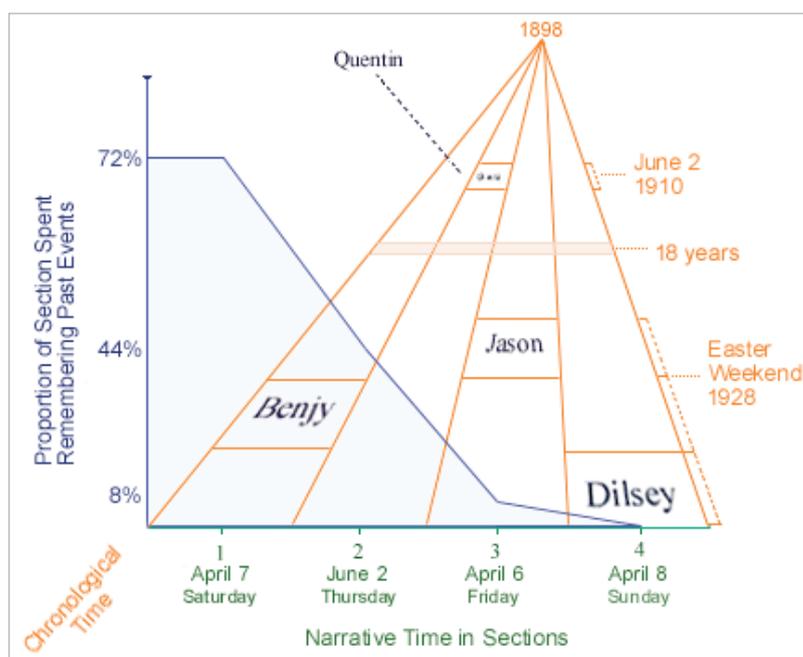


Figure 3. (Case 13) 3-D display of time of William Faulkner's novel *The Sound and the Fury*, by Joel Deshayé and Peter Stoicheff (2003)

methods that are bar chart derivatives (cases 4, 5, 6, 9 and 11). Three cases use curves connecting parts of the texts: two arcs and one radial diagrams (cases 1, 8 and 14).

Discussion

A common method cannot be identified for these whole-text visualizations. Yet, as expected, they all present an axis representing the whole text. In 13 of the 15 cases, the text line is represented by a horizontal or vertical line. The two exceptions use a circle—the case of *Radial word connections* (case 1)—and an iconification of a text on the page—the case of *Texty* (case 7).

Since whole-text visualizations always include an abstraction of the text, referred to as its text line, a question arises: which part of the text is physically present in the whole-text visualization being reviewed? Interestingly, nine of the 15 visualizations do not show a single word (cases 4, 5, 6, 7, 8, 9, 10, 11 and 15). Four cases show a small number of words (cases 1, 2, 12 and 13) (figure 3), while only two cases show all the text (cases 3 and 14).

The most common approach is to show the occurrence of a certain feature—this might be a term, topic, cross-reference or character—within the text as a whole (all cases except 3, 12, 13 and 15). With the exception of Wattenberg's *arc diagrams* (case 14), these occurrences are represented using the same colour.

It is interesting to observe how very similar data are represented in very different ways depending on the case under review. For example, while Viégas and Wattenberg's *History flow* (case 10) and Fry's *Favoured Traces* (case 6) both present document-version histories by section, the former is spatialized and the latter animated. Similarities, however, are seen in the approaches adopted, for example, by *TileBars* (case 15) and *Texty* (case 7). Thus, both highlight words from the text within a rectangular figure that is representa-

tive of the whole text. Other cases use opposite or complementary techniques. Thus, Wattenberg’s *Arc diagram* (case 14) shows repetitions while Hurst’s novelty visualization (case 5) shows only new strings, and no repetitions.

Literature and other complex texts, such as political speeches (case 4) and the *Bible* (case 8), dominate the type of corpora used in this category (10 cases). This is perhaps surprising, as these texts tend to be complex, often presenting a high level of abstraction and little formal structure. Arguably, when opting to introduce or test a new approach, it would make more sense to work with simpler, more structured texts (such as scientific papers, patents, health diagnostics, etc.) that present greater regularity in terms of their vocabulary, text length, discourse structure and register. Given the inherent freedoms associated with literature, novelists are under no obligation to adhere to any pattern or rule that might help us give structure to the unstructured.

However, depending on how the text is treated and processed, the nature of the text is not always relevant. For example, Matthew Hurst (case 5) tracks the introduction of new terms in literary texts. Yet the tool can be applied to any other text type, its results being unrelated to the complexity of the text given the ubiquity of the method. Having said this, it would be interesting to apply the technique to scientific papers in which the style is much more clearly defined. Similar arguments can be applied to *Radial word connections* (case 1), *Sentence bar diagrams* (case 4) and *Literature fingerprints* (case 9).

b) Partial-text visualizations

- 16) Literature. *Novel views: Les misérables. Characteristic verbs* by Jeff Clark (2013)
- 17) Any. *Wordle* by Jonathan Feinberg (2009)
- 18) Books. *DocuBurst* by C. Collins, S. Carpendale and G. Penn (2009)
- 19) Literature. *Phrase nets* by Frank van Ham, Martin Wattenberg and Fernanda B. Viégas (2009)

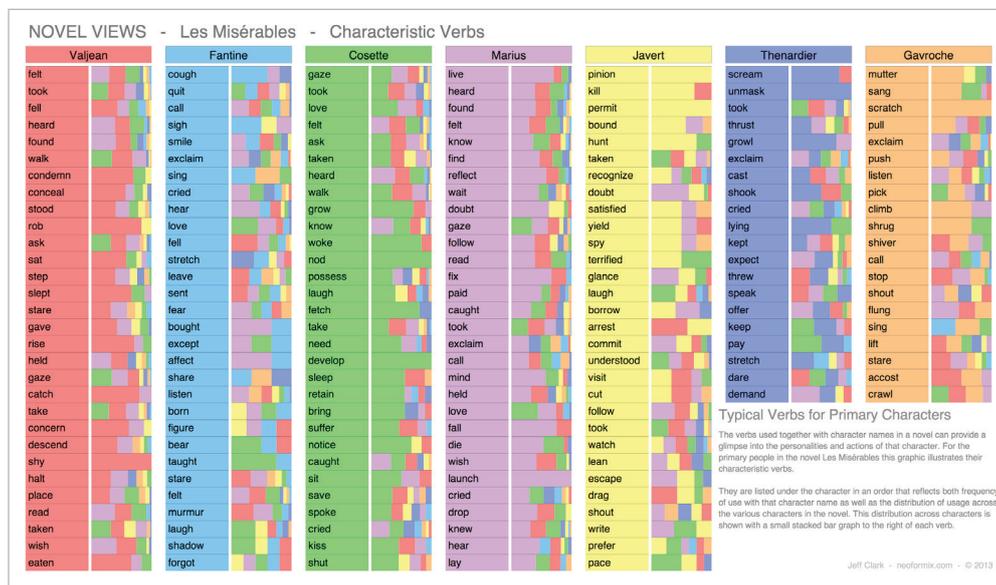


Figure 4. (Case 16) *Novel views: Les misérables. Characteristic verbs* by Jeff Clark (2013)

- 20) *Google* data. *Word spectrum: Visualizing Google’s bi-gram data* by Chris Harrison (2008)
- 21) *Google* data. *Word associations: Visualizing Google’s bi-gram data* by Chris Harrison (2008)
- 22) Literature/songs. *Document arc diagrams* by Jeff Clark (2007)
- 23) Any book. *Gist icons* by P. DeCamp, A. Frid-Jimenez, J. Guinness, D. Roy (2005)

Description

- Number of cases: We identify eight cases that can be categorized as partial-text visualizations.
- Years: The cases were published over an eight-year period from 1995 to 2013.
- Authors and datasets: Two cases by Jeff Clark (cases 16 and 22) and one by the creative team of Wattenberg and Viégas in collaboration with van Ham (case 19) use literary texts. The two cases by Chris Harrison use large *bi-gram* datasets published by *Google*. One case is not dependent on the nature of the text: *Wordle* (case 17), the very popular “word cloud” method introduced by Feinberg. Finally, two interactive approaches involving large datasets are presented: *DocuBurst* (case 18) and *Gist icons* (case 23).
- Methods: In six of the eight cases (cases 16, 17, 18, 19, 22 and 23), the dataset is reduced to what is called a bag of words and only these words are present in the visualization. Cases 20 and 21 are representations of all bi-grams that pit two primary terms against each other.

Discussion

Partial-text visualization is a successful, popular way to draw a text, presumably because of the way in which a long text can be effectively represented using a small set of words. Simple statistical methods, such as word frequency counts, are readily interpretable. A list of variously sized words is a direct way of communicating with any user, from beginner to expert. Most of the partial-text approaches available online use statistical methods to extract the part from the whole.

It is our contention that extracting part of the corpora can be affected by the structure and complexity of the whole. In the visualizations under review, half present unstructured text corpora, but the criteria used in extracting the part from the whole are well defined and include lists of verbs (*Characteristic verbs*, case 16), words occurring in the text in an “X and Y” pattern (*DocuBurst*, case 18) and lists of words not included in a list of predefined empty words (*Google’s bi-gram data*, case 21).

Clearly, extraction processes based on word or phrase functionality, as opposed to those that use statistical methods, are more closely affected by the nature of the text. Here, we focus on these cases because they are more interesting in terms of our research goals. They include the cases of *Novel views: Les misérables*. *Characteristic verbs* (case 16), which represents only verbs, *DocuBurst* (case 18) which uses the crowd-sourced lexical database *Wordnet* as a human-like backup, and *Phrase net* (case 19) and the two *Google bi-gram visualizations* (cases 20 and 21).

A common pattern detected in the partial-text visualizations reviewed is that once a part of the text has been extracted all except one (*Document arc diagrams*, case 22) discard any reference to the original text sequence in the visualization. See the following point for a more detailed discussion of this idea.

c) Other subcategories

Here we include sequential and non-sequential visualizations, discourse and syntactic structures visualizations, search results and datasets dependent on time visualizations.

Sequential visualizations

Sixteen of the 23 single-text visualizations maintain a similar sequence to that of the original text. Seven of these visualize the sequence using a discourse structure (primarily chapters), while the remaining nine use syntactic elements to represent the original sequence of the text (primarily words).

Strikingly, only one partial-text visualization, Clark's *Document arc diagrams* (case 22) (figure 5), follows the original text sequence, whereas all the whole-text visualizations are sequential. It would thus appear that sequentiality is intrinsic to whole-text visualization. Whole-text visualizations do not literally represent every word of the text, but rather present a graphical metaphor of the whole: a text line. This text line may represent either a discourse structure or a syntactic structure of the text; but, whatever the case, graphically a line or area is used to represent the length of the text.

The sequentiality of the visualization means it can be read both backwards and forwards, as can the text. In the case of a long text, such as a book (nine of the 16 cases), the visualization can serve as a map or guide to the text.

Non-sequential visualizations

Five cases use non-sequential visualizations: three use word clouds (cases 17, 20 and 21), one a net of phrases (case 19) and one visualizes all the verbs in the text (case 16).

Discourse structures in the visualization

Cases: 1, 2, 5, 6, 8, 11, 12 and 13

The eight visualizations that follow the discourse structure of the text are sequential –no cases being found in which

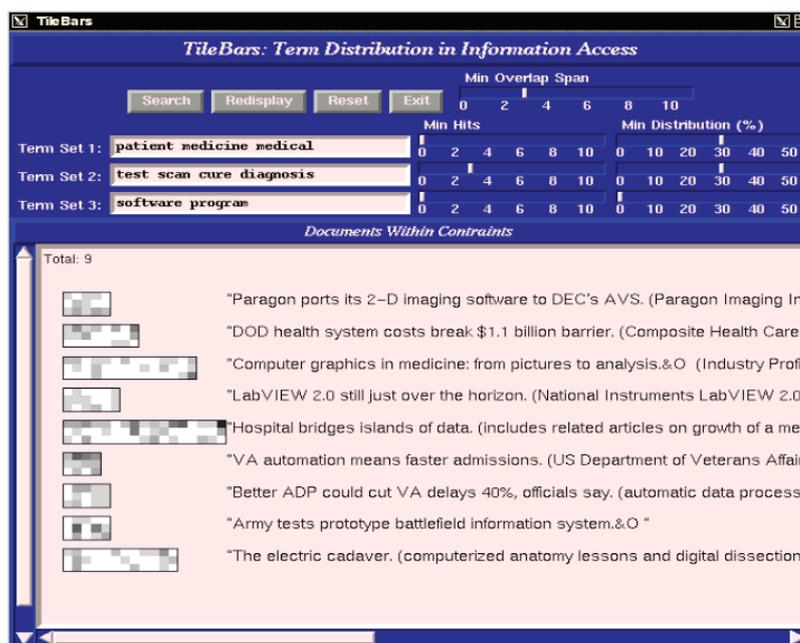


Figure 5. (Case 15) *TileBar* search on (patient medicine medical AND test scan cure diagnosis AND software program) with stricter distribution constraints.

the discourse structure appeared out of sequence with regards to the text. This is perhaps unsurprising, as those cases in which the text is divided into chapters and each chapter represented as a separate entity were considered as text collection visualizations (e.g., *Sentence bar diagrams*, case 4). For this reason, all the cases in this section represent the parts of a text ordered and aligned (in a curve or line). Of the eight visualizations, five represent chapters or sections of a book, two represent complete volumes, while one (*Colour-coded chronological sequencing*, case 11) divides the text in colours according to narrative topics and scenes. Indeed, case 11 is the only one we have identified that uses discourse structure elements that are more deeply embedded than chapters, sections, books and volumes. In all likelihood, more deeply embedded methods than these, such as, narrative topics, would require manual text line segmentation.

Syntactic structures in the visualization

Cases: 3, 16, 4, 7, 18, 9, 22 and 23.

The other eight sequential visualizations use intrinsic text elements, including groups of words (cases 7, 18, 22 and 23), verbs (case 16), sentences (cases 4 and 9) and a complete text analysis (case 3). Syntactic analysis requires either word-by-word parsing of the text (using a database of lexical or semantic word lists) or sentence and paragraph parsing. Syntactic-structure visualization is less dependent on the nature of the text in the sense that the methodology is unaffected by the complexity of the text. Typically, the software automatically extracts or marks the chosen syntactic elements.

Search-result visualizations

Cases: 15, 18 and 23

The three search-result visualizations were presented as web applications and were, therefore, interactive – the user being able to query the visualization system and obtain a

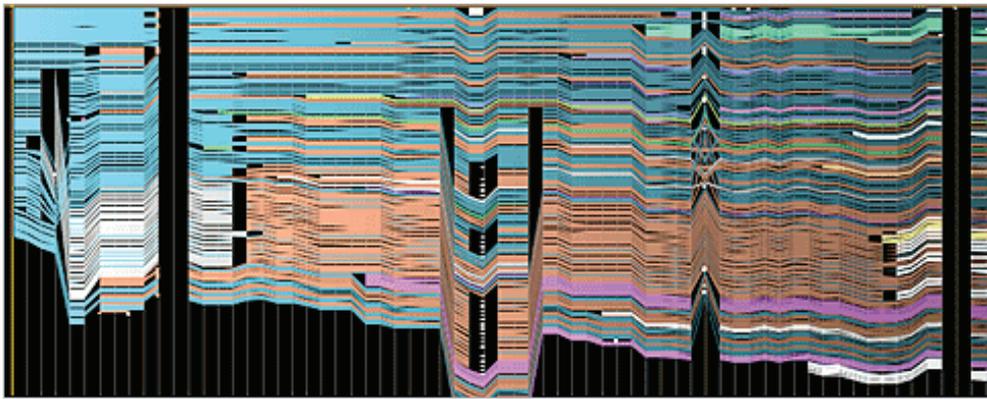


Figure 6. (Case 10) *History flow* by Fernanda Viégas and Martin Wattenberg researchers at IBM's Visual Communication Lab (2003)

over time. A dynamic text visualization demonstrates that data visualization may be the only way to solve certain tasks and that it is not just one more method of pure data advocacy. For example, it is extremely challenging to show how a *Wikipedia* entry evolves over time in line with the editors' participation (*History flow*, case 10) (figure 6). *History flow* provides a solution to

unique representation for each search. The three cases, however, are no longer available online. *DocuBurst* (case 18) is a *Prefuse* application that can be downloaded (Collins et al., 2009). *Prefuse* is a set of software tools for creating rich interactive data visualizations.

TileBars is a classic case of visualization (cited 625 times by *Google Scholar*) designed by a leading expert in visualization and search engine interfaces, Marti Hearst. *DocuBurst* and *Gist icon* are interactive radial visualizations, the latter being one of the references and main influences on the development of *DocuBurst*, as explained in the *DocuBurst* paper cited.

Partial-text visualization is a successful, popular way to draw a text, presumably because of the way in which a long text can be effectively represented using a small set of words

Search-result visualization approaches have not been widely implemented in information retrieval systems and most result outputs are one-dimensional lists of itemized texts (Nualart; Pérez-Montoro, 2013). The three cases reviewed here are each applied to large datasets and, starting with a search query, present an improved search output designed to help the user read and filter the results. All three are particularly concerned with distinguishing between similar items: *TileBars* searches *PubMed* (more than 20 million papers); *DocuBurst* uses the *WordNet* lexical database (155,287 words organized in 117,659 synsets for a total of 206,941 word-sense pairs) to classify the visualized text; and, *Gist icons* use, among others, the complete dataset of approximately 7 million *USpto* patents and the *Enron* email dataset comprising 500,000 emails.

In the text collection category below, we present nine further search-result visualizations.

Time dependent datasets

Cases: 6 and 10.

We present two cases in which the visualization approaches can be used to understand or follow the evolution of a text

this problem and sheds light on the complex collaborative process of *Wikipedia*.

In the second case (*Favoured traces*, case 6), an animated visualization demonstrates how Darwin's ideas evolved through successive editions of the *Origin of Species*. In Ben Fry's words: "The first English edition was approximately 150,000 words and the sixth is a much larger 190,000 words. In the changes are refinements and shifts in ideas —whether increasing the weight of a statement, adding details, or even a change in the idea itself."

2.2.2. Text collections

We present text collections grouped as pure item visualizations, aggregation visualizations and other subcategories. The latter includes data as a landscape layer and search result visualizations. Each subsection adheres to the following structure: list of cases, description of the group and discussion.

a) Item visualizations

- 24) Literature (Note: this converts a single text into a collection). *Novel views: Les misérables. Segment word clouds* by Jeff Clark (2013)
- 25) Literature. *Grimm's fairy tale network* by Jeff Clark (2013)
- 26) Twitter. *Spot* by Jeff Clark (2012)
- 27) Science. *Word storm* by Quim Castella and Charles Sutton (2012)
- 28) Literature. *Topic networks in Proust. Topology* by Elijah Meeks and Jeff Drouin (2011)
- 29) Wikipedia. *Notabilia* by D. Taraborelli, G. L. Ciampaglia and M. Stefaner (2010)
- 30) Media art. *X by Y* by Moritz Stefaner (2009)
- 31) Search engine. *Search clock* by Chris Harrison (2008)
- 32) Online media. *Digg rings* by Chris Harrison (2008)
- 33) Science. *Royal Society Archive* by Chris Harrison (2008)
- 34) Wikipedia. *WikiViz: Visualizing Wikipedia* by Chris Harrison (2007)
- 35) Visualization. *Area* by Jaume Nualart (2007)
- 36) *Chromograms* by M. Wattenberg, F.B. Viégas and K. Hollenbach (2004)

37) Search engines. *KartOO/Ujiko* by Laurent Baleyrier and Nicholas Baleyrier (2001)

38) Search engines. *Touchgraph* by TouchGraph, LLC. (2001)

39) Internet. *HotSauce* by Ramathan V. Guha (1996)

Description

- Number of cases: We identify 16 cases that can be categorized as item visualizations.
- Years: The cases were published over a 17-year period from 1996 to 2013.
- Authors: The most prolific authors in this category are Chris Harrison (cases 13, 32, 33 and 34) and Jeff Clark (cases 24, 25 and 26), followed by Moritz Stefaner with two cases (29 and 30).
- Disciplines and datasets: Interestingly, nine cases are datasets taken from the Internet: *Wikipedia* (cases 29, 34 and 36), search engines (cases 31, 37 and 38), *Twitter* (case 26), online media (case 32), web pages (case 39). Only three cases use literary texts (cases 24, 25 and 28). Finally, two cases visualize scientific papers (cases 27 and 33), one case uses media art datasets (case 30) and one represents non-specific collections (case 35).

Discussion

The main difference between single-text and text-collection visualizations lies in the nature of the text. In the case of the latter, most of the texts do not originate from literature and are accessible online. Yet, the nature of the text appears to be less important when the goal is the representation of the collection rather than of the text itself.

Item visualizations use methods that are independent of the nature of the items themselves. Once the text collections have been itemized, the dataset can be considered a general case of data visualization and not a pure case of text visualization. For this reason, in this category, the methods are generally well known and used in other fields of visualization. Thus, we find six network visualizations (cases 25, 28, 34, 37, 38 and 39), three timelines (cases 31, 32 and 33) and three cases that likewise use timelines but which also permit categorization-based groupings (cases 26, 30 and 35) (figure 7).

Finally, four cases are, we believe, quite specific to text visualization. Two are concerned with item com-

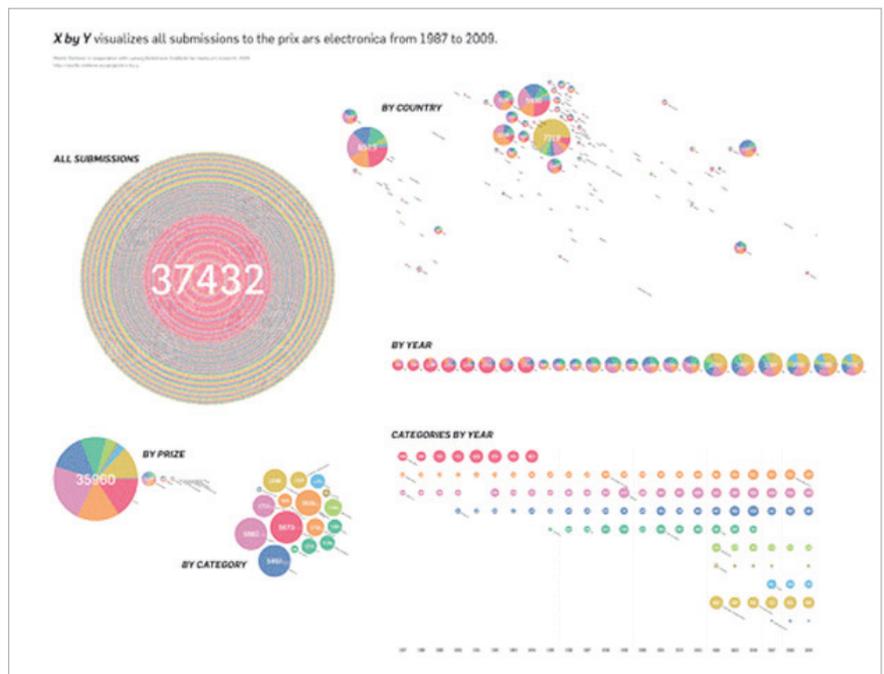


Figure 7. (Case 30) *X by Y* by Moritz Stefaner (2009)

parison: *Segment word clouds* (case 24) and *Word storm* (case 27). *Segment word clouds* transforms a single text into a text collection. Specifically, it is used to represent the chapters of *Les misérables* as word cloud items, thus facilitating their comparison. It also uses colour to identify words as they acquire prominence in the text.

Word storm is a reinvention of word cloud, or more specifically a variation of *Wordle* (case 17) that allows word clouds to be compared. This is achieved by assigning a fixed position to each word. This simple idea makes it visually easy to compare word clouds while maintaining the usual word cloud features.

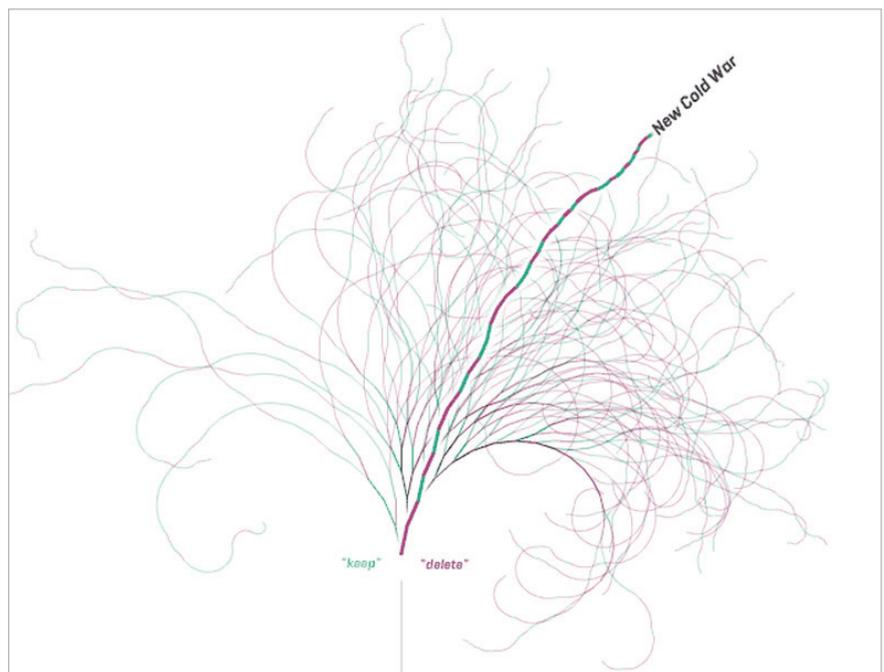


Figure 8. (Case 29) *Notabilia. 100 longest Article for deletion [AfD] discussions on Wikipedia* by Dario Taraborelli, Giovanni-Luca Ciampaglia (data and analysis) and Moritz Stefaner (visualization) (2010)

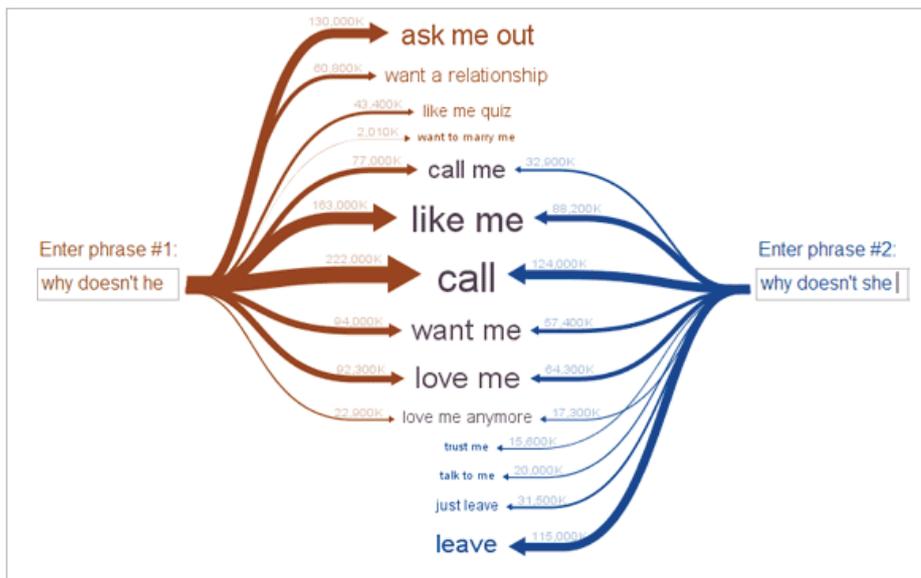


Figure 9. (Case 43) *Web seer* by Fernanda Viégas & Martin Wattenberg (2009)

To conclude, *Notabilia* (case 29) and *Chromograms* (case 36) are two highly original cases that deserve mention. The very specific design of *Notabilia* shows the evolution of “Article for deletion” discussions of Wikipedians (figure 8), discussions that are sometimes more like “flame wars” given the controversies that rage over the simple existence of certain definitions. *Notabilia* visualizes the evolution of the hundred longest discussions and their final outcomes. Moritz Stefaner’s visualization constitutes an interactive bush tree, the branches of which are highlighted when moused over. The shape of the branches informs the reader about the nature of the discussion: cyclical, straight or never-ending.

Chromograms is also based on *Wikipedia* data, providing an analysis of the comments of editors for each edition of a *Wikipedia* entry. Visually it produces colour-coded stripes that in a small space rapidly inform the reader about the edit history of *Wikipedia* entries.

It might prove more effective to apply visualization techniques to texts that have a more formal register and/or predefined outline and a well-defined vocabulary

b) Aggregation visualizations

- 40) Literature. *Grimm’s fairy tale metrics* by Jeff Clark (2013)
- 41) Topic models. *Termite* by J. Chuang, C.D. Manning and J. Heer (2012)
- 42) *Wikipedia*. *Pediameter* by Müller-Birn, Benedix and Hantke (2011)
- 43) *Google* suggestions. *Web Seer* by Fernanda Viégas & Martin Wattenberg (2009)
- 44) *Google* n-grams. *Web trigrams: visualizing Google’s trigram data* by Chris Harrison (2008)

45) Political speech. *FeatureLens* by A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman and C. Plaisant (2007)

46) Online news. *Newsmap* by Marcos Weskamp (2004)

47) Email conversation. *Themail* by Fernanda B. Viégas, Scott Golder, Judith Donath (2006)

48) Search engine. *WebBook* by S.K. Card, G.G. Robertson and W. York (1996)

49) Any texts. *Dotplot applications* by Jonathan Helfman (1994)

Description

- Number of cases: We identify 10 cases that can be categorized as aggregation visualizations.
- Years: The cases were published over a 19-year period from 1994 to 2013.
- Authors and datasets: Only Fernanda B. Viégas participated in more than one of the 10 cases in this category (cases 43 and 47); the rest participated in just one case each. The texts are very similar in nature to those in the item visualization category. Five cases are corpora that can be found online (*Wikipedia*, case 42; *Google*, cases 43 (figure 9) and 44; online news, case 46; search engine results, case 48). The standard unstructured texts include one from literature (*Sentence Bar Diagrams*, case 4), one from political speeches (*FeatureLens*, case 45) and one from a year’s worth of email conversations between two correspondents (*Themail*, case 47). Finally, there are two quite unique cases: *Termite* (case 41) and *Dotplot* (case 49). All the cases are discussed below.

Discussion

Aggregation visualizations is the category with the greatest variation in the methods employed. Thus, apart from visualizing text collections, the only thing the 10 cases assigned to this category have in common is that they do not represent specific items.

Given these circumstances, we comment on each case separately:

Sentence bar diagrams (case 40) provide a matrix (or table-like) visualization that allows rows to be sorted by clicking on columns. The columns provide a quantitative definition of 13 metrics related to the 62 stories making up Grimm’s fairy tales. It is a powerful tool for analysing, understanding and comparing the tales.

Termite (case 41) is a case that represents an intermediary dataset known as topic models. Topic models are a “cleverer” way of obtaining a bag-of-words from a text than applying a typical word-frequency statistical analysis. *Termite* does not visualize texts but it does compare parts of

texts. As such, the tool can be used to compare topic models.

Pediameter (case 42) is a specific interface that uses bar charts to show *Wikipedia* editions in real time. It is most remarkable for using a device known as an *Arduino* to detect editions and transcribe them to a physical indicator, merging digital and material worlds.

Web Seer (case 43) is another specific visualization method that shows the most popular search queries based on *Google* suggestions. The approach allows queries to be compared by representing the suggestions with trees and then connecting the matching branches. The simplicity of this case contrasts with its power of communication: rapid and user friendly.

Google's tri-gram data (case 44) uses a similar visualization method to that used by *Web seer*. It draws on the huge *Google* n-gram dataset and represents and compares three-word sentences (tri-grams).

FeatureLens (case 45) is an interactive, dashboard-style interface for comparing texts. The central representation uses a visualization of frequent concepts similar to that used by *Texty* (case 7) and *TileBars* (case 15). It allows text browsing and shows line graphs of frequent words found throughout a text.

Newsmap (case 46) uses treemap visualization to offer a new method for reading and monitoring the news in real-time, employing online *Google* news feeds. It is totally customizable in terms of topic, country and publication time. The software, which is available free of charge online, can also be used for news searches.

TheMail (case 47) is an experiment in which a highly specific interface was developed to follow and analyse the evolution of an email correspondence between two people over the course of one year. It visualizes the words that characterize each of the writers and their evolution over time.

When first developed in 1996, *WebBook* (case 48) (figure 10) was a somewhat surprising application, as it transformed search engine results in a multimedia (text and images, primarily) mash-up based on the metaphor of the book. The application was a pure text (web pages) collection visualization that presented the results as aggregations of text and images.

Finally, *Dotplot* (case 49) was an innovative visualization application with multiple uses, not unlike *Arc diagrams* (case 14). The main use of *Dotplots* is for text comparisons, including multi-language, text version and programming code comparisons.

c) Other subcategories

Here we include landscape data layers, search-result visualizations and time-dependent datasets.

Landscape as an additional data layer

Cases: 40, 26, 28, 33, 47, 37, 38 and 49.

The typical concept of landscape data is a network visualization comprising two layers of data, as in *Topic networks* (case 28). In this specific case, the first layer is provided by



Figure 10. (Case 48) *WebBook* by Stuart K. Card, George G. Robertson, and William York (1996)

the Marcel Proust texts represented as items and the second layer by a network of topic models of these texts. The positions of the nodes of both layers are optimised so that proximity indicates more strongly related nodes. This definition of landscape can also be found in the defunct search engine results provided by *KartOO/Ujiko* (case 37) and *TouchGraph* (case 38).

All the other cases included in this category present text collections in combination with more data. This is the case of *Dotplot*, which represents the coincidence or otherwise of strings in various texts, and of *Grimm's fairy tale metrics*, which combines a list of texts in rows with various parameters listed in columns. These parameters do not form a direct part of the text, but rather they are recalculated features related to the text, including, for example, length, lexical diversity and the presence of different groups of words that represent entities (for example: body -> hand, head, heart, eyes and foot) in each tale.

“Landscape visualizations allow to compare a collection of texts simultaneously with a second parameter”

A third kind of landscape is based on the representation of timed metadata, as exemplified by *Spot* (case 26), the *Royal Society Archive* (case 33) and *TheMail* (case 47).

A common feature of landscape visualizations is their capacity to compare a collection of texts simultaneously with a second parameter, while their main limitation is the number of items represented so that large numbers create problems of overlapping items.

Search result visualizations

Cases: 26, 43, 35, 45, 47, 46, 37, 38 and 48.

Compared to single-text visualizations, text-collection visualizations include considerably more cases offering search capacities (three vs. nine). Common sense suggests that when presenting a text collection, a natural feature of such an approach will be a way of selecting part of that collection based on given criteria, i.e., filter and search features.

All the cases included in this category allow search queries and output a unique visualization for each query. All the cases include a search box and a search button.

Time-dependent datasets

Cases: 42, 29, 36 and 46.

The four cases included in this category allow the user to monitor the evolution of the texts in the collection over time. Only one is designed for use in real-time (*Newsmap*, case 46), but potentially all of them can visualize the collection on a specific date and at a specific time.

One obstacle faced by an approach that represents changes in text collections over time is providing access to an updated feed or an accessible API. It is presumably for this reason that three of the four use *Wikipedia* data and the other uses *Google* news. In all cases, they are online sources that have long allowed public access to their feeds.

3. Conclusions

The diversity of approaches developed in different disciplines, the wide diffusion of publications or, on occasions, the absence of formal publications of innovative ideas, represent a considerable challenge to the undertaking of a comprehensive survey of the work completed in this field. Thus, some of the visualizations we present here have been unearthed in highly specific publications, the case for example of Joel Deshayé and Peter Stoicheff and their work on representing Faulkner (cases 11, 12 and 13). If we read Stoicheff's working notes it is apparent that their visualizations were developed to facilitate the study of William Faulkner's narrative timelines. There are no additional references to the application of these interesting ideas to other texts, suggesting that more works remain hidden in the depths of other fields.

Text visualization, as we have argued throughout this review, may be considered a subfield of data visualization. Yet, the boundaries of the discipline are not always clearly defined. This is readily illustrated, for example, by the case of Harrison's *Search clock* (case 31), in which the text corpora comprise an enormous dataset of search engine queries. Can this dataset really be considered a collection of texts when each of them, in most instances, is no more than one or two words in length? Does a text have to satisfy a minimum length in order to be considered a text? Here, we opted to treat case 31 as a collection of texts, short ones admittedly but, ultimately, *texts*.

Clearly, the critical decision to be made throughout this review has been how to classify the cases identified. As few papers have attempted to review only text visualization approaches, we turned to classic data visualization reviews (e.g., Shneiderman, 1996) as well as to more recent ones (e.g., Collins *et al.*, 2009). In all these instances, the classifications were based on tasks that the visualization approach can solve rather than on the explicit aspects of the visualization themselves. For this reason we chose to propose our own classification, which, while far from perfect, we hope will be useful for undertaking a classification based on visual features.

We conclude with a list of insights, as well as shortcomings, that we have identified to date:

- Single-text visualizations have been applied mainly to literature, a field that, apart from being characterized by complex combinations of words, can present high levels of human abstraction and freedom of structure and experimentation. As such it might prove more effective to apply visualization techniques to texts that have a more formal register and/or predefined outline and a well-defined vocabulary, such as legal texts, scientific papers, template-based texts and communications, etc.
- We have identified only one single/partial-text visualization that is sequential (*Document arc diagrams*, case 22). Most partial-text visualizations extract the essence of the text based on one or more criteria and so the original sequence of the text is lost. Since sequential visualization approaches present certain advantages, it seems that partial-visualization approaches that maintain the original text sequence should be encouraged.
- Text-collection visualizations tend to employ methods that are used for data visualization in general. Hence, there is a need for further experimentation in applying more standard data visualization methods and approaches to the specific subfield of text visualization.
- Text collection aggregations is the category in which the most specific designs and ideas have been developed. More work needs to be undertaken to identify any common approaches in this kind of visualization.

And, finally, we pose the following question:

- Why is it that most of the cases reviewed here that are more than five years old are no longer available online? If the software used is no longer (or was never) in use, we should perhaps question its effectiveness. While we have not investigated just how many cases form part of commercial software products and how many, following publication, have simply been forgotten, the question remains as to why some apparently magnificent ideas did not establish themselves as new standards. Our challenge to researchers is to produce applications that will be adopted in one field or another, or which can solve a problem for a certain group of users; indeed, as the cases reviewed here highlight, adoption seems to represent a considerable challenge.

Acknowledgement

This work is part of the project "Active audiences and journalism. Interactivity, web integration and findability of journalistic information". CSO2012-39518-C04-02. *National plan for R+D+i, Spanish Ministry of Economy and Competitiveness*.

4. References

- Anglin, Gary J.; Vaez, Hossein; Cunningham, Kathryn L. (2004). "Visual representations and learning: The role of static and animated graphics". *Handbook of research on educational communications and technology*, 2, pp. 865-916.
- Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier *et al.* (1999). *Modern information retrieval*. New York: ACM press, vol. 463.

- Baeza-Yates, Ricardo; Broder, Andreiz; Maarek, Yoelle** (2011). "The new frontier of web search technology: Seven challenges". *Search computing*, v. 6585 of *Lecture notes in computer science*, pp. 3-9.
http://dx.doi.org/10.1007/978-3-642-19668-3_1
- Benavides, David; Segura, Sergio; Ruiz-Cortés, Antonio** (2010). "Automated analysis of feature models 20 years later: A literature review". *Information systems*, v. 35, n. 6, pp. 615-636.
<http://dx.doi.org/10.1016/j.is.2010.01.001>
- Collins, Christopher; Carpendale, Sheelagh; Penn, Gerald** (2009). "DocuBurst: Visualizing document content using language Structure". *Computer graphics forum* (Procs. of the Eurographics/IEEE-VGTC Symposium on visualization, EuroVis), v. 28, n. 3, pp. 1039-1046.
<http://dx.doi.org/10.1111/j.1467-8659.2009.01439.x>
- Feldman, Ronen; Sanger, James** (2006). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press. ISBN: 13 978 0 521 83657 9
- Grobelnik, Marko; Mladenić, Dunja** (2002). "Efficient visualization of large text corpora". In: *Procs of the 7th seminar*. Dubrovnik, Croatia.
<http://ailab.ijs.si/dunja/SiKDD2002/papers/GrobelnikSep02.pdf>
- Hearst, Marti A.** (2003). *What is text mining?*
<http://people.ischool.berkeley.edu/~hearst/text-mining.html>
- Hearst, Marti A.** (2009). "Search user interfaces", Chapter 1. ISBN: 9780521113793
<http://searchuserinterfaces.com/book>
http://searchuserinterfaces.com/book/sui_ch1_design.html
- Hearst, Marti A.** (2011). "Natural search user interfaces". *Communications of the ACM*, v., 54, n. 11, November, pp. 60-67.
<http://cacm.acm.org/magazines/2011/11/138216-natural-search-user-interfaces/fulltext>
<http://dx.doi.org/10.1145/2018396.2018414>
- Heer, Jeff** (2010). "A conversation with Jeff Heer, Martin Wattenberg, and Fernanda Viégas". *Queue*, v. 8, n. 3, 10 pp., March.
<http://doi.acm.org/10.1145/1737923.1744741>
- Iliinsky, Noah** (2013). *Choosing visual properties for successful visualizations*. IBM Software. Business Analytics.
<http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03323usen/YTW03323USEN.PDF>
- Kitchenham, Barbara** (2004). *Procedures for performing systematic reviews*. Keele, UK, Keele University, 33 pp.
- Levie, W. Howard; Lentz, Richard** (1982). "Effects of text illustrations: A review of research". *ECTJ*, v. 30, n. 4, pp. 195-232.
- Mann, Thomas M.** (2002). *Visualization of search results from the world wide web*.
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.91.2535>
- Meeks, Elijah** (2011). *Digital humanities specialist*. Documents.
<https://dhs.stanford.edu/comprehending-the-digital-humanities/documents>
- Nualart-Vilaplana, Jaume** (2013). *How we draw texts: a visualization of text visualization tools*.
<http://research.nualart.cat/textvistools>
- Nualart, Jaume; Pérez-Montoro, Mario** (2013). "Texty, a visualization tool to aid selection of texts from search outputs". *Information research*, v. 18, n. 2, June.
<http://www.informationr.net/ir/18-2/paper581.html>
- Shneiderman, Ben** (1996). "The eyes have it: A task by data type taxonomy for information visualizations". In: *Visual Languages*. Proceedings IEEE Symposium, pp. 336-343.
<http://dx.doi.org/10.1109/VL.1996.545307>
- Šilić, Artur; Dalbelo-Bašić, Bojana** (2010). "Visualization of text streams: A survey". *Knowledge-based and intelligent information and engineering systems*, v. 6277 of *Lecture notes in computer science*, pp. 31-43. Berlin, Heidelberg: Springer.
http://dx.doi.org/10.1007/978-3-642-15390-7_4
- Stefaner, Moritz** (2013). *Gender balance visualization*.
<http://moritz.stefaner.eu/projects/gender-balance/#NUM/NUM>
- Strecker, Jacqueline** (2012). *Data visualization in review: summary*. International Development Research Centre (IDRC), Ottawa, ON, Canada.
<http://idl-bnc.idrc.ca/dspace/bitstream/10625/49286/1/IDL-49286.pdf>
- Times Higher Education. *World university rankings 2012-2013*.
<http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking>
- Tufte, Edward R.; Graves-Morris, P. R.** (1983). *The visual display of quantitative information*, v. 2. Cheshire, CT: Graphics Press, 199 pp.



Diego-Alonso Gómez-Aguilar, Francisco-José García-Peñalvo y Roberto Therón



Diego-Alonso Gómez-Aguilar es doctorando en la *Universidad de Salamanca*. Su investigación se centra en la combinación de la informática, la inteligencia artificial, el diseño gráfico, la analítica visual, *e-learning*, los sistemas adaptativos, la ingeniería web y web semántica. Actualmente forma parte del *Grupo de Investigación de Analítica Visual y Visualización de la Información (VisUsal)* y del grupo de investigación *Grial (Grupo de Investigación en Interacción y e-learning)*. Trabaja en el análisis visual de la información en *Moodle*.

<http://orcid.org/0000-0003-1915-4758>

dialgoag@usal.es



Francisco-José García-Peñalvo es doctor por la *Universidad de Salamanca* y director del grupo de investigación *Grial (Grupo de investigación en Interacción y e-learning)*. Sus principales intereses de investigación se centran en el *e-learning*, computadores y educación, sistemas adaptativos, ingeniería web, web semántica y reutilización de software. Ha dirigido y participado en más de 15 proyectos de innovación e investigación. Fue Vicerrector de Innovación Tecnológica de la *Universidad de Salamanca*, ha publicado más de 100 artículos en revistas y conferencias internacionales y revisor de varias revistas internacionales.

<http://orcid.org/0000-0001-9987-5584>

fgarcia@usal.es



Roberto Therón es doctor por la *Universidad de Salamanca*. Su investigación se centra en la combinación de campos como la informática, la inteligencia artificial, la estadística, el diseño gráfico y la visualización de datos como medio para la comprensión de datos complejos. Es director del *Grupo de Investigación de Analítica Visual y Visualización de la Información (VisUsal)*, dedicado al desarrollo de herramientas avanzadas que ayudan a los usuarios a comprender conjuntos de datos complejos provenientes de una variedad de campos. Ha participado y/o gestionado numerosos proyectos de excelencia en los ámbitos regional, nacional y europeo. Es autor de más de 100 artículos publicados en revistas y conferencias internacionales.

<http://orcid.org/0000-0001-6739-8875>

theron@usal.es

*Universidad de Salamanca. Facultad de Ciencias
Departamento de Informática y Automática
Plaza de los Caídos, s/n. 37008 Salamanca, España*

Resumen

Las tecnologías utilizadas en los procesos de aprendizaje implican el registro de todas las actividades realizadas. Estos datos se pueden aprovechar para la evaluación de estudiantes, profesores y de los propios procesos. Sin embargo, aunque existe esta gran cantidad de datos, sigue siendo difícil para los profesores (y otras partes interesadas) verificar hipótesis, extraer conclusiones o tomar decisiones basadas en hechos o situaciones detectadas. Se presenta un modelo de análisis de datos educativos basado en analítica visual, analítica del aprendizaje y analítica académica. Por medio de una herramienta de software permite realizar análisis de datos exploratorios y confirmatorios, en interacción con la información obtenida de un sistema típico de gestión de aprendizaje. El objetivo principal es el descubrimiento de nuevo conocimiento sobre el proceso de aprendizaje educativo que, a su vez, posibilite la mejora de éste.

Palabras clave

Minería de datos educativa, *E-learning*, Realimentación, Analítica del aprendizaje, Análisis de redes sociales, visualización, Analítica visual.

Title: Visual analytics in e-learning

Abstract

Current technologies used in *e-learning* processes require the logging of all activities performed. These data can be used in the assessment of students, instructors and the processes themselves. However, although this wealth of data exists, it is still difficult for those who teach (and interested stakeholders) to verify a hypothesis, extract conclusions, or make decisions based on the facts or situations that are discovered. This paper introduces an educational data analysis model based on visual analytics, learning analytics and academic analytics and using a software tool that allows performing confirmatory and exploratory data analysis through interaction with the information gathered from a typical Learning Management System. The main goal is to enable the discovery of knowledge on the specific learning process that, in turn, will permit its improvement.

Keywords

Educational data mining, Feedback, E-learning, Learning analytics, Inferring social network, Visualization, Visual analytics.

Gómez-Aguilar, Diego-Alonso; García-Peñalvo, Francisco-José; Therón, Roberto (2014). "Análítica visual en e-learning". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 236-245.

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

La analítica en la educación

La informática y la tecnología ubicuas han provocado una evolución en la educación hacia nuevos paradigmas de enseñanza, concediendo a los usuarios acceso a aulas virtuales y a bases de datos de conocimiento desde cualquier lugar y en cualquier momento. Como consecuencia, los procesos de aprendizaje están en continua evolución, apareciendo nuevos enfoques en la universidad hacia la docencia basada en medios electrónicos y sociales (Berlanga; Peñalvo; Sloep, 2010).

El análisis de datos educativos no es nuevo; el término "analítica académica" (en inglés, *academic analytics*, AA) lo acuñaron Goldstein y Katz (2005) definiéndolo como la aplicación de herramientas de inteligencia de negocios para el área de aprendizaje. La principal aplicación de la AA es la de ir más allá de un reporte simple de información y sugerir decisiones. A diferencia de la AA, donde se utilizan los datos capturados para tomar decisiones a nivel de institución, el objetivo de la "analítica del aprendizaje" (en inglés, *learning analytics*, LA), es el uso de estos datos y de cualquier otra observación adicional de la que se pueda obtener información, para tener un impacto directo sobre los estudiantes, los profesores y el proceso de aprendizaje (Long; Siemens, 2011).

Las herramientas de analítica visual hacen posible obtener un modelo mental de datos complejos y nuevo conocimiento

Siguiendo con el objetivo de mejora del aprendizaje, la minería de datos educativos (en inglés, *educational data mining*, EDM) es un campo que persigue un objetivo similar al de la LA. La EDM propone el uso de técnicas como el análisis estadístico, aprendizaje automático (del inglés, *machine learning*), minería de datos, etc., para resolver los problemas de investigación educativa y entender el entorno en el que los estudiantes aprenden (Baker; Yacef, 2009; Romero; Ventura, 2010). Entre estas técnicas se encuentran las regresiones (Agudo-Peregrina et al., 2014) como método

utilizado para este propósito. En la misma línea, la socialización y colaboración tienen un papel clave en los procesos de aprendizaje (Brown; Adler, 2008). Junto a esto, existe una fuerte evidencia de que en las actividades bien estructuradas, los procesos de construcción del conocimiento alcanzan niveles superiores de pensamiento crítico, donde los estudiantes son capaces de establecer y sostener grupos cohesivos. Esto justifica la necesidad del "análisis de redes sociales" (del inglés, *social network analysis*, SNA) y de herramientas que sean capaces de ofrecer en tiempo real análisis para los tutores y profesores (Bakharia; Dawson, 2011).

De acuerdo con Barry y Fulmer (2004), el grado en que otras personas pueden recuperar y leer mensajes de un individuo puede afectar su uso. En el "sistema de gestión de aprendizaje" (en inglés, *learning management system*, LMS), Silva y Figueira (2012) describen un sistema listo para utilizarse acoplado con un análisis de contenido para rechazar mensajes irrelevantes y recopilar métricas de redes sociales. Las métricas ayudan a obtener una mejor medición de las características de la red. Existen diversas aplicaciones que tratan el análisis de redes sociales desde diferentes técnicas visuales, y no sólo en el área educativa, (González-Teruel; Andreu-Ramos, 2013; Van-Ham; Schulz; Dimicco, 2009; Viégas; Donath, 2004; Von-Landesberger; Görner; Schreck, 2009).

Por otra parte la "analítica visual" (en inglés, *visual analytics*, VA) es un área emergente de la investigación y su práctica tiene como objetivo apoyar el razonamiento analítico a través de interfaces visuales interactivas (Thomas; Cook, 2006). En ésta, la visualización se integra como núcleo y, en conjunto con otras disciplinas científicas, mejora la división del trabajo entre el hombre y la máquina, combina el razonamiento analítico con visualización interactiva, es decir, representaciones visuales y técnicas de interacción (que explotan las capacidades perceptivas del ancho de banda del ojo humano para permitir a los usuarios ver, explorar y entender grandes cantidades de información de forma simultánea), así como técnicas de apoyo a la producción, presentación y difusión de resultados analíticos que comuniquen información en el contexto adecuado para una variedad de audiencias (Keim; Zhang, 2011).

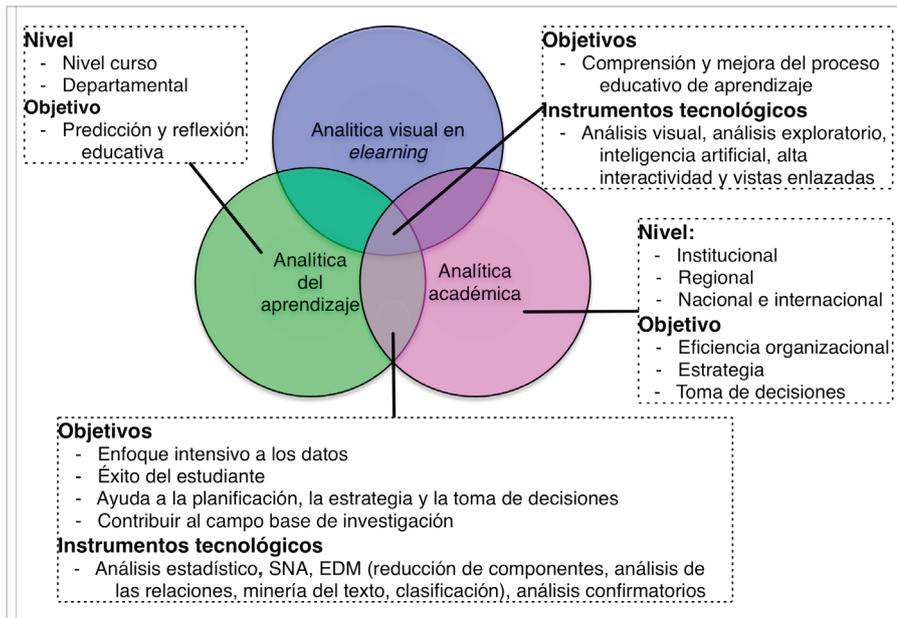


Figura 1. Modelo teórico de la visualización analítica en e-learning

Las herramientas de VA hacen posible obtener un modelo mental de los datos complejos y nuevo conocimiento (Keim et al., 2008; Thomas; Cook, 2005). El objetivo es ayudar a los usuarios a descubrir las anomalías inesperadas y sorprendentes, cambios de patrones y relaciones que luego son evaluadas para desarrollar nuevas ideas (Cook; Earnshaw; Stasko, 2007). En los últimos años, han sido aplicadas algunas de las técnicas basadas en el análisis visual: Gibbs, Olexa y Bernas (2006) presentan un software que representa el esquema temporal de las relaciones en los debates; De-Laat utiliza métodos del análisis de redes sociales para estudiar la naturaleza de los patrones de interacción dentro de una comunidad de aprendizaje en red (De-Laat, 2007; Haythornthwaite; De-Laat, 2010). Por último y más cercano al presente trabajo, Govaerts et al. (2012) presentan en paralelo un histograma y un diagrama de barras que evidencian las diferencias entre los datos.

En el presente trabajo de investigación se pretende mejorar la eficiencia del proceso de análisis de aprendizaje. El objetivo es proponer y validar un modelo que tome en cuenta técnicas de analítica visual y la información existente en un LMS, con el fin de mejorar el proceso de LA y AA, sin olvidar asegurar el éxito y desarrollo de la asignatura. Para ello, a manera de comprobación del modelo, se ha implantado un sistema y aplicado sobre un LMS para validar su expresividad.

Tabla 1. Abreviaturas utilizadas en este trabajo

AA	Analítica académica	Academic analytics
LA	Analítica del aprendizaje	Learning analytics
VA	Analítica visual	Visual analytics
EDM	Minería de datos educativos	Educational data mining
LMS	Sistema de gestión de aprendizaje	Learning management system
VeLA	Visualización analítica en e-learning	Visual e-learning analytics

1. Definición de un modelo de analítica visual en e-learning

Al igual que con la mayoría de las áreas de investigación, la terminología que identifica y diferencia EDM, AA y LA no es tan homogénea como sería deseable (Van-Barneveld; Arnold; Campbell, 2012), sin embargo, las tres áreas tienen en común, además del objetivo de mejorar y comprender el proceso de aprendizaje, la necesidad de una gran cantidad de observaciones de dicho proceso.

Por otro lado, Siemens et al. plantean que las comunidades de LA y EDM tienen un solapamiento considerable (tanto en términos de investigación como en sus investigadores), y que ambas

comunidades creen firmemente en la realización de investigaciones y aplicaciones que beneficien a los estudiantes, así como la mejora de las ciencias del aprendizaje; además plantean algunas diferencias entre estas comunidades (Siemens; Baker, 2012); Siemens et al., 2011).

Las técnicas, métodos y objetivos que diferencian e identifican la LA de la EDM se encuentran mencionados y aplicados en diversas áreas de la ciencia dentro del alcance de la VA, por tanto realizar una aplicación de VA en el área de la educación, más que ser excluyente a la AA y a LA, las complementa. La notable intersección de estas áreas y sus diferencias da lugar al presente modelo de “visualización analítica en e-learning” (en inglés, *visual e-learning analytics*, VeLA). El presente trabajo considera la EDM y SNA como parte de los métodos, técnicas e instrumentos tecnológicos que se utilizan en VeLA. En la figura 1 se muestra como las áreas de LA y AA confluyen en un conjunto de objetivos e instrumentos tecnológicos y cómo el área de VeLA complementaría y enriquecería el proceso de AA y LA (Chen et al., 2006; Keim; Mansmann; Thomas, 2010; Keim; Zhang, 2011; Von-Landesberger; Görner; Schreck, 2009), lo que se plantea como contribución del presente trabajo.

El modelo de la figura 1 (VeLA) se basa en el modelo de referencia de LA de Chatti et al. (2012), la teoría del ciclo de LA de Clow (2012) y el proceso de la exploración visual de datos de Keim (Keim et al., 2010; Keim; Zhang, 2011), y además toma en cuenta las restricciones y elementos del modelo de dimensiones críticas de la LA de Greller y Drachsler (2012). Por tanto, la aportación teórica principal de la utilización de la VA en el área de la educación es que el proceso de VeLA combina métodos de analíticas automáticas así como visuales exploratorias, que con una alta interacción disponible para el usuario adoptan la finalidad de obtener nuevo conocimiento de los datos educativos.

Ben Shneiderman (1996) propuso un paradigma de la búsqueda de la información que ha sido extendido por Daniel

Keim *et al.* (2008) para dar más capacidad de análisis en el bucle de antes y después de la generación de representaciones visuales interactivas. Con éstas, el usuario explora los datos para ayudarse a extraer modelos abstractos a partir de conjuntos de datos que son demasiado grandes o demasiado complejos para ser analizados de una manera directa, de tal forma que las preguntas del usuario se respondan y se creen durante el ciclo de exploración. Asimismo el paradigma de Keim, aplicado a la analítica educativa, puede extenderse a un paso final, la intervención, proporcionando al ciclo analítico una retroalimentación, quedando de la siguiente forma:

“Analyze first; Show the important; Zoom, filter and analyze further; Details on demand, Intervention”

Este ciclo está representado en la figura 2 en la parte central del proceso. Se muestra una visión general abstracta de las diferentes etapas (representadas a través de círculos) y sus transiciones (flechas) en el proceso de VeLA. Dentro de cada etapa se encuentran representadas las dos áreas de la analítica en la educación, la AA y LA, y su respectiva descripción en cada etapa. Como se aprecia en el diagrama, el modelo contempla distintos perfiles de usuario, la diferencia entre éstos es el objetivo y, por tanto, el tipo de información de su interés.

Dado que el tipo de dato que se maneja para el análisis tiene una temporalidad variable, una socialización y una alta cantidad de actividades sin restricción de tiempo o espacio, este trabajo, con el fin de mejorar el reconocimiento de patrones, aborda específicamente las siguientes necesidades y desafíos en el proceso de VeLA:

Temporalidad

- Analizar el contenido del curso de *e-learning* a través del tiempo y no sólo un curso, sino también un campus o toda una universidad.
- Definir una representación temporal compacta, a fin de encontrar patrones (en meses, horas o semanas), a través de interacción.
- Personalizar el análisis por agrupación de actividades, tiempo, persona específica y perfil.

Análisis del contenido semántico

- Determinar un diseño nube de etiquetas que represente su evolución en el tiempo, en múltiples niveles de detalle, equilibrando así la coherencia semántica, de contenido y la estabilidad espacial de la visualización.
- Plantear un diseño compacto para la legibilidad del *tag cloud* de forma que la importancia o la frecuencia y la evolución en el tiempo de una etiqueta se codifiquen directamente en espacio del tamaño de la palabra.

Análisis de redes sociales

- Realizar la comparación de la estructura jerárquica de la plataforma y las relaciones que se crean entre las personas (foros y contribuyentes compartidos entre foros).
- Definir una herramienta gráfica social que brinde la posibilidad de encontrar la relación directamente proporcional del desempeño de los estudiantes con la frecuencia de la lectura en los foros y/o recursos.
- Ofrecer la posibilidad de ocultar elementos de la red social a fin de ver más claramente la estructura y la relación de los elementos que interesen.

Métricas estadísticas

- Definir un sistema que permita múltiple selección de subconjuntos y/o el uso de diferentes categorías personalizadas de actividades, posibilitando diferenciarlas por su coloración con el objetivo de la identificación de patrones. Además de permitir generar un archivo con ellas posibilitando la importación y exportación.
- Soportar métodos de interacción, como el filtro de búsqueda, ocultación y reordenación de las diferentes métricas.

Con el fin de comprobar la utilidad y expresividad del modelo y su aplicabilidad se ha implementado un sistema de software de VeLA, como resultado de dos ciclos de investigación-acción, en febrero de 2013 y 2014, basados en los trabajos mencionados y en el modelo anteriormente descrito.

Estos ciclos tuvieron como realimentación el cuestionario realizado a los estudiantes del Máster Universitario en TICs en Educación de la *Universidad de Salamanca* (ver resultado de las encuestas en figura 3). Este cuestionario consta de 30 preguntas sobre usabilidad, practicidad y eficiencia del sistema, expuesto a los 35 participantes.

2. Sistema de VeLA

La propuesta está dedicada a la solución del problema de análisis de esa gran cantidad de información. El enfoque

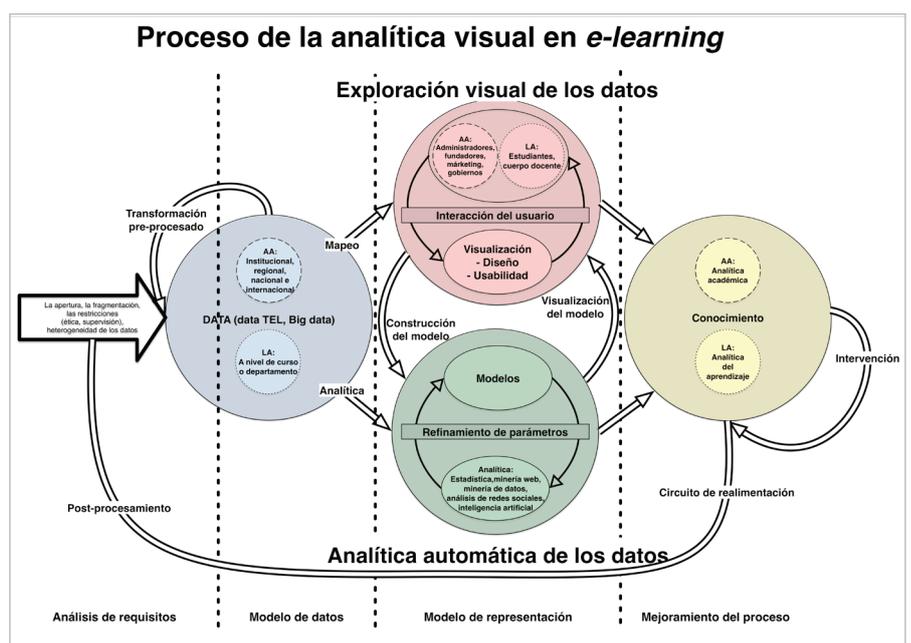


Figura 2. Proceso de analítica visual en e-learning

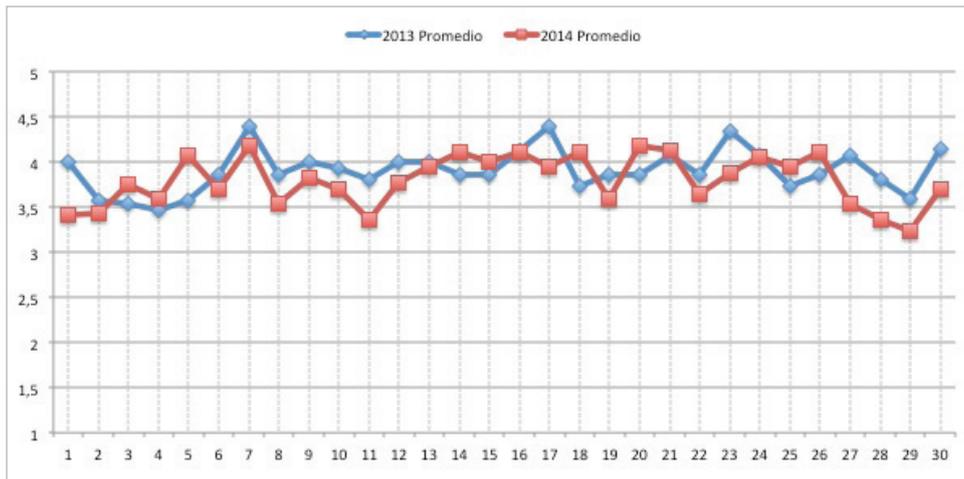


Figura 3. Promedio de las encuestas realizadas de los ciclos de investigación. Los valores van de 1 al 5, siendo 1 una respuesta “muy mala” y 5 “excelente”.

proporciona nuevos conocimientos útiles a través de una herramienta de *software* eficiente que soporta el preprocesamiento de la información, análisis, visualización e interacción. Además, se utiliza de forma independiente de la versión del LMS, y puede ser utilizada offline, online, en web o como aplicación de escritorio. Está dirigida tanto a estudiantes, como profesores y académicos de una institución educativa. Los datos a los que el usuario puede acceder dependen de los permisos que tenga en el LMS, y éstos pueden variar desde un curso hasta un campus universitario completo. En este caso, ha sido adaptado para trabajar con Moodle, tiene capacidad de exploración y comparación usando cuatro técnicas de visualización, basadas en líneas de tiempo (Gómez-Aguilar; Therón; García-Peñalvo, 2009), nubes de etiquetas (Gómez-Aguilar et al., 2011) (Gómez-Aguilar; García-Peñalvo; Therón, 2013a), coordenadas paralelas (Gómez-Aguilar; García-Peñalvo; Therón, 2013b) y gráficos de redes sociales (Gómez-Aguilar; Therón; García-Peñalvo, 2013), y se describe a continuación:

Línea de tiempo en espiral

Es una representación visual interactiva que se utiliza para analizar el uso de un LMS a través del tiempo. Su objetivo es proporcionar una representación compacta de la actividad global del LMS y su aspecto global se puede observar en la figura 4. Este punto de vista se puede adaptar a las necesidades del usuario, por lo que se pueden explorar todos los datos temporales disponibles o pasar de la vista general al detalle (curso, persona, tiempo o actividad) de una persona o actividad en un período de tiempo determinados.

La visualización consta de tres paneles visuales que proporcionan diferentes vistas de datos. La representación principal es la línea de tiempo en espiral, en la parte

central, que en su forma más simple, no es más que una secuencia de eventos codificados por colores. Estos están ordenados siguiendo las agujas del reloj, colocando los datos más antiguos en el centro de la espiral y los eventos más recientes más alejados.

Cuenta con una técnica visual para equilibrar detalle y contexto en la visualización de datos, que se conoce como función de zoom semántico (Modjeska, 1997) o interfaces de múltiples escalas además del tradicional zoom gráfico o físico. La distribución de interfaz gráfica, más allá de la zona central donde se encuentra la espiral, se presenta de la siguiente manera: los otros dos paneles proporcionan vistas adicionales de los datos mostrados en la espiral, en la vista superior se dibuja de forma lineal, mostrando los datos actuales seleccionados. En el lado derecho se encuentra una vista general (el contexto), que también incluye un control deslizante para seleccionar el período del enfoque.

Se pretende descubrir y analizar los tiempos en los que la participación del usuario es más representativa aplicando una línea de tiempo en espiral

Representación de nube de palabras

Las nubes de etiquetas proporcionan un medio para que los usuarios se formen una impresión general del conjunto de los contenidos y la esencia de lo que se trata. Son una lista de las etiquetas que, por lo general se muestran en orden

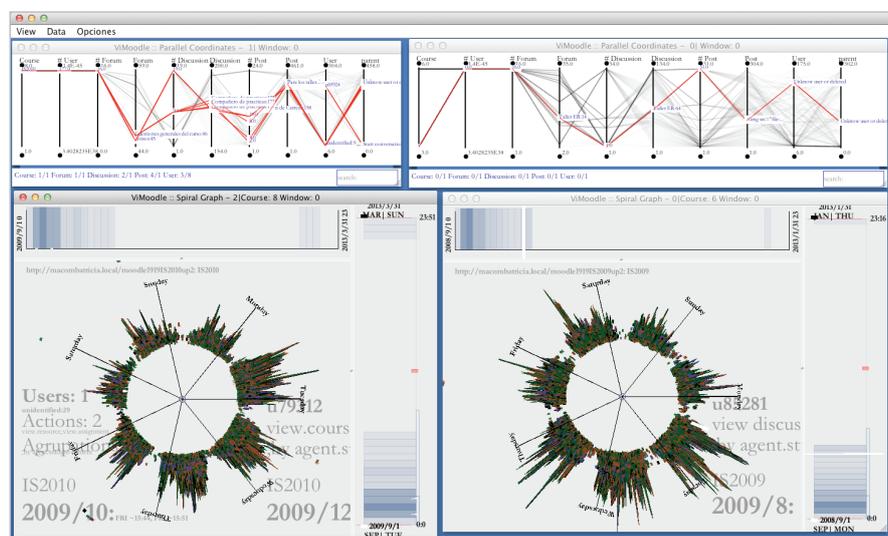


Figura 4. Representaciones de las coordenadas paralelas y la línea de tiempo en espiral

alfabético y ponderadas visualmente dependiendo de su popularidad. El tamaño de las etiquetas se corresponde con su frecuencia de uso (figura 5).

La popularidad de una etiqueta se expresa por el tamaño de la fuente (en relación con las otras etiquetas) y, por tanto, se reconoce fácilmente (Hassan-Montero; Herrero-Solana, 2006). A veces otras propiedades visuales, como color de la fuente, intensidad o peso, también se manipulan (Bateman; Gutwin; Nacenta, 2008). Las nubes de palabras pueden evolucionar a medida que los datos asociados cambian a través del tiempo. Esta necesidad de estudiar las tendencias y comprender cómo el contenido del texto o temas evolucionan en un período temporal ha sido el propósito de otras visualizaciones, por ejemplo los mini gráficos (Bongshin et al., 2010; Willett; Heer; Agrawala, 2007). Para transmitir mejor la evidencia del cambio a través de múltiples nubes de etiquetas, en esta propuesta se desarrolla un nuevo tipo de etiqueta en la nube que integra, en sí misma, un gráfico de onda y gráfico de barras (figura 5).



Figura 5. Representaciones de las coordenadas paralelas y la nube de palabras/etiquetas. Se observa que la etiqueta PRÁCTICA fue escrita al inicio del curso (expresado por la onda en azul sobre la palabra) y sin embargo fue leída hacia el final del curso (expresado por la onda en rojo).

El análisis de la nube de palabras de los contenidos del foro pone en evidencia cuáles son los conceptos más representativos para los estudiantes o los recursos más leídos, hilos y foros

Esta visualización se adapta a las necesidades del usuario, explorando todas las discusiones, foros, cursos y datos de los usuarios disponibles, y pasando de la vista general al detalle de una determinada persona, curso o discusión dentro de un período de tiempo.

Además habilitan al usuario tanto el zoom físico como el semántico. Cuando los usuarios utilizan el zoom semántico, dependiendo del contexto, la herramienta selecciona el foro, comentarios y contenido relacionados con la palabra seleccionada para su análisis.

Análisis de redes sociales

Aunque los usuarios de los LMS han construido estructuras masivas de conectividad social, las interfaces web típicas de estas construcciones sociales siguen siendo relativamente pobres. Esto causa problemas a sus miembros para explorar su comunidad en línea y evaluar la escala y los individuos a los que está expuesta su información personal de auto medición y auto reporte. La visualización de la red social se utiliza para representar el mapa de relaciones, con enlaces y frecuencias de actividades de los estudiantes y profesores, en función del

tamaño y el color de los iconos en la misma gráfica. En la visualización se implantaron como un nodo algunas entidades codificadas según se presenta en la figura 6.

La posición de cada elemento en el gráfico se basa en una simulación de la física de las fuerzas que interactúan entre los nodos, generando atracciones y repulsiones entre sí. Estas fuerzas dependen del peso de los elementos y, a su vez, del número de relaciones que cada uno de ellos tiene. Además el tamaño de estos íconos depende de su grado de entrada y salida y de su jerarquía de profundidad. Cabe mencionar que estos grados están directamente relacionados con la frecuencia de interacción del usuario, repercutiendo directamente en la distancia de los nodos hacia el punto central, el curso; quedando más alejados del nodo del curso los nodos menos activos.

La visualización de la red social se utiliza para representar el mapa de relaciones, con enlaces y frecuencias de actividades de los estudiantes y profesores

Por otra parte, en la red social diferentes nodos se pueden ocultar o cambiar de color, de forma individual o en grupo. En el caso del nodo de las personas, hay tres tipos como se puede ver en la figura 7: profesores (en azul), estudiantes (rojo a la izquierda y naranja en la parte derecha de dicha figura). Además, la red está dotada de función de búsqueda, un cuadro de texto en la base de la visualización.

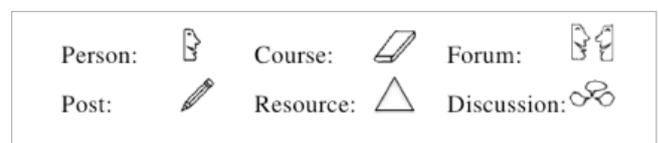


Figura 6. Elementos de la red social

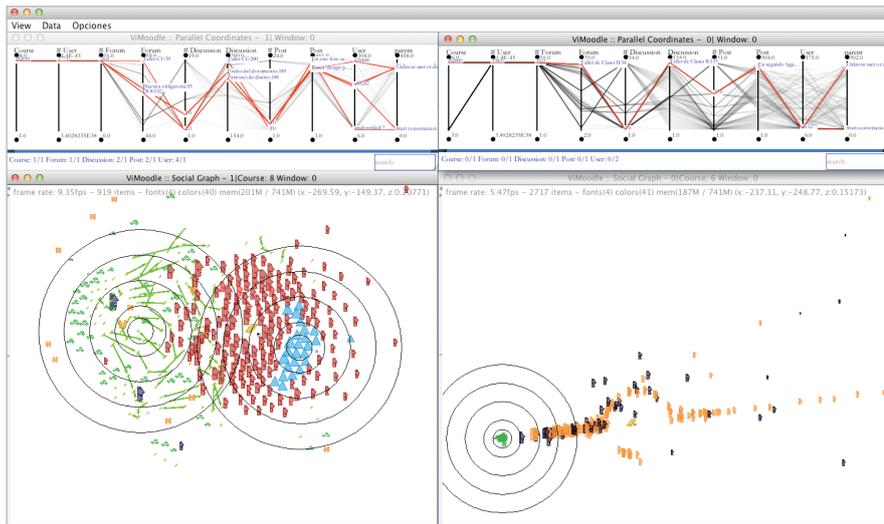


Figura 7. Representación de las coordenadas paralelas y de la red social.

Por último, dado que se ha demostrado en otros estudios la relación entre la frecuencia de la lectura y la escritura en los foros con el rendimiento de los estudiantes, y entre la frecuencia de los recursos de lectura y el rendimiento de los estudiantes (Agudo-Peregrina; Hernández-García; Iglesias-Pradas, 2012; Lipponen *et al.*, 2003), siguiendo esta idea a través del menú contextual, el usuario puede ver la distribución de los estudiantes en función de dichas relaciones (figura 7 a la derecha).

Representación de las coordenadas paralelas

Son una forma común de visualizar los datos de alta dimensión en espacio limitado y permitir la exploración de las tendencias (Inselberg; Dimsdale, 1990). En Govaerts *et al.* (2012), así como en el presente prototipo, los ejes verticales representan diferentes métricas, siendo cada estudiante una línea horizontal polígona; de igual forma para el caso de los cursos, las métricas serían el promedio de uso de los estudiantes, el promedio de notas de éstos, el número de foros, discusiones y posts, etc. (figura 8).

Los cuatro puntos de vista (nubes de palabras, red social, línea de tiempo y coordenadas paralelas) están interrelacionados. Esto significa que, dado que todos los puntos de vista son diferentes maneras de transmitir información, cuando el usuario interactúa con uno de ellos y cambia la representación, los otros puntos de vista también cambian dependiendo de la acción realizada originalmente.

3. Conclusiones

Se han aplicado cuatro representaciones para la información recuperada de los LMS en el sistema de VeLA. En concreto,

las herramientas aplicadas están orientadas a analizar y descubrir las diferentes dimensiones de la información: los tiempos en los que la participación del usuario es más representativa aplicando una línea de tiempo en espiral. Esto permite definir los momentos más importantes durante el curso académico con el fin de introducir el nuevo personal o herramientas, y proporciona información sobre otros momentos con mayor y menor participación.

Además, el análisis de la nube de palabras de los contenidos del foro pone en evidencia cuáles son los conceptos más representativos para los estudiantes o los recursos más leídos, hilos y foros. Esta información hace posible tomar decisiones sobre la inclusión de nuevos temas, herramientas, actividades, y cuando estos elementos ayudan a los estudiantes a comprender conceptos difíciles.

La aplicación de técnicas de VA facilita la toma de decisiones sobre cómo y dónde desplegar los recursos, permitiendo la optimización de los procesos de aprendizaje

Por otra parte, en muchas actividades de aprendizaje la participación se considera en la evaluación del estudiante

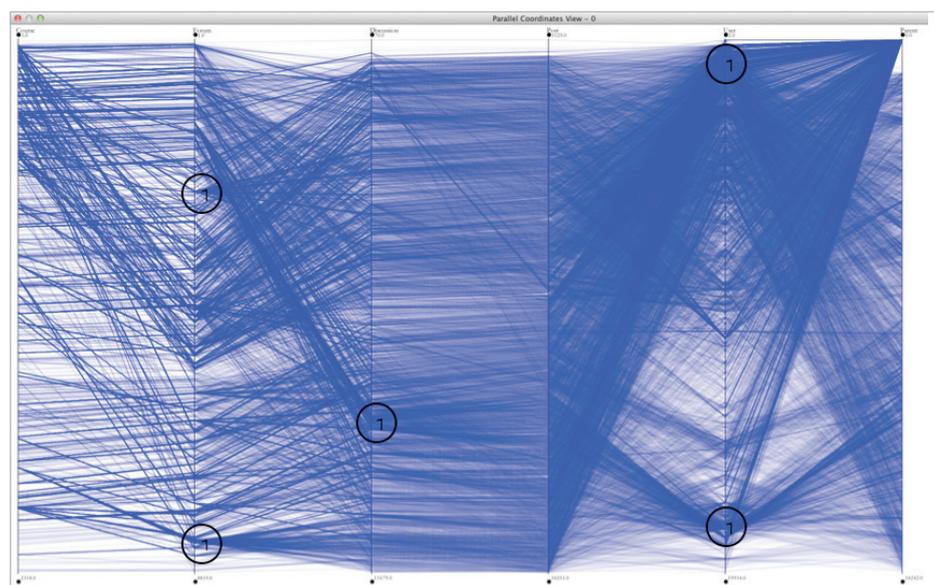


Figura 8. Representación de las coordenadas paralelas. Se representa una vista global de un campus completo de una universidad. Nótese a simple vista que con la ayuda de la transparencia en las líneas y las tonalidades, se perciben las zonas donde se acumula mayor frecuencia (indicadas con círculos negros sobre la imagen), ya sea de foros de un curso (segunda columna) o de posts de un usuario específico (quinta columna) o de discusiones de un foro (tercera columna) a simple vista y sin haber realizado ninguna interacción.

(por lo general con puntos adicionales por su intervención). Sin embargo, es difícil medir de manera eficaz esta interacción, las visualizaciones de la red social son muy útiles en este sentido, proporcionando información sobre cuáles son los usuarios más activos, las discusiones más importantes, los foros más relevantes y el nivel de participación en cada uno de ellos. En la visualización propuesta de la red social se representa la distribución de los estudiantes (figura 7 a la derecha) en función de la relación entre la frecuencia de lectura y escritura en foros y recursos, proporcionando una predicción de su rendimiento y la posible nota del curso. Obviamente cuanto mayor sea la actividad global que se realice más precisa será la predicción, posibilitando definir estrategias para mejorar la participación de los estudiantes teniendo en cuenta las experiencias anteriores y, por tanto, su rendimiento.

Gracias a estos conocimientos, la aplicación de técnicas de VA, que combina la representación visual con una interacción sobre la visualización, en el área de LA (como se describe en el modelo) ayuda sin duda a mejorar los cursos, proporciona información que de otro modo no sería conocida y facilita la toma de decisiones sobre cómo y dónde desplegar los recursos, permitiendo la optimización de los procesos de aprendizaje en la materia.

4. Agradecimientos

Los autores agradecen el apoyo del *Ministerio de Ciencia e Innovación* (proyectos FFI2010-16234 y TIN2010-21695-C02-01) y la *Consejería de Educación* de la *Junta de Castilla y León* (proyecto SA294A12-2).

5. Bibliografía

Agudo-Peregrina, Ángel F.; Hernández-García, Ángel; Iglesias-Pradas, Santiago, (2012). "Predicting academic performance with learning analytics in virtual learning environments: a comparative study of three interaction classifications". En: *International Symposium on Computers in Education (SIIE)*, 29-31 Oct., pp. 1-6.

Agudo-Peregrina, Ángel F.; Iglesias-Pradas, Santiago; Conde-González, Miguel Ángel; Hernández-García, Ángel, (2014). "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning". *Computers in human behavior*, v. 31, pp. 542-550. <http://dx.doi.org/10.1016/j.chb.2013.05.031>

Baker, Ryan S.J.D.; Yacef, Kalina, (2009). "The state of educational data mining in 2009: a review and future visions". *Journal of educational data mining*, v. 1, n. 1. <http://educationaldatamining.org/JEDM/index.php/JEDM/article/viewFile/8/2>

Bakharia, Aneesa; Dawson, Shane, (2011). "SNAPP: a bird's-eye view of temporal participant interaction". En: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (Alberta, Canada), pp. 168-173. <http://dx.doi.org/10.1145/2090116.2090144>

Barry, Bruce; Fulmer, Ingrid-Smithey (2004). "The medium and the message: the adaptive use of communication media in dyadic influence". *The academy of management review*, v.

29, n. 2, pp. 272-292.

<http://dx.doi.org/10.2307/20159033>

Bateman, Scott; Gutwin, Carl; Nacenta, Miguel, (2008). "Seeing things in the clouds: the effect of visual features on tag cloud selections". En: *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia* (Pittsburgh, USA), pp. 193-202.

<http://hci.usask.ca/publications/2008/jp039-bateman.pdf>

<http://dx.doi.org/10.1145/1379092.1379130>

Berlanga, Adriana J.; Peñalvo, Francisco G.; Sloep, Peter B., (2010). "Towards e-learning 2.0 university". *Interactive learning environments*, v. 18, n. 3, pp. 199-201.

<http://dx.doi.org/10.1080/10494820.2010.500498>

Bongshin, Lee; Riche, Nathalie-Henry; Karlson, Amy K.; Carpendale, Sheelagh (2010). "SparkClouds: visualizing trends in tag clouds". En: *IEEE Transactions on Visualization and Computer Graphics*, v. 16, n. 6, pp. 1182-1189.

<http://research.microsoft.com/EN-US/UM/REDMOND/GROUPS/cue/publications/TVCG2010-SparkClouds.pdf>

<http://dx.doi.org/10.1109/TVCG.2010.194>

Brown, John-Seely; Adler, Richard P. (2008). "Minds on fire: open education, the long tail, and learning 2.0". *Educause quarterly*, v. 42, n. 6, pp. 16-32.

<http://www.educause.edu/ero/article/minds-fire-open-education-long-tail-and-learning-20>

Chatti, Mohamed-Amine; Dyckhoff, Anna-Lea; Schroeder, Ulrik; Thüs, Hendrik (2012). "A reference model for learning analytics". *International journal of technology enhanced learning*, v. 4, n. 5/6, pp. 318-331.

http://learntech.rwth-aachen.de/dl1139%7CCDST12_IJTEL.pdf

<http://dx.doi.org/10.1504/ijtel.2012.051815>

Chen, Chaomei; Ibekwe-SanJuan, Fidelia; SanJuan, Eric; Weaver, Chris (2006). "Visual analysis of conflicting opinions". En: *IEEE Symposium on Visual Analytics Science And Technology*, pp. 59-66.

<http://fidelia1.free.fr/chen.pdf>

<http://dx.doi.org/10.1109/VAST.2006.261431>

Clow, Doug, (2012). "The learning analytics cycle: closing the loop effectively". En: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, Canada), pp. 134-138.

<http://oro.open.ac.uk/34330>

<http://dx.doi.org/10.1145/2330601.2330636>

Cook, Kris; Earnshaw, Rae; Stasko, John (2007). "Guest editors' introduction: discovering the unexpected". *Computer graphics and applications*, Sep-Oct, v. 27, n. 5, pp. 15-19.

<http://dx.doi.org/10.1109/MCG.2007.126>

De-Laat, Maarten; Lally, Vic; Lipponen, Lasse; Simons, Robert-Jan (2007). "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for social network analysis". *International journal of computer-supported collaborative learning*, v. 2, n. 1, pp. 87-103.

<http://dx.doi.org/10.1007/s11412-007-9006-4>

Gibbs, William J.; Olexa, Vladimir; Bernas, Ronan S. (2006). "A visualization tool for managing and studying online com-

munications". *Journal of educational technology & society*, v. 9, n. 3, pp. 232-243.

http://www.ifets.info/journals/9_3/20.pdf

Goldstein, Philip J.; Katz, Richard N. (2005). *Academic analytics: the uses of management information and technology in higher education*. Educause Center for Applied Research, v. 8.

<https://net.educause.edu/ir/library/pdf/ers0508/rs/ers0508w.pdf>

Gómez-Aguilar, Diego-Alonso; Conde-Gonzalez, Miguel-Ángel; Therón, Roberto; García-Peñalvo, Francisco-José (2011). "Revealing the evolution of semantic content through visual analysis". En: *11th IEEE International Conference on Advanced Learning Technologies (ICALT)* (6-8 July 2011), pp. 450-454.

<http://dx.doi.org/10.1109/ICALT.2011.141>

Gómez-Aguilar, Diego-Alonso; García-Peñalvo, Francisco-José; Therón, Roberto, (2013a). "Evaluación visual de las relaciones entre participación de los estudiantes y sus resultados en entornos de e-learning". En: *XV Simposio Internacional de Tecnologías de la Información y las Comunicaciones en la Educación Conjuntamente con el Congreso Español de Informática (CEDI 2013)*. ISBN: 978 84 695 8362 3

<http://grialdspace.usal.es:443/handle/grial/270>

Gómez-Aguilar, Diego-Alonso; García-Peñalvo, Francisco-José; Therón, Roberto (2013b). "Tap into visual analysis of the customization of grouping of activities in e-learning". En: *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality* (Salamanca, Spain), pp. 253-259.

<http://dx.doi.org/10.1145/2536536.2536575>

Gómez-Aguilar, Diego-Alonso; Therón, Roberto; García-Peñalvo, Francisco-José (2009). "Semantic spiral timelines used as support for e-learning". *Journal of universal computer science (j-uics)*, April, v. 15, n. 7, pp. 1526-1545.

<http://dx.doi.org/10.3217/jucs-015-07-1526>

Gómez-Aguilar, Diego-Alonso; Therón, Roberto; García-Peñalvo, Francisco-José (2013). "Reveal the relationships among students participation and their outcomes on e-learning environments: case study". En: *IEEE 13th International Conference on Advanced Learning Technologies (ICALT)* (15-18 July 2013), pp. 443-447.

<http://dx.doi.org/10.1109/ICALT.2013.136>

González-Teruel, Aurora; Andreu-Ramos, Carolina (2013). "Investigación del comportamiento informacional a través del análisis de redes sociales". *El Profesional de la Información*, v. 22, n. 6, pp. 522 - 528.

http://eprints.rclis.org/20686/1/Gonza%CC%81lez-Teruel_Andreu-Ramos_2013.pdf

<http://dx.doi.org/10.3145/epi.2013.nov.04>

Govaerts, Sten; Verbert, Katrien; Duval, Erik; Pardo, Abelardo (2012). "The student activity meter for awareness and self-reflection". En: *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA), pp. 869-884.

<https://lirias.kuleuven.be/bitstream/123456789/343059/1/sigchi2012-casestudy.pdf>
<http://dx.doi.org/10.1145/2212776.2212860>

Greller, Wolfgang; Drachler, Hendrik (2012). "Translating learning into numbers: a generic framework for learning analytics". *Journal of educational technology & society*, v. 15, n. 3, pp. 42-57.

http://www.ifets.info/journals/15_3/4.pdf

Hassan-Montero, Yusef; Herrero-Solana, Victor (2006). "Improving tag-clouds as visual information retrieval interfaces". En: *1 International Conference on Multidisciplinary Information Sciences and Technologies* (Mérida, Spain), pp. 25-28.

http://www.yusef.es/improving_tagclouds.pdf

Haythornthwaite, Caroline; De-Laat, Maarten (2010). "Social networks and learning networks: using social network perspectives to understand social learning". En: *Proceedings of the 7th International Conference on Networked Learning 2010* (2010).

<http://www.lancaster.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Haythornwaite.pdf>

Inselberg, Alfred; Dimsdale, Bernard, (1990). "Parallel coordinates: a tool for visualizing multi-dimensional geometry". En: *Proceedings of the First IEEE Conference on Visualization*, (23-26 Oct. 1990), pp. 361-378.

http://www.ifs.tuwien.ac.at/~mlanzenberger/teaching/ps/ws04/stuff/auth/00146402.pdf?origin=publication_detail
<http://dx.doi.org/10.1109/VISUAL.1990.146402>

Keim, Daniel; Andrienko, Gennady; Fekete, Jean-Daniel; Görg, Carsten; Kohlhammer, Jörn; Melançon, Guy (2008). "Visual analytics: definition, process, and challenges". En: Kerren, Andreas; Stasko, John; Fekete, Jean-Daniel; North, Chris (eds.) *Information visualization*, Springer Berlin Heidelberg, pp. 154-175. ISBN: 978 3 540 70955 8

http://hal-lirmm.ccsd.cnrs.fr/docs/00/27/27/79/PDF/VChapter_final.pdf

http://dx.doi.org/10.1007/978-3-540-70956-5_7

Keim, Daniel A.; Kohlhammer, Jörn; Ellis, Geoffrey; Mansmann, Florian (2010). "Mastering the Information age-solving problems with visual analytics". Florian Mansmann. <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf>

Keim, Daniel A.; Mansmann, Florian; Thomas, Jim (2010). "Visual analytics: how much visualization and how much analytics?". *SIGKDD Explorations newsletter*, v. 11, n. 2, pp. 5-8.
<http://dx.doi.org/10.1145/1809400.1809403>

Keim, Daniel; Zhang, Leishi (2011). "Solving problems with visual analytics: challenges and applications". En: *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies* (Graz, Austria), pp. 1-4.

<http://bib.dbvis.de/uploadedFiles/343.pdf>

<http://dx.doi.org/10.1145/2024288.2024290>

Lipponen, Lasse; Rahikainen, Marjaana; Lallimo, Jiri; Hakkarainen, Kai (2003). "Patterns of participation and discourse in elementary students' computer-supported collaborative learning". *Learning and instruction*, v. 13, n. 5, pp. 487-509.

[http://dx.doi.org/10.1016/S0959-4752\(02\)00042-7](http://dx.doi.org/10.1016/S0959-4752(02)00042-7)

Long, Phil; Siemens, George (2011). "Penetrating the fog: analytics in learning and education". *Educause review*, v. 46, n. 5, pp. 30-32.

<https://net.educause.edu/ir/library/pdf/ERM1151.pdf>

Modjeska, David (1997). *Navigation in electronic worlds: research review for depth oral Exam*. University of Toronto. Department of Computer Science.

Romero, Cristóbal; Ventura, Sebastián (2010). "Educational data mining: a review of the state of the art". *IEEE Transactions on Systems, Man, and Cybernetics. Part C: applications and reviews*, v. 40, n. 6, pp. 601-618. <http://dx.doi.org/10.1109/TSMCC.2010.2053532>

Shneiderman, Ben (1996). "The eyes have it: a task by data type taxonomy for information visualizations". En: *Proceedings IEEE Symposium on Visual Languages*, pp. 336-343. <http://dx.doi.org/10.1109/VL.1996.545307>

Siemens, George; Baker, Ryan S.J.d. (2012). "Learning analytics and educational data mining: towards communication and collaboration". En: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (Vancouver, Canada), pp. 252-254. <http://users.wpi.edu/~rsbaker/LAKs%20reformatting%20v2.pdf> <http://dx.doi.org/10.1145/2330601.2330661>

Siemens, George; Gasevic, Dragan; Haythornthwaite, Caroline; Dawson, Shane; Buckingham Shum, Simon; Ferguson, Rebecca; Duval, Erik; Verbert, Katrien; Baker, Ryan S.J.d. (2011). *Open learning analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. <http://solaresearch.org/OpenLearningAnalytics.pdf>

Silva, André; Figueira, Álvaro (2012). "Visual analysis of online interactions through social network patterns". En: *IEEE 12th International Conference on Advanced Learning Technologies (ICALT)* (4-6 July), pp. 639-641. <http://dx.doi.org/10.1109/ICALT.2012.57>

Thomas, James J.; Cook, Kristin A. (2005). *Illuminating the path: the research and development agenda for visual*

analytics. National Visualization and Analytics Center. http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf

Thomas, James J.; Cook, Kristin A. (2006). "A visual analytics agenda". *Computer graphics and applications*, v. 26, n. 1, pp. 10-13. <http://dx.doi.org/10.1109/MCG.2006.5>

Van-Barneveld, Angela; Arnold, Kimberly, E.; Campbell, John, P. (2012). *Analytics in higher education: establishing a common language*. <http://qa.itap.purdue.edu/learning/docs/research/ELI3026.pdf>

Van-Ham, Frank; Schulz, Hans-Jörg; Dimicco, Joan M. (2009). "Honeycomb: visual analysis of large scale social networks". En: *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (Uppsala, Sweden), pp. 429-442. <http://www.informatik.uni-rostock.de/~hs162/pdf/interactpaper.pdf> http://dx.doi.org/10.1007/978-3-642-03658-3_47

Viégas, Fernanda B.; Donath, Judith (2004). "Social network visualization: can we go beyond the graph". En: *Workshop on Social Networks for Design and Analysis: Using Network Information in CSCW' 04* (Chicago, USA), pp. 6-10. <http://wiki.commres.org/pds/CommunicationTheory/SocialNetworkVisualization.pdf>

Von-Landesberger, Tatiana; Görner, Melanie; Schreck, Tobias (2009). "Visual analysis of graphs with multiple connected components". En: *IEEE Symposium on Visual Analytics Science and Technology. VAST 2009* (12-13 Oct.), pp. 155-162. <http://dx.doi.org/10.1109/VAST.2009.5333893>

Willett, Wesley; Heer, Jeffrey; Agrawala, Maneesh (2007). "Scented widgets: Improving navigation cues with embedded visualizations". En: *IEEE Transactions on Visualization and computer graphics*, v. 13, n. 6, pp. 1129-1136. http://vis.berkeley.edu/papers/scented_widgets

Da visibilidad a tu trabajo depositándolo en e-LIS, el mayor repositorio internacional sobre biblioteconomía, documentación y comunicación

<http://eprints.rclis.org>

HACIENDO VISIBLE LO INVISIBLE: VISUALIZACIÓN DE LA ESTRUCTURA DE LAS RELACIONES EN RED EN TWITTER POR MEDIO DEL ANÁLISIS DE REDES SOCIALES

Miguel Del-Fresno-García



Miguel Del-Fresno-García es profesor en la *Universidad Nacional de Educación a Distancia (UNED)*, doctor en sociología (*UNED*), DEA y master en sociedad de la información y el conocimiento (*UOC*), MBA y executive Master en e-business (*IE Business School*) y licenciado en Filosofía (*UCM*). Imparte docencia en masters de comunicación en la *Universidad del País Vasco (UPV)*, *UCM*, *Universidad de Sevilla* y *Universidad Jaume I*. Profesor visitante en la *Universidad de la República de Uruguay (Montevideo)*. Investigador visitante en *U. C. Berkeley (EUA)*, *U. C. San Diego (EUA)*. Anteriormente ha sido director de marketing y comunicación y director de unidad de negocio en *Elsevier España*; director de marketing y comunicación de la cadena de librerías *Casa del Libro*; director de marketing y miembro del equipo fundador de *casadellibro.com*. Sus últimos libros sobre investigación social online son *El consumidor social* y *Netnografía*.
<http://orcid.org/0000-0003-2746-3681>

Universidad Nacional de Educación a Distancia
Obispo Trejo, s/n. 28040 Madrid, España
mdelfresno@der.uned.es

Resumen

Vivimos inmersos en una densa red social de interacciones que nos conectan a personas, información, eventos y lugares facilitando o restringiendo el flujo de información, ideas y percepciones, en un instantáneo y masivo sistema de comunicación en red. Creamos artefactos cognitivos que integran la visualización de datos y los procesos cognitivos para incrementar nuestras capacidades para comprender la creciente complejidad del mundo externo, al que se han sumado los social media de internet. Se combina *Twitter* como campo de investigación, y el análisis de redes sociales (ARS) como marco analítico y conjunto de métodos para capturar, analizar y visualizar datos empíricos tanto de forma estática (imágenes representando redes en ciertos puntos del tiempo) como dinámica (vídeo reflejando datos longitudinales). Se presenta *Flocker*, una herramienta diseñada para la visualización y representación en tiempo real de la actividad en red en *Twitter*.

Palabras clave

Visualización de información, Visualización de redes sociales, Análisis de redes sociales, ARS, *Twitter*, Medios sociales, *Flocker*.

Title: Making the invisible visible: visualizing the network structure of *Twitter* relationships using Social Network Analysis

Abstract

We live in a dense social network of interactions that connect us to people, information, events and places facilitating or constraining the flow of information, ideas, perceptions, etc., into an instant and massive network communication system. We create cognitive artefacts that integrate data visualization and cognitive processes in order to increase our ability to understand the complexity of the external world, made even more complex with the Internet and rise of social media. In this paper, we combine *Twitter*, as a type of fieldwork site, and Social Networks Analysis (SNA), as both an analytic framework and set of methods, to capture, analyze, and visualize empirical data, both statically (figure representing point in time) and dynamically (video reflecting longitudinal data). We present *Flocker*, a tool designed for real-time visualization and representation of network activity using *Twitter* as an illustrative case.

Keywords

Information visualization, Social network visualization, Social network analysis, SNA, *Twitter*, Social media, *Flocker*.

Del-Fresno-García, Miguel (2014). "Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en *Twitter* por medio del análisis de redes sociales". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 246-252.

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

1. Introducción

La era de la información (**Castells**, 2001) o sociedad de la información (**Mattelart**, 2002) se caracteriza por los cambios sociales sucedidos desde las décadas finales del siglo XX, derivados de la evolución de las tecnologías digitales de información y comunicación junto con la emergencia de una estructura social en red, que ha afectado a todos los ámbitos de la actividad humana a escala global.

Los cambios sociales producidos por la innovación tecnológica han provocado “variaciones en la ordenación de las relaciones interpersonales” (**Del-Fresno-García**, 2011a, p. 20) y una creciente tecnologización de la realidad (**Virilio**, 1991), el mundo externo, con un crecimiento exponencial de los datos y la información disponible hasta convertirse en un problema colectivo: el exceso de información o *information overload* (**Toffler**, 1970; **Toffler**; **Toffler**, 1994). Para **Alvin Toffler** el exceso de información desencadena un mecanismo psicológico de defensa que provoca que, individuos y sociedades, tiendan a simplificar el mundo externo y lleva a la selección de aquellos datos que confirman los prejuicios previos, ignorando el resto de los datos existentes. Ello provoca que el acceso, el uso y la comprensión del conocimiento existente sea cada día más difícil.

Si la década de los años 70 del siglo XX, según **Paul Krugman**, fue el punto más alto en la generación “de grandes cantidades de teoría aplicada desde muy pequeñas cantidades de datos” (**Silver**, 2013, p. 7), en los inicios de la segunda década del siglo XXI emerge el *big data*. La promesa implícita del *big data* es que la solución al exceso de información pasa por mayores cantidades de datos. Lo que se ha denominado como *era del petabyte*, “un mundo en el que ingentes cantidades de datos y las matemáticas aplicadas reemplazan cualquier otro instrumento”, lo que implica que “el volumen de los datos obviará la necesidad de teoría, e incluso del método científico” (**Anderson**, 2008).

La cognición como proceso de “adquisición o uso del conocimiento” tiene la ventaja de “centrarse tanto en el objeto de visualización como en el medio” (**Card**; **Mackinlay**; **Shneiderman**, 1999, p. 6). El desarrollo de técnicas y métodos para incrementar nuestras capacidades analíticas para el pensamiento y el razonamiento, cognición externa o *external cognition* (**Scaife**; **Rogers**, 1996), es una necesidad ya que los errores pueden ser producidos también de forma masiva. La crisis económica de finales de 2007, en la cual los datos que mostraban los problemas estaban disponibles con antelación pero no se interpretaron de forma correcta, los atentados del 11-S, del 15-M, etc., muestran la dimensión que puede alcanzar una interpretación deficiente de gran cantidad de datos existentes. Así, el *big data* podría estar ya mostrando dos debilidades significativas:

- una deficiencia metodológica: el foco se pone en los grandes volúmenes de datos y no en el proceso de toma de decisiones correctas;
- ingenuidad epistemológica: el futuro nunca es una prolongación lineal del pasado.

Incluso con las mejores técnicas de *big data* es imposible distinguir hoy por hoy lo imposible de lo altamente impro-

vable. Somos nosotros quienes debemos dotar a los datos de sentido y significado. La demanda debe ser mejores capacidades analíticas para comprender los significados de los datos y no, de forma determinista, más cantidades de datos. Usando la visión para pensar, las técnicas de representación de datos e información permiten crear artefactos cognitivos que posibilitan la integración de la visualización y los procesos cognitivos con el objetivo de amplificar -más y mejor- nuestra comprensión de patrones no identificados en la creciente complejidad del mundo externo. La visualización de forma estática o dinámica de las relaciones, no físicas, basadas las interacciones sociales en los social media de internet, en concreto en *Twitter*, es el objeto del presente trabajo de investigación.

2. Social media: nosotros los medios

Internet es el fenómeno socio-tecnológico más disruptivo para la comunicación, al menos, desde la llegada de la televisión. Las relaciones que se dan en internet y sus social media forman parte destacada de ese mundo externo, se producen de forma masiva y tienden a permanecer invisibles. Capturar y representar esas relaciones nos enfrenta con nuestros límites cognitivos ante la necesidad de comprender la sociabilidad en un contexto sin precedentes como el ciberespacio, “un lugar sin espacio” (**Del-Fresno-García**, 2011a, p. 19). El resultado es que hoy todo está conectado: personas, información, eventos y lugares, creando relaciones entre sí y formando de manera agregada una vasta red social donde “ya no es ni significativa ni oportuna una separación ontológica ni fenomenológica entre el mundo online y offline” (**Del-Fresno-García**, 2011a, p. 61) como parte de un mismo continuum social.

Las relaciones que se dan en internet y sus social media se producen de forma masiva y tienden a permanecer invisibles

2.1. Autocomunicación de masas y micromedios

Desde el punto de vista social y relacional, internet y su arquitectura han permitido la evolución y uso de plataformas orientadas de manera específica a la interacción personal (*Facebook*, *Twitter*, blogs, foros, agregadores de opiniones, medios de comunicación profesionales sólo online, etc.). Desde un punto de vista técnico, los social media son aplicaciones que permiten la creación e intercambio de contenidos generados por los usuarios (**Kaplan**; **Haenlein**, 2009) por medio de la interacción, síncrona y asíncrona, con un alcance sin precedentes en la historia de la humanidad. Ello supone un complejo entramado relacional entre prácticas comunicativas y sociales e infraestructura tecnológica dando lugar a un nuevo ecosistema de comunicación interpersonal colectiva (**Del-Fresno-García**, 2012).

El estadio básico de captura y representación de datos de internet son las técnicas de recuperación de información (*information retrieval*) por medio de herramientas de monitorización (**Astigarraga et al.**, 2011; **Del-Fresno-García**,

2011b; **González-Gálvez; Rey-Martín; Cavaller-Reyes, 2011; Sagarminaga-Gutiérrez; Canals, 2011**), lo que se puede considerar un *press clipping* online, que carece, sin embargo, de la posibilidad de mostrar cómo se comparte la información en red entre entidades (individuos, organizaciones, etc.).

A través de un nuevo conjunto de métodos que unen la captura y visualización de las interacciones en los social media y el análisis de redes sociales (ARS) o *social network analysis* (SNA) es posible representar y analizar esas interacciones como redes con una profundidad y escala nunca antes posible. Una de las plataformas más destacadas por su dinamismo y crecimiento, local y global, es *Twitter*.

2.2. Por qué *Twitter*

Twitter es un medio social online global y gratuito, sin un claro modelo de negocio definido desde sus inicios, que combina elementos de los blogs, mensajes de texto y emisión de información. Los usuarios escriben mensajes cortos limitados a 140 caracteres, conocidos como *tweets*, que se envían a todos aquellos que han optado por recibir los *tweets* del emisor. En cada *tweet* es posible enlazar o incrustar vídeo, imágenes y *hashtags* (una palabra o frase prefijada con el símbolo # que las convierte en metadatos). Aunque puede parecer difícil una comunicación significativa en *Twitter*, los usuarios “han encontrado formas creativas para conseguir el máximo rendimiento de cada *tweet* mediante el uso de diferentes herramientas de comunicación” (**Lovejoy; Waters; Saxton, 2012, p. 313**). En *Twitter* se puede capturar una extensa variedad de información en tiempo real, y de forma retrospectiva, lo que supone un acceso sin precedentes a los registros de la actividad humana online en el tiempo.

Twitter puede ser definido como un *microblogging* conversacional y, además de millones de personas, muchas empresas, organizaciones, medios de comunicación, ONGs o gobiernos lo utilizan para fines de comunicación tales como: marketing, atención al cliente, noticias, o incluso, activismo. Dadas sus propiedades aparentemente simples e inherentemente flexibles, en particular la brevedad y la velocidad, *Twitter* seguirá adaptándose a nuevos usos y usuarios en el futuro a pesar de los pronósticos de los escépticos y es muy probable que siga creciendo impulsado por el crecimiento de las aplicaciones para móviles, para los que *Twitter* es especialmente adecuado (**Arceneaux; Schmitz, 2010**) y ser un medio de intersección de *every media & medium* (**Dorsey, 2012**).

El interés de *Twitter* para esta investigación es doble:

- 1) permite el intercambio de comunicación de forma masiva e interpersonal que puede ser capturada, almacenada, representada y analizada;
- 2) los usuarios se expresan al menos por medio de dos tipos de conexiones explícitas: *retweets* (RTs) y menciones o respuestas (MTs).

Se trata de dos tipos de conexiones explícitas, más relevantes que el indicador común de número de seguidores/seguidos. MTs y en especial RTs, son actos significativos al demostrarse que “sólo el 6% de todos los *tweets* son republicados” (**Elmer, 2013, p. 19**) y “menos de uno cada 200 *tweets* son *retweeteados* después de una hora de ser publicados” (**Geere, 2010**).

3. Análisis y visualización de redes sociales online

Vivimos inmersos en una densa red de interacciones y relaciones sociales que nos unen a otros y facilitan el flujo de información, ideas, percepciones, etc., en un instantáneo y masivo sistema de comunicación, global y local, en red. Como metáfora, el término red o *network* es muy gráfico, ya que evoca imágenes de nodos (individuos) conectados por medio de líneas (relaciones). Una red social consiste en “un conjunto finito de actores y la relación definida entre esos actores” (**Wasserman; Faust, 1994, p. 20**) considerados como unidades interdependientes y que pueden ser representadas por medio de un grafo.

La teoría y aplicación del ARS ofrece un conjunto de métodos para el análisis de las interacciones sociales que permiten de forma específica investigar las estructuras relacionales y su representación como redes. Los investigadores en ARS -desde las matemáticas, la física y las ciencias computacionales- han aportado conceptos y métricas aplicables al estudio sistemático de procesos sociales impensables hace unos años para las ciencias sociales. El ARS proporciona tanto acceso a nuevo conocimiento como la representación de las estructuras relacionales y como éstas pueden ser consecuencia de la acción individual y colectiva. Para **Linton C. Freeman** (1979; 2000) el ARS tiene cuatro características singulares:

- a) asume una intuición estructural de las relaciones sociales;
- b) la recogida de datos empíricos ha de realizarse de forma sistemática;
- c) los modelos matemáticos son una parte fundamental del análisis con la ayuda de ordenadores como herramientas para su explotación y visualización;
- d) crear y compartir las visualizaciones de las relaciones, los patrones de las interacciones, permite la generación de *insights* estructurales significativos y su comunicación a otros.

Su aplicación al contexto de los social media o web 2.0 aún está en fase de desarrollo embrionario (**Xu; Lin, 2013; Del-Fresno-García; Daly, 2014**) y es de esperar que se desarrollen más investigaciones empíricas en breve.

“*Degree* o grado es una medida simple que mide el número de conexiones de un nodo y esta relacionado con la influencia y capacidad de difusión en red”

4. Metodología

Se presentan dos métodos, directo e indirecto, de capturar y extraer los datos relacionales de *Twitter* a partir de conexiones explícitas (RTs y MTs) para su posterior representación, visualización y análisis de forma estática (imágenes representando redes en ciertos puntos del tiempo) y dinámica (vídeo reflejando datos longitudinales) como redes por medio del ARS.

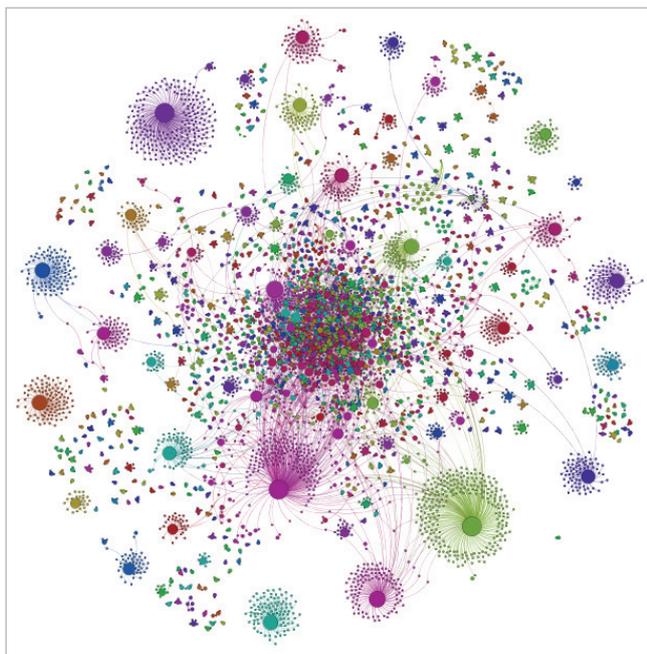


Figura 1. Red de RT con *in-degree*

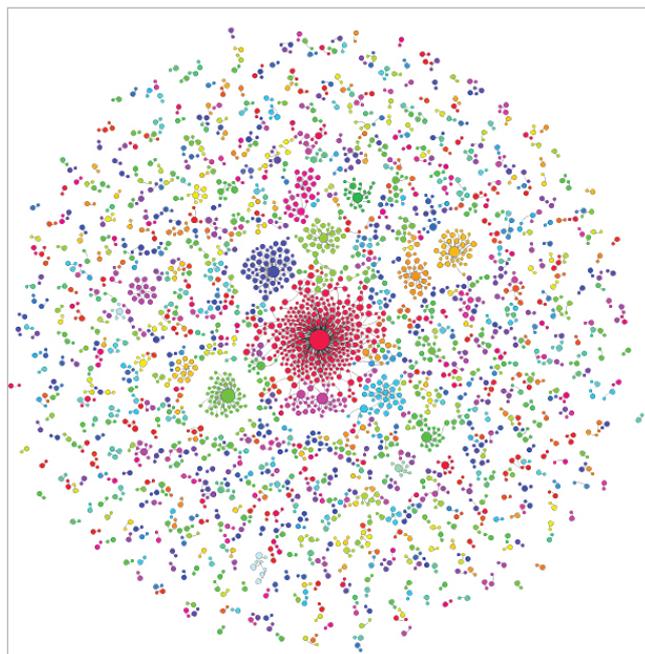


Figura 2. Red de MT con *in-degree*

El concepto de centralidad en ARS aborda la cuestión de qué nodos son los más importantes o centrales en una red. *Degree* o grado es una medida simple que mide el número de conexiones de un nodo y esta relacionado con la influencia y capacidad de difusión en red. El *degree* de un nodo se calcula como la suma de *in-degree* (grado entrante) y *out-degree* (grado saliente). El *in-degree* es el número de conexiones que apuntan hacia un nodo y *out-degree* es el número de conexiones que se originan en un nodo apuntando hacia otros nodos. En ARS existen numerosos indicadores para interpretar diferentes roles de los nodos en una red. En las imágenes presentadas el tamaño de los nodos está relacionado con su *degree* o *in-degree*.

4.1. Captura directa de datos contra la API de *Twitter*

La captura y extracción de datos desde *Twitter* se realiza directamente contra la API (*application programming interface*) de *Twitter*¹. Tras la definición previa del objeto de estudio por medio de una palabra o frase clave, *hashtag*, etc., se define el campo de estudio:

- por el número de *tweets* total que se desean capturar -desde miles a millones-
- por un lapso temporal -desde horas a meses-.

Esta amplitud en volumen de datos y lapso de tiempo sólo es viable con este método.

Tras la captura y extracción se genera un fichero con la totalidad de las interacciones del que se generan dos nuevos ficheros (red de RTs y red de MTs). Los *nicknames* de los usuarios pueden ser sustituidos con un ID único para garantizar el anonimato. Los ficheros (la red de RTs, la red de MTs y la red total) se generan en formatos importables desde software de ARS² para su representación, visualización y explotación analítica.

4.2. *Flocker*: captura indirecta de datos de *Twitter*

Para evitar requerimientos técnicos de programación se ha

desarrollado *Flocker*³, una herramienta online específica para la visualización y representación en tiempo real de la actividad en *Twitter*. Tras la definición del objeto de estudio por medio de una palabra clave (en el ejemplo: EPA), el campo se define:

- por el número de *tweets* a capturar
- por el lapso temporal que el usuario deberá definir.

Una vez obtenida la red es posible exportar las imágenes de la red obtenida (.svg o .png) y la propia red (.gexf). *Flocker* también permite el acceso a datos adicionales: laboratorio de datos, *tag cloud* de la red y *rankings* de *tweets*, usuarios y urls con más RTs.

<http://flocker.outliers.es>

5. Resultados

5.1. Captura directa de datos y representación en red en imágenes

La representación de las redes de RT y MTS (figuras 1 y 2):

- 1) Figura 1. Red RTs: n=6.230 nodos y 5.568 conexiones, aplicando el algoritmo de *in-degree*.
- 2) Figura 2. Red MTs: n=3.036 nodos y 1.999 conexiones, aplicando el algoritmo de *in-degree*.
- 3) Figura 3. Red de imágenes (nodos en rojo) más compartidas entre usuarios (nodos en azul) de una red por medio de RTs: n=4.514 nodos y 6.040 conexiones, aplicando el algoritmo de *in-degree*.

5.2. Captura indirecta de datos y representación en red en imágenes

Para la captura y representación de las redes con *Flocker* (figura 3) y *Gephi*⁴ (figura 4) con la *keyword* de búsqueda "EPA" de un total de 9.976 *tweets* genera la siguiente red.

- 1) Figuras 4 y 5. Red: n=5.248 nodos y 6.212 conexiones, aplicando el algoritmo de *in-degree*.

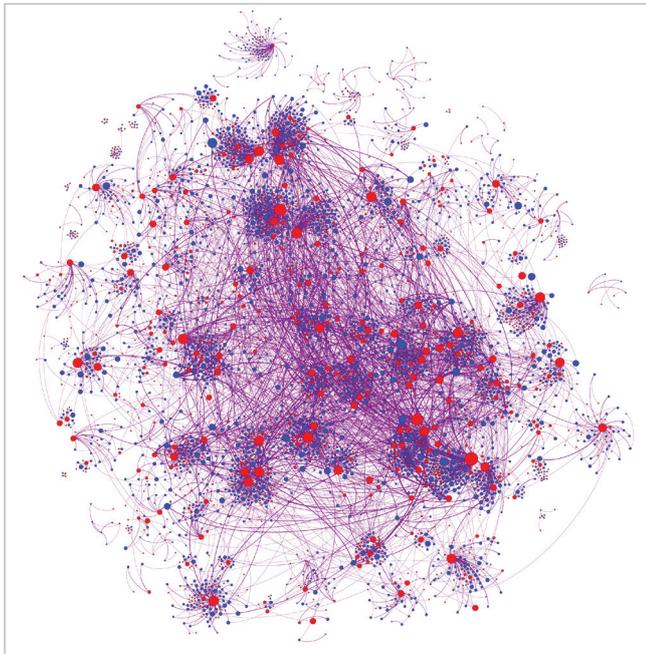


Figura 3. Red usuarios (azul) e imágenes (rojo) como más RTs con *in-degree*

5.3. Captura directa de datos y representación en vídeo

A partir de los datos del apartado 5.1 y utilizando el método propuesto por **Esteban Moro** (2012) se generaron dos vídeos de la red completa presentada en 5.1:

- red de tweets instantánea: SNA ARS *instantaneous 1*
<https://www.youtube.com/watch?v=ofY9OG-LKQI>
- red de tweets agregada: SNA ARS *aggregated 2*
<https://www.youtube.com/watch?v=suz1YoC5Qek>

5.4. Captura indirecta de datos y representación en vídeo

Utilizando *Flocker* y la *keyword* de búsqueda “New York” se capturaron 5.000 *tweets*, y con software de captura de pantalla se generó un vídeo (velocidad 8X) de la creación de la red y sus relaciones: *New York around 5.000 tweets*
<https://www.youtube.com/watch?v=rUkkq-YHi20>

6. Discusión

Gracias a la creciente capacidad computacional, el desarrollo de tecnologías específicas y la arquitectura de internet es posible la captura, representación y análisis de los datos relacionales online, con el objetivo de amplificar nuestra cognición sobre el mundo externo que son internet y sus social media. Las redes sociales de las que formamos parte juegan un papel central en nuestras actividades y despliegue social. Visualizar y comprender la influencia de estos patrones de interacción y su relación con una amplia variedad de comportamientos es algo clave para diferentes campos del conocimiento. El ARS aplicado a los social media de internet, en este caso a *Twitter*, permite hacer visible lo que hasta hace poco era impensable visualizar: la estructura de las relaciones en redes en el ciberespacio. Una de las aplicaciones prácticas directas del ARS a diversos ámbitos de la comunicación es la identificación de nuevos tipos de *social media influencers* (**Del-Fresno-García; Daly, 2014**)

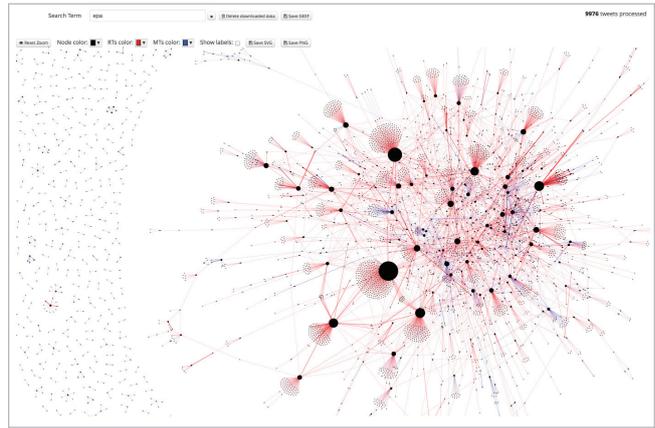


Figura 4. Red de tweets con *in-degree* (*Flocker*)

El ARS aplicado a este nuevo ámbito como la sociabilidad online facilita además de lo ya mencionado:

- identificación de individuos clave en esas redes;
- segmentación e identificación de grupos, cliques y comunidades;
- seguimiento de los comportamientos de los miembros de la red;
- cambios en las redes en una doble vertiente: cómo el despliegue social de los individuos modifica la estructura de la red y cómo la estructura de la red condiciona las posibilidades de sus miembros.

Ello abre nuevas vías de investigación sobre formas emergentes de desigualdad en red (**Del-Fresno-García, 2014**) y la posición, central o periférica, que ocupamos en la misma.

Se ha presentado un método intuitivo y empírico para la representación y visualización de la estructura de las redes y nodos clave en *Twitter*. No obstante, existen limitaciones ya que la representación de la estructura y relaciones presenta un déficit: la no comprensión de los significados que

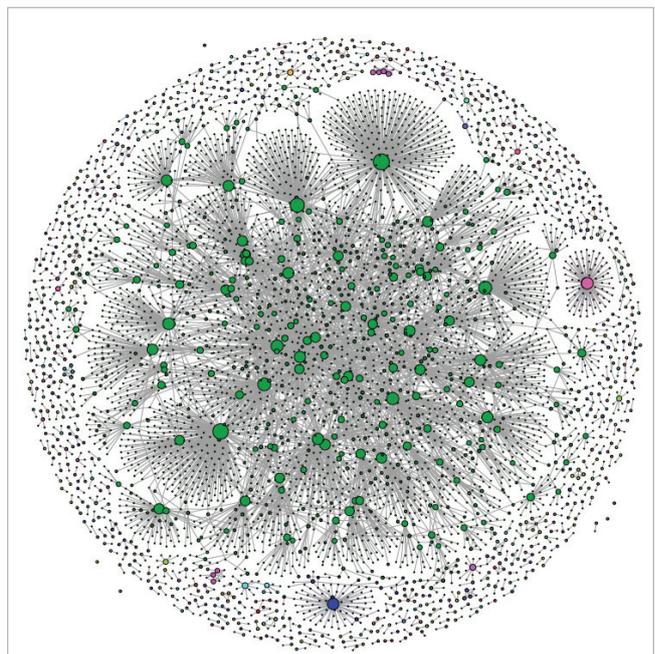


Figura 5. Red de tweets *in-degree* en *Gephi*

circulan por la red y las motivaciones de los individuos que interactúan con otros. Mientras que el ARS proporciona una manera de medir con precisión la cantidad de relaciones, no aporta una explicación cualitativa alrededor de los intercambios que se producen en red. Nuevas investigaciones y metodologías deben incorporar ambas aproximaciones. Cabe destacar la *network ethnography* (Howard, 2002) o la *netnografía* (Del-Fresno-García, 2011b; Del-Fresno-García; López, 2014), un enfoque interdisciplinario entre la sociología, antropología, lingüística y comunicación para representar y comprender la creciente complejidad, local y global, de la comunicación y sociabilidad en red.

El ARS puede ser aplicado a distintas disciplinas desde los siguientes enfoques:

- a) la identificación de individuos clave a la hora de promover la difusión de cambios de comportamiento en redes;
- b) la segmentación e identificación de grupos, cliques y comunidades;
- c) el cambio de comportamientos o adopción de innovaciones a través de los miembros internos de la red y no líderes jerárquicos;
- d) la influencia en el seno de las redes puede ser bidireccional: individuos-red y red-individuos.

Ello que abre un amplio campo de exploración a diferentes disciplinas a la hora de identificar individuos y roles clave en el seno de las múltiples redes de las que formamos parte.

7. Notas

1. El servidor de *Flocker* (Python) establece una comunicación con el API de *Twitter* (vía autorización *OAuth*) usando las credenciales de un usuario. Gestiona la entrada de datos del *streaming* de *Twitter* y genera otro *streaming* con el navegador del usuario. Este proceso aísla al usuario de la complejidad técnica con la API *Twitter*. El *javascript* en la parte cliente (navegador) recibe y procesa los *tweets* y genera y visualiza la red con *D3.js*. El factor limitador en el proceso es el propio sistema del usuario.

2. Para un listado actualizado de los programas disponibles http://en.wikipedia.org/wiki/Social_network_analysis_software

3. *Flocker* ha sido desarrollado por *Outliers*. <http://outliers.es>

4. *Gephi* es un software gratuito de código abierto para la visualización interactiva, exploración y análisis de grafos de redes. <https://gephi.org>

8. Bibliografía

Anderson, Chris (2008). "The end of theory: the data deluge makes the scientific method obsolete". *Wired*, Jun 23. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Arceneaux, Noah; Schmitz, Amy (2010). "Seems stupid until you try it: press coverage of Twitter, 2006-9". *New media & society*, v. 12, n. 8, pp. 1262-1279.

http://mc7290.bgsu.wikispaces.net/file/view/Arceneaux_Noah.pdf

<http://dx.doi.org/10.1177/1461444809360773>

Astigarraga, Eneko; Azpillaga, Josu; Fernández, Luis; Naberan, Aitzol (2011). "Umap, inteligencia colectiva extraída de las redes sociales". *El profesional de la información*, v. 20, n. 5, pp. 542-547.

<http://eprints.rclis.org/16177>

<http://dx.doi.org/10.3145/epi.2011.sep.08>

Card, Stuart; Mackinlay, Jock; Shneiderman, Ben (1999). *Readings in information visualization. Using vision to think*. San Francisco: Academic Press, 712 pp. ISBN: 978 1558605336

Castells, Manuel (2001). *La era de la información: economía, sociedad y cultura*. Madrid: Alianza editorial, 645 pp. ISBN: 978 8420647401

Del-Fresno-García, Miguel (2011a). *Netnografía*. Barcelona: UOC, 112 pp. ISBN: 978 84 9788 385 6

Del-Fresno-García, Miguel (2011b). "Infosociabilidad monitorización e investigación en la web 2.0 para la toma de decisiones". *El profesional de la información*, v. 20, n. 5, pp. 548-554.

<http://eprints.rclis.org/16150>

<http://dx.doi.org/10.3145/epi.2011.sep.09>

Del-Fresno-García, Miguel (2012). "Comprendiendo los social media y mass media: un modelo para el estudio de la comunicación interpersonal colectiva en tiempos de internet". *Derecom*, n. 11, pp. 99-109.

<http://derecom.com/numeros/pdf/delfresno.pdf>

Del-Fresno-García, Miguel (2014). "Connecting the disconnected: social work and social network analysis. A methodological approach for identifying within network peer leaders". *Arbor*, 2014 (en prensa).

Del-Fresno-García, Miguel; López, Antonio (2014). "Social work and netnography: the case of Spain and generic drugs". *Qualitative social work*, v. 13, n. 1, pp. 85-107.

<http://migueldefresno.com/2014/02/netnography-social-work-the-case-of-spain-and-generic-drugs.html>

<http://dx.doi.org/10.1177/1473325013507736>

Dorsey, Jack (2012). "Twitter takes the pulse of the planet. It's the intersection of every media & médium". *Twitter*, November 15.

<https://twitter.com/TwitterAds/status/269129576318386177>

Elmer, Greg (2013). "Live research: twittering an election debate". *New media & society*, v. 15, n. 1, pp. 18-30.

<http://dx.doi.org/10.1177/1461444812457328>

Freeman, Linton (1979). "Centrality in social networks: conceptual clarification". *Social networks*, v. 1, pp. 215-239.

<http://moreno.ss.uci.edu/27.pdf>

Freeman, Linton (2000). "Visualizing social networks". *Journal of social structure*, v. 1, n. 1.

http://www.bebr.ufl.edu/files/Visualizing%20social%20networks_0.pdf

Geere, Duncan (2010). "It's not just you: 71 percent of

tweets are ignored". *Wired*, November 11.
<http://www.wired.com/2010/10/its-not-just-you-71-percent-of-tweets-are-ignored/>

González-Gálvez, Pilar; Rey-Martín, Carina; Cavaller-Reyes, Víctor (2011). "Redes sociales para la inteligencia competitiva. Propuesta de un índice sintético". *El profesional de la información*, v. 20, n. 5, pp. 525-532.
<http://dx.doi.org/10.3145/epi.2011.sep.06>

Howard, Philip (2002). "Network ethnography and the hypermedia organization: new media, new organizations, new methods". *New media & society*, v. 4, n. 4, pp. 550-574.
<http://dx.doi.org/10.1177/146144402321466813>

Kaplan, Andreas; Haenlein, Michael (2010). "Users of the world, unite! The challenges and opportunities of social media", *Business horizons*, v. 53, n. 1, pp. 59-68.
<http://dx.doi.org/10.1016/j.bushor.2009.09.003>

Lovejoy, Kristen; Waters, Richard; Saxton, Gregory (2012). "Engaging stakeholders through Twitter: how nonprofit organizations are getting more out of 140 characters or less". *Public relations review*, v. 38, p. 313-318.
<http://arxiv.org/pdf/1106.1852.pdf>
<http://dx.doi.org/10.1016/j.pubrev.2012.01.005>

Mattelart, Armand (2002). *Historia de la sociedad de la información*. Barcelona: Paidós, 193 pp. ISBN: 978 84 493 2042 2

Moro, Esteban (2012). "Temporal networks with igraph and R (with 20 lines of code!)". *Implicit none*, November 11.
<http://www.r-bloggers.com/temporal-networks-with-igraph-and-r-with-20-lines-of-code>

Sagarminaga-Gutiérrez, Naiara; Canals, Agustí (2011). "Re-

cuperación y tratamiento de opinión movilizada para el análisis del clima social". *El profesional de la información*, v. 20, n. 5, pp. 510-515.

<http://eprints.rclis.org/16153>
<http://dx.doi.org/10.3145/epi.2011.sep.04>

Scaife, Mike; Rogers, Yvonne (1996). "External cognition: how do graphical representations work?". *International journal of human-computer studies*, v. 45, n. 2, pp. 185-213.
<http://users.mct.open.ac.uk/yvonne.rogers/papers/externalcognition.pdf>
<http://dx.doi.org/10.1006/ijhc.1996.0048>

Silver, Nate (2013). *The signal and the noise: why so many predictions fail — but some don't*. New York: Penguin Books, 544 pp. ISBN: 978 1594204111

Toffler, Alvin (1970). *Future shock*. New York: Bantam Books, 286 pp. ISBN: 978 0553277371
<http://cdn.preterhuman.net/texts/literature/general/Alvin%20Toffler%20-%20Future%20Shock.pdf>

Toffler, Alvin; Toffler, Heidi (1994). "Surfing the third wave". *Micro times*, n. 118, January 3.
<http://mdelf.es/xx4n>

Virilio, Paul (1991). *The aesthetics of disappearance*. New York: Semiotext(e), 128 pp. ISBN: 978 1570270413

Wasserman, Stanley; Faust, Katherine (1994). *Social network analysis: methods and applications*. New York and Cambridge: Cambridge University Press, 857 pp. ISBN: 978 0521387071

Xu, Guandong; Lin, Li (2013). *Social media mining and social network analysis: emerging research*. IGI Global, 347 pp. ISBN: 978 1466628069

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



TENDENCIAS EN MAPAS DE LA CIENCIA: CO-USO DE INFORMACIÓN CIENTÍFICA COMO REFLEJO DE LOS INTERESES DE LOS INVESTIGADORES



Daniel Torres-Salinas, Evaristo Jiménez-Contreras y Nicolás Robinson-García



Daniel Torres-Salinas es doctor en documentación científica por la *Universidad de Granada*. Trabaja como técnico de gestión en la *Universidad de Navarra (UN)* donde realiza auditorías sobre el rendimiento científico y es investigador en el ámbito de la bibliometría y miembro del *Grupo EC3 (Evaluación de la Ciencia y la Comunicación Científica de la Universidad de Granada)*. Es co-creador de diferentes herramientas de evaluación de la investigación como *Cientificacvn*, los *Rankings I-UGR de Universidades* o la *Clasificación integrada de revistas científicas (CIRC)*. Es cofundador y CEO de la spin-off de asesoría científica *EC3metrics*. Habitualmente imparte cursos sobre comunicación científica, web 2.0 y ciencia o servicios de investigación en bibliotecas.

<http://orcid.org/0000-0001-8790-3314>

torressalinas@gmail.com



Evaristo Jiménez-Contreras es catedrático de bibliometría en la *Facultad de Comunicación y Documentación* de la *Universidad de Granada*, y director del grupo de investigación *EC3 (Evaluación de la Ciencia y de la Comunicación Científica)*. Sus líneas de investigación se centran en la evaluación de la actividad científica con metodología bibliométrica. Es uno de los promotores de los índices de impacto de las revistas españolas de ciencias sociales, jurídicas y humanas (respectivamente *In-Recs, In-Recj, In-Rech*), *CIRC (Clasificación integrada de revistas científicas)*, *Metaranking EC3*, y *Científica*.

<http://orcid.org/0000-0001-5668-7057>

evaristo@ugr.es



Nicolás Robinson-García es licenciado en documentación y máster en información científica por la *Universidad de Granada*. Es miembro del grupo *EC3 (Evaluación de la Ciencia y la Comunicación Científica)* de la misma universidad, donde se encuentra haciendo el doctorado. En la actualidad sus intereses se centran en la investigación sobre bibliometría y la evaluación de la calidad investigadora.

<http://orcid.org/0000-0002-0585-7359>

elrobin@ugr.es

*Universidad de Granada, EC3 Research Group
Campus Cartuja. 18071 Granada, España*

Resumen

Se explora la posibilidad de elaborar mapas de la ciencia basados en el co-uso de la información científica que hacen los usuarios de una institución académica. Se define co-uso como la co-ocurrencia de consultas de información científica por parte de pares de usuarios en una plataforma de revistas científicas. Sobre la base de los datos de consulta de la *Universidad de Navarra* en 2012 a la plataforma de revistas científicas *ScienceDirect* se analiza el potencial de dicha aproximación metodológica. Se destaca la viabilidad de dicha metodología para ahondar en los intereses investigadores de una institución académica y las relaciones que se producen entre distintas disciplinas.

Palabras clave

Mapas de la ciencia, Visualización de la información, Co-uso, Consumo de información científica, Bibliografía científica, Metodologías, Recuperación de información, Intereses de investigación, *Universidad de Navarra*.

Title: Trends in science mapping: Co-use of scientific literature as evidence of researchers' interests

Abstract

The possibility of constructing science maps based on co-use of scientific literature by academic users is explored. We define co-use as the co-occurrence of scientific information requests by pairs of users recorded in platforms of scientific journals. We used search data from the *University of Navarra* to the *ScienceDirect* platform in 2012 in order to test the validity and to analyze the possibilities of this methodological approach. We conclude by emphasizing the viability of this methodology when exploring the research interests of an academic institution along with the relations among different disciplines.

Artículo recibido el 30-01-2014

Aceptación definitiva: 21-05-2014

Keywords

Science maps, Information visualization, Co-use, Consumption of scientific literature, Methodologies, Information retrieval, Research interests, *University of Navarre*.

Torres-Salinas, Daniel; Jiménez-Contreras, Evaristo; Robinson-García, Nicolás (2014). "Tendencias en mapas de la ciencia: co-uso de información científica como reflejo de los intereses de los investigadores". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 253-258.

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

Introducción

El formato electrónico ha transformado la forma de consumo de la información científica. En 2005 la quinta parte de los artículos que consultaban los investigadores eran en formato electrónico; actualmente son más de la mitad (**Tenopir et al.**, en prensa). Además los hábitos de lectura han cambiado: los investigadores leen un mayor número de trabajos, la lectura se suele hacer de manera transversal dedicando menos tiempo por trabajo y se tiende a prestar una mayor atención a los trabajos más breves (**Van-Noorden**, 2014).

Este escenario plantea nuevas oportunidades y retos para bibliotecarios y bibliómetras, ya que resulta más sencillo monitorizar la actividad de los usuarios y analizar el consumo que hacen de la información científica en el medio digital. Se comienza a profundizar en el análisis de nuevos indicadores que puedan complementar o sustituir a los tradicionales (préstamos o lecturas en sala en el caso de los bibliotecarios, y citas en el de los bibliómetras). Surgen ambiciosos proyectos como *Counter*, dirigido a bibliotecarios y editores o *Mesur*, dirigido a la comunidad científica.

<http://www.projectcounter.org>

<http://mesur.informatics.indiana.edu>

De manera paralela los sectores profesional y académico emprenden la búsqueda de nuevos indicadores (**Borrego; Urbano**, 2005; **Torres-Salinas; Cabezas-Clavijo**, 2013), como: marcadores sociales (**Haustein; Siebenlist**, 2011), consultas en bibliotecas digitales (**Bollen et al.**, 2005), hiperenlaces (**Aguillo et al.**, 2006) o menciones en redes sociales (**Torres-Salinas; Cabezas-Clavijo; Jiménez-Contreras**, 2013). Se observan principalmente tres frentes de investigación que analizan el uso de la literatura científica utilizando indicadores alternativos a los tradicionales:

1) La webmetría (**Almind; Ingwersen**, 1997): considera internet como el mayor repositorio mundial de documentos y analiza relaciones entre ellos mediante hiperenlaces del mismo modo que lo hacen las citas (**Theilwall**, 2008). Así resulta posible analizar por ejemplo la visibilidad de las instituciones académicas en la Web (**Aguillo et al.**, 2006).

2) La bibliometría basada en el uso o *usage bibliometrics* (**Kurz; Bollen**, 2010): nació a mediados de los noventa y se deriva del análisis de las colecciones electrónicas para determinar las políticas de suscripción de las bibliotecas a paquetes de revistas y asesorar a los editores en la mejora de sus servicios (**Nicholas; Huntington; Watkinson**, 2005; **Rodríguez-Bravo et al.**, 2012). Posteriormente, **Bollen** planteó su uso como herramienta de evaluación científica (**Bollen et al.**, 2005).

3) Las *altmetrics* o métricas alternativas (**Priem et al.**, 2010): la más reciente de estas propuestas. Propone utilizar indicadores de las redes sociales y demás servicios de la web 2.0 como medidas de la atención social que reciben los trabajos científicos.

En todos estos casos se plantean dos tipos de estudios: comparar la relación de los mismos con las citas o adaptar las técnicas bibliométricas a estos nuevos indicadores. En este proceso de extrapolación metodológica uno de los aspectos al que se ha prestado menos atención es la visualización y el mapeo de la ciencia. Se trata de técnicas que se han visto revitalizadas desde los años noventa con el avance de la tecnología y su implementación como herramientas de asesoramiento de la política científica (**Noyons**, 2004). Si bien los mapas de la ciencia están basados en indicadores relacionales como la citación (**Small**, 1973) o la coautoría (**Luukkonen; Persson; Sivertsen**, 1992), en los mapas de co-uso el elemento relacional se establecería cuando dos trabajos han sido consultados en una misma sesión o por un mismo usuario. Esto permite ofrecer una visión más amplia del consumo que se hace de la bibliografía científica al plantearla desde la perspectiva del lector y no de los autores (**Rowlands; Nicholas**, 2007).

“ El objetivo de este trabajo es explorar la posibilidad de realizar mapas de la ciencia de categorías temáticas basados en co-uso de información ”

Este trabajo explora cómo trasladar las técnicas de mapeo de co-citación al entorno de las plataformas online de revistas científicas con el objetivo de generar mapas de la ciencia de co-uso. Con esta perspectiva el objetivo esencial es explorar la posibilidad de realizar mapas de la ciencia de categorías temáticas basados en co-uso de información. Se explora asimismo que estos mapas ayuden a identificar los intereses científicos de una determinada comunidad, en concreto de una universidad.

Material y métodos

Como estudio de caso se ha seleccionado la *Universidad de Navarra* con el fin de elaborar un mapa de la ciencia basado en las consultas de artículos, principalmente descargas y visualización de ficheros html, que hizo su personal en 2012. Los datos de las consultas provienen de *ScienceDirect* y fueron cedidos a través del *Elsevier bibliometric research program*¹. La plataforma *ScienceDirect* contiene más de 2.500

revistas, que cubren las grandes áreas del conocimiento científico. Para cada consulta realizada por un usuario se obtuvo un registro que incluía la IP desde la que se realizó la consulta, el identificador de sesión, la fecha y diferentes datos identificados de la revista. Para asignar una categoría temática a cada revista los datos se cruzaron con el *Scimago journal rank (SJR)*. En la tabla 1 se muestra la información básica del dataset.

A continuación se establecieron las relaciones entre categorías temáticas a partir del co-uso de los artículos. Las sesiones comienzan cada vez que un usuario accede a *ScienceDirect* y finalizan cuando éste abandona la plataforma o bien ha pasado cierto tiempo de inactividad. Cada sesión corresponde con una única demanda de información, ya que no puede haber más de un usuario en una misma sesión. Establecemos así un posible vínculo temático o de contenido entre los artículos consultados en un modo similar al que ocurre con la co-citación, donde se asume que dos trabajos citados conjuntamente están relacionados. El término co-uso hace mención por tanto a las relaciones que se establecen entre documentos consultados en una única sesión. Si dos trabajos de categorías temáticas diferentes han sido consultados, co-descargados o co-visualizados en la misma sesión, entre dichas categorías se genera el mencionado vínculo temático. Partiendo de esta idea se ha creado una matriz de distancias de co-uso entre categorías *SJR* normalizadas según el *índice de Jaccard (Leydesdorff, 2008)*. Finalmente se representó la matriz a través de *Pajek*² manteniendo sólo las relaciones más significativas y denotando el número de consultas a través del tamaño de los nodos.

Índice de Jaccard

También conocido como coeficiente de similitud fue definido por Paul Jaccard (1868-1944), es una medida de la semejanza entre dos particiones o grupos de objetos (por ejemplo, resultantes de aplicar dos tipos de categorización al conjunto).

Sea un conjunto de documentos, de los cuales se forman dos grupos:

M: los que tienen el descriptor m

N: tienen el descriptor n

Definimos:

r = número de parejas de documentos que se encuentran ambas tanto en M como en N (o sea, están simultáneamente en M y en N)

s = número de parejas de documentos que se encuentran en M pero no en N

t = número de parejas de documentos que se encuentran en N pero no en M

Índice de Jaccard = $r / (r+s+t)$

Expresado de otra manera, es la intersección de 2 conjuntos $M \cap N$

N dividida por su reunión $M \cup N$

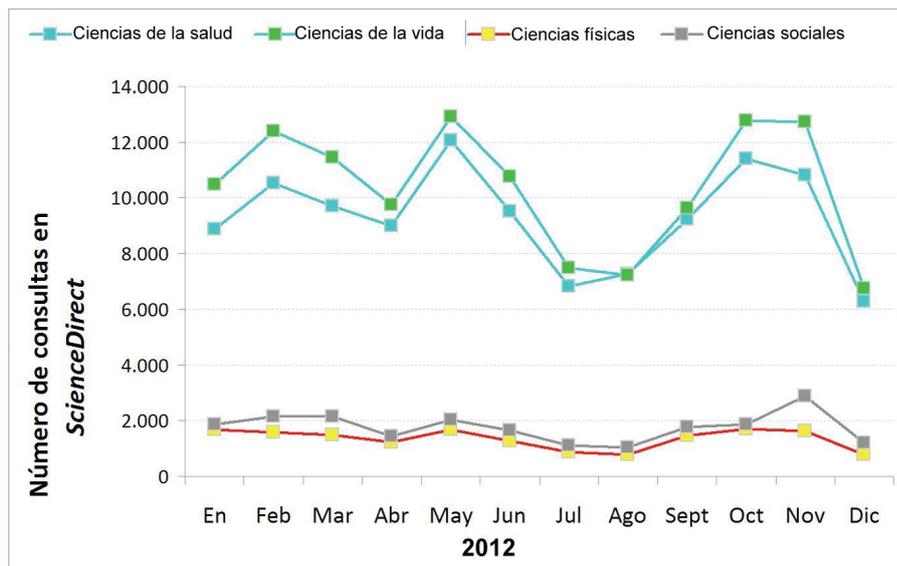


Figura 1. Distribución temporal por grandes áreas de las consultas realizadas por la Universidad de Navarra a ScienceDirect en 2012

Resultados: mapas de la ciencia de co-uso

El 75,98% de las consultas realizadas desde la Universidad de Navarra en 2012 fueron sobre trabajos pertenecientes a las áreas de ciencias de la vida y ciencias de la salud. Como se observa en la figura 1, los trabajos publicados en revistas pertenecientes al área de ciencias de la vida fueron los más consultados seguidos por ciencias de la salud. Por su parte ciencias sociales y humanidades representan poco más del 8% de las consultas recibidas mientras que ciencias físicas apenas superan el 6%.

Esta sobrerrepresentación de ciencias de la vida y de la salud se constata en el mapa de co-uso que se muestra en la figura 2. Vemos un mapa caracterizado por el predominio de las categorías de estas dos grandes áreas que quedan representadas en la parte superior. La parte superior izquierda del mapa está poblada por categorías pertenecientes a ciencias de la vida que conectan a través de farmacia y bioquímica con medicina (parte superior derecha). En el margen inferior derecho están las categorías relacionadas con economía, negocios y ciencias sociales unidas a medicina a través de enfermería. Ingenierías y ciencias de la computación conectan ciencias sociales con ciencias medioambientales (margen inferior izquierda) y ciencias físicas, que se sitúan en el centro del mapa conectando física con química y ésta con bioquímica, que cierra el círculo conectando nuevamente con ciencias de la salud.

Tabla 1. Descripción general de las consultas realizadas desde la Universidad de Navarra a ScienceDirect en 2012

Indicador	Resultado
Consultas	259.547
Sesiones	79.693
Números IP	1.845
Consultas por sesión	3,26
Sesiones por IP	43,21

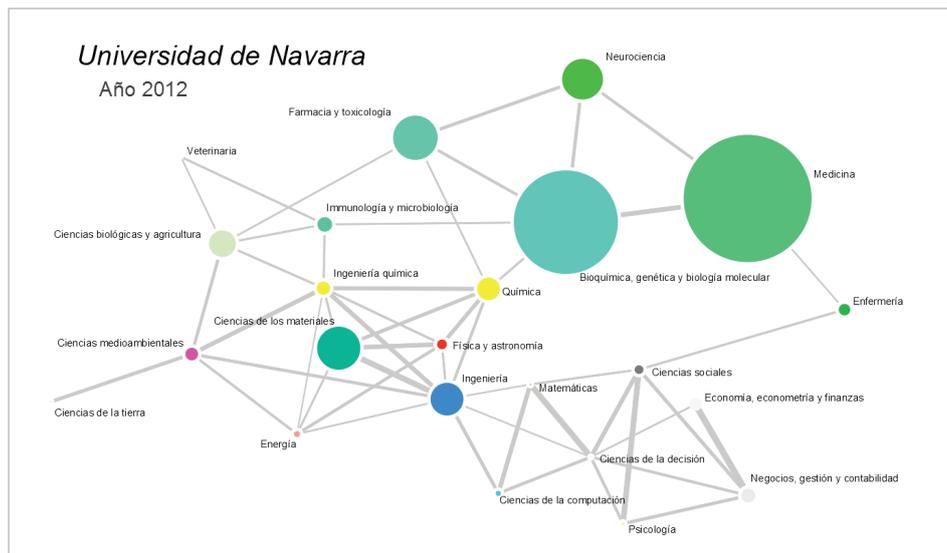


Figura 2. Mapa general de co-uso de la *Universidad de Navarra* basado en las consultas realizadas a *ScienceDirect* en 2012

Al bajar a un nivel de descripción más detallado, podemos analizar las disciplinas que conforman cada una de las grandes áreas. En la figura 3 centramos el foco en las disciplinas pertenecientes al área de la medicina. Aquí se observan tres ramas que reciben un mayor interés por parte del personal de la *Universidad de Navarra*, todas ellas unidas por la cirugía en el centro de la figura. La primera es la relacionada con la investigación oncológica, el cáncer, etc., que se sitúa en el margen superior derecho. La segunda está formada por disciplinas relacionadas con la neurología clínica como principal foco de atención. De esta rama y a través de anestesiología y medicina del dolor, cuelga reumatología y medicina del deporte. Finalmente, la tercera es la relacionada con endocrinología y diabetes y metabolismo. En este caso sin embargo la presencia de una categoría miscelánea enturbia el dibujo. Por último las disciplinas de parasitología y enfermedades infecciosas forman un clúster separado del resto de la red.

La figura 4 muestra una comparativa entre la producción de la universidad en el mismo año en el que se tomaron los datos de consulta de acuerdo con la base de datos *Scopus*. Se observa cierta coherencia en ciencias de la salud entre la producción y el uso de literatura científica. Para las otras tres áreas se producen ciertos desajustes que pueden deberse o bien a diferencias en los patrones de uso entre ramas científicas o a sesgos inherentes a las fuentes de infor-

mación utilizadas (*Scopus* y *ScienceDirect*). No obstante, ciencias de la vida y de la salud vuelven a ser las dos áreas en las que esta universidad es más productiva, frente a ciencias físicas y ciencias sociales.

La *Universidad de Navarra* presenta altas tasas de especialización temática muy por encima de la media española en medicina (2,13) y economía (2,02) (Torres-Salinas; Delgado López-Cózar; García-Moreno-Torres; Herrera, 2011). Estas dos áreas quedan perfectamente retratadas en la figura 2. También el tamaño en función de las consultas online parece estar bien representado ya que

precisamente son las áreas de medicina y bioquímica las más productivas en 2012 en *Scopus*, representando entre las dos el 46,1% de la producción de la institución.

“ El término co-uso hace mención a las relaciones que se establecen entre documentos consultados en una única sesión ”

Consideraciones finales

En este trabajo se explora la posibilidad de representar mapas que ilustren el co-uso de las publicaciones planteándose una alternativa a técnicas clásicas como la co-citación o el *bibliographic coupling*. Esta aproximación a la creación de mapas de la ciencia basados en el co-uso ya ha sido suge-

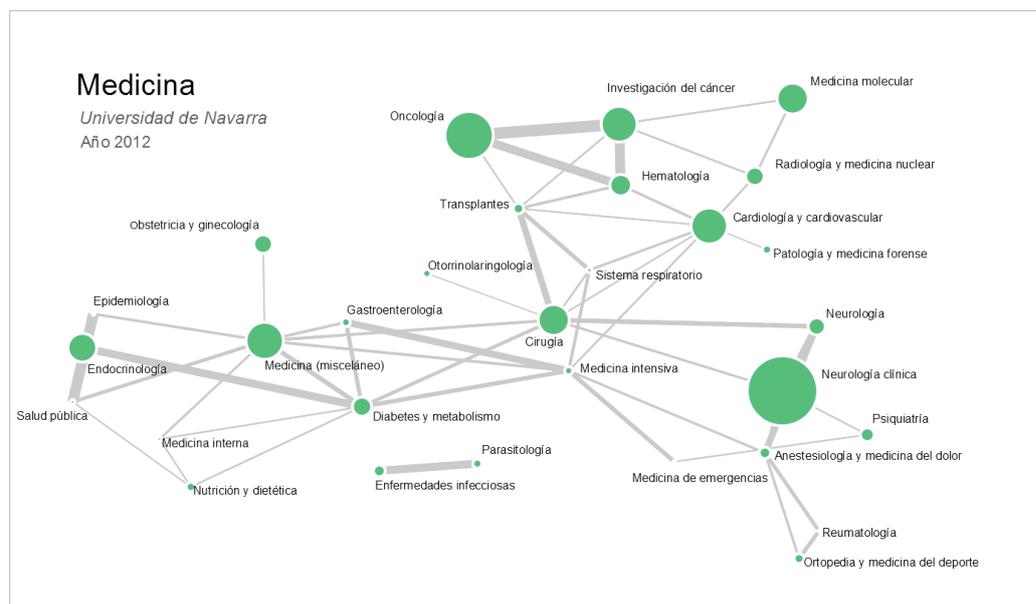


Figura 3. Mapa de co-uso del área de medicina de la *Universidad de Navarra* basado en las consultas realizadas a *ScienceDirect* en 2012

rida por otros autores (Bollen *et al.*, 2005; Brody; Harnad; Carr, 2006), siendo este trabajo el primero que aplica dicho análisis a nivel institucional para representar los intereses de su comunidad de usuarios. Los mapas basados en el uso enriquecen mucho más el análisis al incluir a otros actores pasivos en el consumo de la literatura científica como son estudiantes, profesionales o docentes.

En líneas generales se observan patrones similares de producción y de consumo de información científica (figura 4), especialmente en el área de las ciencias de la salud. La plataforma de revistas electrónicas *ScienceDirect* es ampliamente utilizada tanto en ésta área como en las ciencias experimentales (Rodríguez-Bravo *et al.*, 2012), lo que hace que se adapte perfectamente a la universidad empleada como caso de estudio al tratarse de una universidad especializada precisamente en el ámbito biomédico. Por otra parte en este artículo empleamos una definición de uso de la información científica muy amplio, considerando como tal la consulta de bibliografía científica. En un futuro sería de interés analizar y comparar estos resultados con definiciones de uso más restrictivas como descargas, visualizaciones del texto completo, etc.

La metodología de visualización mediante co-uso puede fácilmente extrapolarse a otras plataformas y contextos de la web social académica

Del mismo modo, aunque los resultados preliminares son prometedores, todavía sigue siendo necesario ahondar más en el significado de este tipo de mapas y su posible utilidad como complemento a los tradicionales mapas de la ciencia. Su carácter más inclusivo, al contemplar otras dimensiones distintas de la investigación propiamente, consulta profesional o docentes, apunta a tres posibles usuarios:

- El equipo de dirección de bibliotecas: los mapas permiten visualizar de forma rápida el consumo de *ScienceDirect* y el uso real que hacen de ella sus usuarios así como la interconexión que existe entre los intereses de su comunidad de usuarios.
- El equipo de gestión científica: la monitorización en tiempo real de la actividad intelectual de sus investigadores puede resultar de gran utilidad de cara a la toma de decisiones en materia de política científica.
- Editoriales científicas: conocer los hábitos de lectura de sus usuarios les permitirá mejorar la calidad de sus productos, analizando las preferencias de lectura de los investigadores (por ejemplo pdf frente a html).

Finalmente la metodología de visualización mediante co-uso puede fácilmente extrapolarse a otras plataformas y contextos del mundo de la web social académica. La gran mayoría de las plataformas que permiten el cálculo de indicadores altmétricos pueden ser susceptibles de ofrecer mapas de co-uso. Este es el caso de *Mendeley*, donde pueden elaborarse matrices de similaridad a partir de los artículos leídos por un mismo lector. La expansión de la red social así como el trasvase al medio digital muestran un nuevo panorama para la redefinición del concepto de uti-

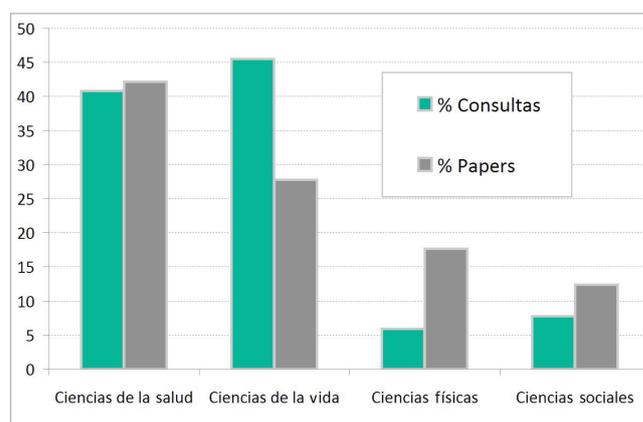


Figura 4. Comparativa entre el porcentaje de consultas realizadas a *ScienceDirect* y la producción según *Scopus* de la Universidad de Navarra en 2012

lidad de cara a la visualización de la ciencia. La utilización de nuevos indicadores ofrece nuevas vías para explorar las relaciones entre disciplinas y entre los intereses de los investigadores.

Notas

1. La información de *Elsevier bibliometric research program* está disponible en:
<http://ebrp.elsevier.com>

El proyecto en el que está basado este trabajo es *Viability of co-downloading data analysis form mapping interdisciplinary research at institutional level* cuya descripción se puede consultar en la página web anterior.

2. *Pajek* (araña, en esloveno) es un software para el análisis y visualización de redes sociales, desarrollado en la Universidad de Ljubljana, Eslovenia, por Vladimir Batagelj, Andrej Mrvar y la contribución de Matjaž Zaveršnik. Es un software libre para uso no comercial y puede descargarse de:
<http://vlado.fmf.uni-lj.si/pub/netwoks/pajek>

Los mapas basados en el uso enriquecen el análisis al incluir a otros actores pasivos en el consumo de la literatura científica como son estudiantes, profesionales o docentes

Agradecimientos

Los autores quieren agradecer a Gali Halevi y Henk Moed la cesión de los datos de uso y a Víctor Sanz Santacruz, director de la Biblioteca de la Universidad de Navarra por permitir el acceso a los mismos.

Bibliografía

Aguillo, Isidro F.; Granadino, Begoña; Ortega, José-Luis; Prieto, José-Antonio (2006). "Scientific research activity measured with cybermetrics indicators". *Journal of the American Society for Information Science and Technology*, v. 57, n. 10, pp. 1296-1306.

http://departments.icmab.es/utc/images/SEO/pdf/2006_JASIST_Webometrics.pdf

<http://dx.doi.org/10.1002/asi.20433>

Almind, Thomas; Ingwersen, Peter (1997). "Informetric analyses on the world wide web: methodological approaches to 'webometrics'". *Journal of documentation*, v. 53, n. 4, pp. 404-426.

<http://dx.doi.org/10.1108/EUM0000000007205>

Bollen, Johan; Van-de-Sompel, Hebert; Smith, Joan A.; Luce, Rick (2005). "Toward alternative metrics of journal impact: a comparison of download and citation data". *Information processing & management*, v. 41, n. 6, pp. 1419-1440.

<http://public.lanl.gov/herbertv/papers/ipm05jb-final.pdf>

<http://dx.doi.org/10.1016/j.ipm.2005.03.024>

Borrego, Ángel; Urbano, Cristóbal (2005). "Estadísticas e indicadores de rendimiento de colecciones y servicios bibliotecarios de carácter electrónico: estudio de caso de las revistas electrónicas". *El profesional de la información*, v. 14, n. 1, pp. 30-38.

<http://www.elprofesionaldeinformacion.com/contenidos/2005/enero/4.pdf>

Brody, Tim; Harnad, Stevan; Carr, Leslie (2006). "Earlier web usage statistics as predictors of later citation impact". *Journal of the American Society for Information Science and Technology*, v. 57, n. 8, pp. 1060-1072.

<http://dx.doi.org/10.1002/asi.20373>

Haustein, Stefanie; Siebenlist, Tobias (2011). "Applying social bookmarking data to evaluate journal usage". *Journal of informetrics*, v. 5, n. 3, pp. 446-457.

<http://dx.doi.org/10.1016/j.joi.2011.04.002>

Kurz, Michael J.; Bollen, Johan (2010). "Usage bibliometrics". *Annual review of information science and technology*, v. 44, pp. 1-64.

<http://dx.doi.org/10.1002/aris.2010.1440440108>

Leydesdorff, Loet (2008). "On the normalization and visualization of author co-citation data: Salton's cosine versus the Jaccard index". *Journal of the American Society for Information Science and Technology*, v. 59, n. 1, pp. 77-85.

<http://arxiv.org/abs/0911.1447>

<http://dx.doi.org/10.1002/asi.20732>

Luukkonen, Terttu; Persson, Olle; Sivertsen, Gunnar (1992). "Understanding patterns of international scientific collaboration". *Science, technology & human values*, v. 17, n. 1, pp. 101-126.

<http://dx.doi.org/10.1177/016224399201700106>

Nicholas, David; Huntington, Paul; Watkinson, Anthony (2005). "Scholarly journal usage: the results of deep log analysis". *Journal of documentation*, v. 61, n. 2, pp. 248-280.

<http://dx.doi.org/10.1108/00220410510585214>

Noyons, Christiaan M. (2004). "Science maps within a science policy context". En: Moed, Henk F.; Glänzel, Wolfgang;

Schmoch, Ulrich (eds). *Handbook of quantitative science and technology research*. Dordrecht: Kluwer Academic Publishers, pp. 237-255. ISBN: 978 1402027550

Priem, Jason; Taraborelli, Dario; Groth, Paul; Neylon, Cameron (2010). "Altmetrics: a manifesto". *Altmetrics.org*.

<http://altmetrics.org/manifesto>

Rodríguez-Bravo, Blanca; Alvite-Díez, María-Luisa; Morán-Suárez, María-Antonia; Marraud, Gerardo (2012). "Impacto de la contratación de colecciones de revistas electrónicas en la productividad de la Universidad de Vigo". *El profesional de la información*, v. 21, n. 6, pp. 585-594.

<http://dx.doi.org/10.3145/epi.2012.nov.05>

Rowlands, Ian; Nicholas, David (2007). "The missing link: journal usage metrics". *Aslib proceedings*, v. 59, n. 3, pp. 222-228.

<http://dx.doi.org/10.1108/00012530710752025>

Small, Henry (1973). "Co-citation in the scientific literature: A new measure of the relationship between two documents". *Journal of the American Society for Information Science and Technology*, v. 24, n. 4, pp. 265-269.

http://polaris.gseis.ucla.edu/gleazer/296_readings/small.pdf

<http://dx.doi.org/10.1002/asi.4630240406>

Tenopir, Carol; King, Donald W.; Christian, Lisa; Volentine, Rachel (en prensa). "Scholarly article seeking, reading and use: A continuing evolution from print to electronic". *Learned publishing*.

Thelwall, Mike (2008). "Bibliometrics to webometrics". *Journal of information science*, v. 34, n. 4, pp. 605-621.

<http://www.scit.wlv.ac.uk/~cm1993/papers/JIS-0642-v4-Bibliometrics-to-Webometrics.pdf>

<http://dx.doi.org/10.1177/0165551507087238>

Torres-Salinas, Daniel; Cabezas-Clavijo, Álvaro (2013). "Altmetrics: no todo lo que se puede contar, cuenta". *Anuario ThinkEPI*, v. 7, pp. 114-117.

Torres-Salinas, Daniel; Cabezas-Clavijo, Álvaro; Jiménez-Contreras, Evaristo (2013). "Altmetrics: nuevos indicadores para la comunicación científica en la web 2.0". *Comunicar*,

v. 41, pp. 53-60.

<http://dx.doi.org/10.3916/C41-2013-05>

Torres-Salinas, Daniel; Delgado-López-Cózar, Emilio; García-Moreno-Torres, José; Herrera, Francisco (2011). "Rankings ISI de las universidades españolas según campos científicos: descripción y resultados". *El profesional de la información*, v. 20, n.1, pp. 111-118.

<http://eprints.rclis.org/15325>

<http://dx.doi.org/10.3145/epi.2011.ene.14>

Van-Noorden, Richard (2014). "Scientists may be reaching a peak in reading habits". *Nature*.

<http://dx.doi.org/10.1038/nature.2014.14658>

ESTUDIO EVOLUTIVO DE LA INVESTIGACIÓN ESPAÑOLA CON CÉLULAS MADRE. VISUALIZACIÓN E IDENTIFICACIÓN DE LAS PRINCIPALES LÍNEAS DE INVESTIGACIÓN

Gisela Cantos-Mateos, María-Ángeles Zulueta, Benjamín Vargas-Quesada y Zaida Chinchilla-Rodríguez



Gisela Cantos-Mateos es licenciada en historia y en documentación y máster en documentación. Becaria predoctoral (programa JAE-Predoc) del *Instituto de Políticas y Bienes Públicos (IPP)* del *Consejo Superior de Investigaciones Científicas (CSIC)*. Realiza su tesis doctoral relacionada con análisis bibliométricos y con la representación y visualización de la información científica. Colabora con el *Grupo SCImago*.

<http://orcid.org/0000-0003-2690-1790>

Consejo Superior de Investigaciones Científicas. Instituto de Políticas y Bienes Públicos (CSIC-IPP)
Albasanz, 26-28. 28037 Madrid, España
gisela.cantos@csic.es



María-Ángeles Zulueta es licenciada en medicina y doctora en ciencias por la *Universidad Autónoma de Madrid*. Profesora titular del *Departamento de Filología, Comunicación y Documentación* de la *Universidad de Alcalá* e investigadora del *Grupo SCImago*. Sus áreas de investigación de interés se enmarcan en el campo de la bibliometría y análisis de dominios científicos. Ha dirigido y colaborado en varios proyectos de investigación financiados con fondos públicos en esta área y es autora de numerosas publicaciones en este campo.

<http://orcid.org/0000-0002-3983-258X>

Universidad de Alcalá. Facultad de Documentación
San Cirilo, s/n. 28804, Alcalá de Henares, Madrid, España
ma.zulueta@uah.es



Benjamín Vargas-Quesada es profesor de técnicas avanzadas de recuperación de información en la *Facultad de Comunicación y Documentación* de la *Universidad de Granada* y miembro del *Grupo SCImago*. Sus líneas de investigación se centran en la visualización de la información científica y en la evaluación de la ciencia y de la comunicación científica.

<http://orcid.org/0000-0001-5115-7460>

Universidad de Granada. Facultad de Comunicación y Documentación
Colegio Máximo de Cartuja, s/n. 18071, Granada, España
benjamin@ugr.es



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

<http://orcid.org/0000-0002-1608-4478>

Consejo Superior de Investigaciones Científicas. Instituto de Políticas y Bienes Públicos (CSIC-IPP)
Albasanz, 26-28. 28037 Madrid, España
zaida.chinchilla@csic.es

Resumen

Se analiza la investigación española en células madre en el *Science Citation Index (SCI)* entre 1997 y 2010 con una división temporal en tres sub-períodos, mediante técnicas de visualización basadas en redes sociales. Se han utilizado los descriptores *KeyWords Plus (KW+)* como unidad de análisis, su coocurrencia como unidad de medida, y *Pajek* y *VOSviewer* como herramientas para la generación y visualización de redes sociales. Los resultados muestran dos imágenes complementarias de la investigación: la estructura estática, donde se distingue investigación clínica y básica, y la dinámica del análisis evolutivo, mostrando las líneas más consolidadas y las emergentes. Se plantea una propuesta metodológica para la visualización y detección de las principales líneas de investigación a lo largo del tiempo, su aplicabilidad y capacidad de predicción en dominios científicos y geográficos.

Artículo recibido el 30-01-2014

Aceptación definitiva: 05-06-2014

Palabras clave

Análisis de *co-words*, Visualización de la información, Análisis de redes, Células madre, Tendencias de investigación, Evolución de la ciencia, España.

Title: Development of Spanish research on stem cells. Visualization and identification of the main research fronts

Abstract

Using visualization techniques based on social networks, this study aims to analyze stem cell research in Spain, as reflected in the *Science Citation Index (SCI)* database between 1997 and 2010, divided into three sub-periods. The selected unit of analysis was the *KeyWords Plus* descriptors (*KW+*), the unit of measurement was their co-occurrence, and the *Pajek* and *VOSviewer* tools were used to generate and display the social networks. The results show two complementary images of research: the static structure, distinguishing between clinical and basic research, and the evolutionary dynamic, analysing both the most established and the emerging lines. The main contribution of this work is to present a methodology for the visualization and detection of the main research lines over time, demonstrating its applicability and its predictability in scientific and geographic domains.

Keywords

Co-word analysis, Network analysis, Information visualization, Stem cells, Science trends, Research trends, Spain.

Cantos-Mateos, Gisela; Zulueta, María-Ángeles; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida (2014). "Estudio evolutivo de la investigación española con células madre. Visualización e identificación de las principales líneas de investigación". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 259-271.

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

1. Introducción

Los estudios con células madre son una de las áreas de investigación que más ha evolucionado. Desde que en 1949 el científico **John Hammond** (1949) descubriese el método para mantener los embriones de ratón en cultivo *in vitro*, el progreso en este campo ha ido en aumento, ofreciendo perspectivas muy esperanzadoras en las últimas décadas para el tratamiento de enfermedades que hasta ahora parecían incurables. La investigación en este campo está teniendo grandes repercusiones científicas y sociales, desde los experimentos como la conocida clonación de la oveja Dolly (**Wilmut et al.**, 1997) que tuvieron un especial impacto en el campo de la transferencia nuclear y la diferenciación y expresión celular en 1997, hasta la aparición entre 2006 y 2007 de las células de Yamanaka (**Takahashi; Yamanaka**, 2006; **Takahashi et al.**, 2007) que ofrecen una alternativa mediante la posibilidad de obtener células pluripotentes a partir de la reprogramación de células adultas. Al mismo tiempo se ha generado un importante debate bioético que afecta a la regulación legislativa y la financiación de la investigación con células madre embrionarias, reglamentando así su desarrollo.

Aunque es difícil establecer una relación directa entre las reformas legislativas y la evolución de la investigación, los hitos científicos podrían estar influenciados por las regulaciones legislativas. Pese a las controversias que suscita el desarrollo de determinadas líneas de investigación con células madre, en los últimos años ha habido iniciativas políticas, económicas y legislativas que manifiestan una clara intención de adecuar los hitos científicos a las estructuras sociales de nuestro tiempo. En Europa se pueden destacar importantes esfuerzos para el impulso de la investigación mediante medidas legislativas adoptadas desde 2001: la in-

centivación económica a través del *VII Programa marco* y la creación del *Registro europeo de células madre embrionarias* con el objetivo de proporcionar información a toda la comunidad científica, sobre las líneas de células disponibles en Europa. España ha ido modificando su legislación adoptando las directivas europeas hasta 2006. Estados Unidos, Reino Unido, Australia y Asia han ido legislando al respecto, y a lo largo de los años las regulaciones han sido sometidas a una continua revisión (*EuroStemCell et al.*, 2013). En cualquier caso, las normativas y la financiación de proyectos de investigación con células madre podrían afectar a los resultados de la producción científica tanto a nivel internacional como nacional. Aunque no es el objeto de este estudio, parece adecuado plantear el marco en el que se desarrolla este campo científico.

“ La idea es presentar un análisis aplicable no sólo a la investigación española con células madre, sino a otros dominios temáticos, geográficos y temporales ”

Todos estos hitos han contribuido a generar importantes resultados de esta actividad científica, susceptibles de ser analizados desde la perspectiva bibliométrica, especialmente porque esta investigación con células madre ha trascendido a otros campos científicos, siendo un tema con un importante carácter multidisciplinar. Por ello es un reto definir temáticamente el dominio y las principales líneas de investigación en las que se está trabajando.

Los estudios bibliométricos ayudan en esta tarea ya que el estudio de nuevas técnicas y herramientas ha mejorado

notablemente la comprensión y análisis de dominios científicos (Leydesdorff; Rafols, 2009; Vargas-Quesada *et al.*, 2010; Rafols; Porter; Leydesdorff, 2010; Boyack; Börner; Klavans, 2009), y especialmente en el campo de la visualización de la información (Börner; Chen; Boyack, 2003; De-Moya-Anegón *et al.*, 2006; Vargas-Quesada *et al.*, 2008). La incorporación del análisis estructural de redes sociales (Wasserman; Faust, 1998) lo ha convertido en una herramienta fundamental para visualizar y analizar las relaciones estructurales que se producen en un contexto científico determinado. En el caso de la delimitación temática ayuda a visualizar la interacción de los aspectos más destacados en el dominio, haciendo posible la identificación de áreas o líneas de investigación.

La obtención de los descriptores *KeyWords Plus* no requiere de minería textual ni de ningún pre-procesamiento lingüístico

Existe gran diversidad de métodos y técnicas a la hora de delimitar y visualizar un campo científico. Hay estudios que utilizan la cocitación de autores (*author co-citation analysis* o *ACA*) (White; Griffith, 1981; White; McCain, 1998; Small; Upham, 2009; Chen; Ibekwe-SanJuan; Hou, 2010; Zhao; Strotmann, 2011), el análisis de emparejamiento bibliográfico (*bibliographic coupling*) (Kessler, 1963; Boyack; Klavans, 2010; Chen *et al.*, 2011; Glänzel; Thijs, 2011), estudios de *co-words* (Callon *et al.*, 1983; Rip; Courtial, 1984, Cambrosio *et al.*, 1993; Van-Raan; Tijssen, 1993; Ding; Chowdhury; Foo, 2001; Lee, 2008; Leydesdorff; Welbers, 2011; Zulueta *et al.*, 2011; Cantos-Mateos *et al.*, 2013; Romo-Fernández; Guerrero-Bote; De-Moya-Anegón, 2013) o los estudios híbridos en los que se combinan tanto *ACA* como *co-words* (Zitt; Bassecouard, 1996; Zitt; Lelu; Bassecouard, 2011).

En vista de las ventajas e inconvenientes que condicionan la elección de una metodología sobre otra, para el presente estudio se ha optado por el análisis de *co-words* (Zitt; Lelu; Bassecouard, 2011). Según Michel Zitt y colaboradores (2011), las palabras, a diferencia de las citas, parecen adaptarse mejor a la “ciencia viva” porque son universales. Además las citas podría considerarse que aluden a temas menos actuales, ya que conforman el trasfondo del documento, mientras que las palabras son coetáneas al propio documento (Soos; Kampis; Gulyás, 2013). Para estos autores las unidades lingüísticas tienen la capacidad de reflejar mejor la incidencia de los contextos científicos, sociales y políticos, propios de los dominios más controvertidos y áreas emergentes.

Se puede apreciar con mayor claridad la dinámica del campo si se aplica una perspectiva temporal. Este tipo de análisis permite detectar la evolución de los principales temas de investigación a lo largo del tiempo (Braam; Moed; Van-Raan, 1991; Gábor, 2006; Leydesdorff; Schank, 2008; Boyack; Börner; Klavans, 2009; Chen; Ibekwe-SanJuan; Hou, 2010; Soos; Kampis; Gulyás, 2013). En este caso las visualizacio-

nes de redes de *co-words* de distintos sub-períodos permiten localizar qué temas persisten, desaparecen o emergen en el campo (Braam; Moed; Van-Raan, 1991).

A partir de este marco, los objetivos de este estudio son de carácter fundamentalmente metodológicos. La idea es presentar un análisis que sea aplicable no sólo a la investigación española con células madre, sino a otros dominios temáticos, geográficos y temporales, distintos al que se aborda en este trabajo. Por tanto la presente propuesta metodológica pretende por un lado identificar las principales líneas de investigación a partir de la localización y visualización de la estructura temática del dominio. Y por otro lado reflejar la dinámica del campo a través de la evolución que experimentan las principales líneas de investigación a lo largo del período de estudio. Todo ello a través de la visualización de redes de *co-words*¹.

2. Material y métodos

2.1. Fuente de información

Los documentos de este estudio proceden de la base de datos *Science Citation Index (SCI)*. La estrategia de búsqueda empleada consistió en utilizar los términos *stem* cell** en el campo *Topic*, *Spain* en el campo *Address* y limitada al período temporal de 1997-2010. Se recuperó un total de 4.148 documentos.

2.2. Unidades de análisis

Se han seleccionado los descriptores *KeyWords Plus (KW+)*² como unidades de análisis. Se trata de un vocabulario automatizado que utiliza el *SCI* para indizar los documentos, que se obtiene mediante un algoritmo de búsqueda que extrae las palabras clave de los títulos de las referencias contenidas en los documentos (Garfield, 1990; Garfield; Sher, 1993). En la investigación con células madre estos descriptores tienen una elevada representatividad (78,25% de los documentos) en comparación con los *Author keywords*³ (52,96%), como ha quedado de manifiesto en trabajos anteriores (Cantos-Mateos *et al.*, 2013).

El uso de descriptores *KeyWords Plus* permite realizar un análisis temático de granularidad fina y contiene terminología más actualizada y específica, en comparación con otros descriptores

El uso de estos descriptores tiene la ventaja de que su obtención no requiere de minería textual ni de ningún pre-procesamiento lingüístico. Permite además realizar un análisis temático de granularidad más fina que otro tipo de unidades como pueden ser las categorías, ya que aluden directamente al contenido conceptual de los documentos (Soos; Kampis; Gulyás, 2013). Por otro lado, al tratarse de términos procedentes de los títulos de los artículos que aparecen en las referencias, son los propios autores quienes seleccionan y comunican cuáles son los términos esenciales para expresar el contenido más importante de sus trabajos de investigación. Esta característica favorece la aparición de una

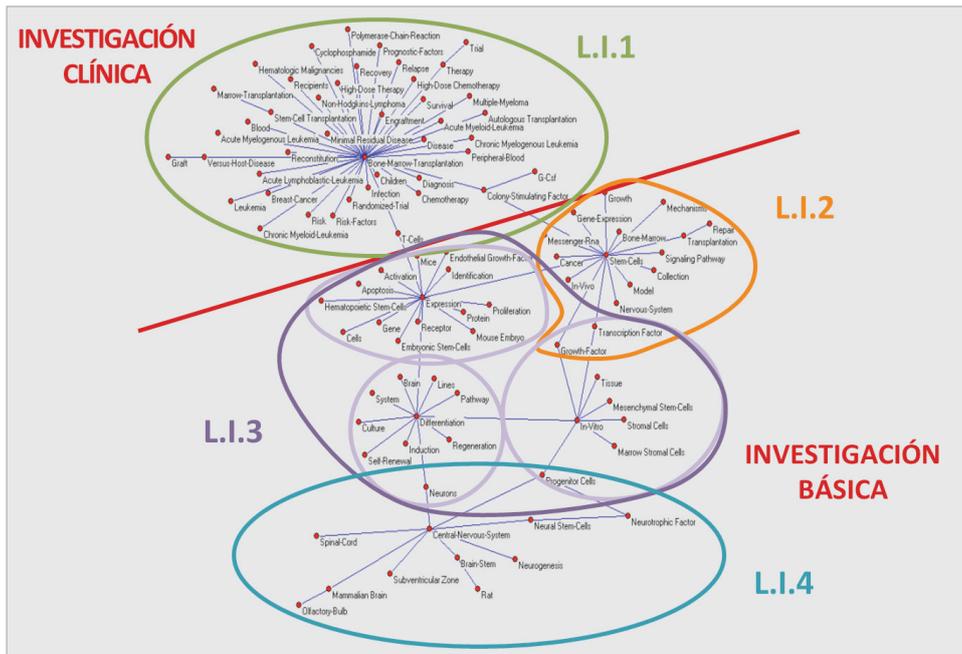


Figura 1. Período completo (1997-2010). Representación en Pajek de los 99 KW+ más frecuentes.

terminología más actualizada y específica en comparación con otros tipos de descriptores, como son los procedentes de un vocabulario controlado (Braam; Moed; Van-Raan, 1991). Como demuestran estudios recientes, estas cualidades contribuyen a reflejar mejor los aspectos temáticos más dinámicos de un dominio.

Los KW+ tienen la desventaja sin embargo de presentar una amplia dispersión terminológica debida a la falta de normalización. Este aspecto hay que resaltarlo como una de las posibles limitaciones de este trabajo. Para paliarla en cierto modo se ha realizado una normalización moderada en la que se han tratado de unificar criterios gramaticales (plurales y singulares) y ortográficos (guiones, mayúsculas y minúsculas).

2.3. Unidades de medida y análisis temporal

El análisis bibliométrico para la detección y visualización de las principales líneas de investigación y su evolución a lo largo del tiempo, ha consistido en calcular las relaciones que se producen entre los KW+ a partir de su aparición conjunta en los documentos, lo que se conoce como *co-words analysis*. Dicho análisis consiste en generar una matriz cuadrada de NxN elementos, donde N es el descriptor KW+ a representar, a partir de las veces que ocurre en los documentos. El

Tabla 1. Períodos temporales analizados

Períodos	Años
Período completo	1997-2010
Primer período	1997-2001
Segundo período	2002-2006
Tercer período	2007-2010

Tabla 2. Número de documentos por años

Período	Primer período					Segundo período					Tercer período			
Años	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Nº docs.	94	142	176	149	163	167	165	296	271	351	395	494	608	677
Total docs.	724					1.250					2.174			

resultado es una matriz de co-ocurrencias que refleja el número de veces que un par de descriptores aparecen conjuntamente en dos documentos.

Para el estudio evolutivo se ha analizado el período completo en su conjunto (1997-2010), y subdivido en otros tres períodos, como muestra la tabla 1. La segmentación temporal ha sido establecida considerando dos criterios complementarios.

El primero son los principales hitos científicos. Partiendo de 1997, cuando tuvieron lugar los exitosos ensayos del *Roslin Institute* de Edimburgo, se desencadenó un tiempo de incertidumbre en la regulación legislativa y las políticas de financiación pública que condicionaron el avance de la investigación en células madre embrionarias.

El proceso comenzó a normalizarse fundamentalmente entre los años 2002 y 2006. A partir de 2007 esta investigación recibió un importante impulso gracias al descubrimiento de las células madre de pluripotencia inducida (células iPS) a partir de células humanas adultas que planteaban una alternativa a los problemas éticos-jurídicos sobre la obtención de células madre embrionarias obtenidas a partir de la masa celular interna de un embrión, lo que suponía la destrucción del mismo (*EuroStemCell et al., 2013*).

El segundo criterio ha sido que el volumen documental de cada período sea lo suficientemente representativo como para permitir la detección de líneas de investigación.

2.4. Visualización de la información

La visualización de las relaciones entre los KW+ se ha hecho siguiendo dos métodos en base al uso de dos programas especializados en visualización y análisis de grandes redes, que permiten obtener unas representaciones que favorecen la identificación de las principales líneas de investigación y su evolución.

Por un lado se ha empleado el programa Pajek (Batagelj; Mrvar, 2010). Se utilizó una matriz de co-ocurrencias de descriptores con datos en bruto en combinación con el algoritmo de poda Pathfinder (De-Moya-Anegón et al., 2007a, 2007b; Zulueta et al., 2011) para la reducción del espacio dimensional. Este algoritmo está basado en el principio de desigualdad del triángulo bajo los parámetros $r=\infty$ y $q=n-1$. Éste ayuda a preservar y poner de relieve las relaciones más

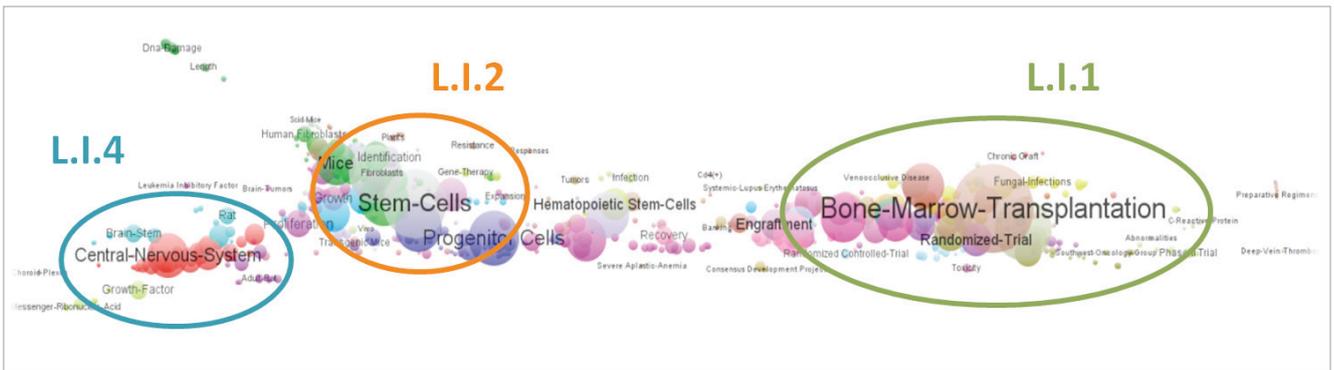


Figura 2. Período completo (1997-2010). Mapa de clusters representado con VOSviewer (número de nodos: 668).

destacadas entre los descriptores (White, 2003; Vargas-Quesada et al., 2008; De-Moya-Anegón et al., 2009). Como algoritmo de visualización se utilizó Kamada-Kawai (Kamada; Kawai, 1989) incluido en Pajek. Aunque este programa permite representar grandes redes, se decidió incorporar únicamente los descriptores más frecuentes. Tras varios ensayos donde se probaron distintos umbrales de frecuencias para la generación de los mapas, se observó que el mejor compromiso entre visualización y análisis, teniendo en cuenta que se trabaja con las limitaciones propias de la pantalla de un ordenador, es el de redes compuestas en torno a los 100 nodos (Zulueta et al., 2011).

Por otro lado se ha utilizado el programa VOSviewer v. 1.5.3 (Van-Eck; Waltman, 2010), que es una alternativa a las representaciones multidimensionales combinando técnicas de visualización y clustering. En este caso se ha construido una matriz de co-ocurrencia normalizada a través de la medida de similitud que utiliza VOSviewer, basada en la fuerza de asociación de los ítems. El umbral de dicha co-ocurrencia ha sido fijo para todas las representaciones. De los cuatro tipos de visualizaciones que ofrece, se han utilizado los mapas etiquetados con nombres, donde cada descriptor está representado por una etiqueta y cuyo tamaño es proporcional a su peso, junto con los denominados “mapas de calor” (o mapas de densidad). Estos últimos se caracterizan porque cada nodo en el mapa se representa con un color que va del rojo al azul, reflejando la densidad de las relaciones entre los descriptores. Cuanto mayor es la densidad, es decir la co-ocurrencia entre descriptores, más se aproximarán a la tonalidad roja; en cambio la mayor aproximación al azul indica una mayor dispersión y por lo tanto menos co-ocurrencia.

Tabla 3. Datos relativos al período completo (1997-2010)

Período completo (1997-2010)			
Documentos	Totales	4.148	
	Indizados con KW+	3.246 (78,25%)	
KW+	Totales	28.733	
	Únicos	Sin normalización	9.833
		Con normalización	9.465
	Por documento	8,85	
	Frecuencia media de aparición	3,04	

3. Resultados y discusión

El número de documentos recuperados fue de 4.148 para el período 1997-2010. La tabla 2 muestra el volumen de producción por tramos temporales.

3.1. Período temporal completo (1997-2010)

De los 4.148 documentos recuperados para todo el período, el 78,25% están indizados con KW+, contabilizándose 28.733 descriptores, de los cuales 9.465 son únicos (tabla 3).

En la figura 1 se aprecian hasta 6 agrupaciones de descriptores KW+ que, por su significado y relación con los demás, podemos considerar que definen líneas de investigación. Se distinguen estructuralmente dos naturalezas temáticas de la investigación claramente diferenciadas. Por un lado la investigación clínica representada por los descriptores agrupados bajo Bone-Marrow-Transplantation (figuras 1 y 3), relacionados con los estudios sobre el uso terapéutico de

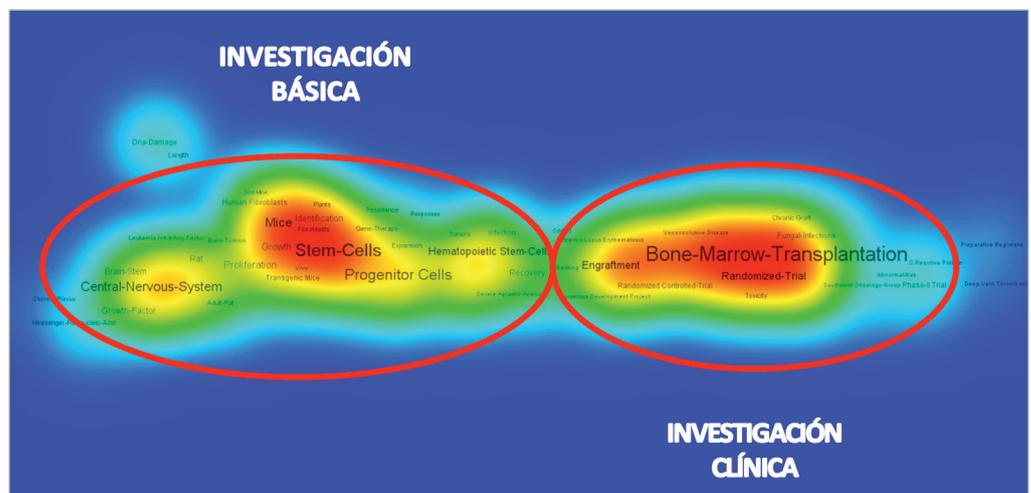


Figura 3. Período completo (1997-2010). Mapa de calor representado con VOSviewer.

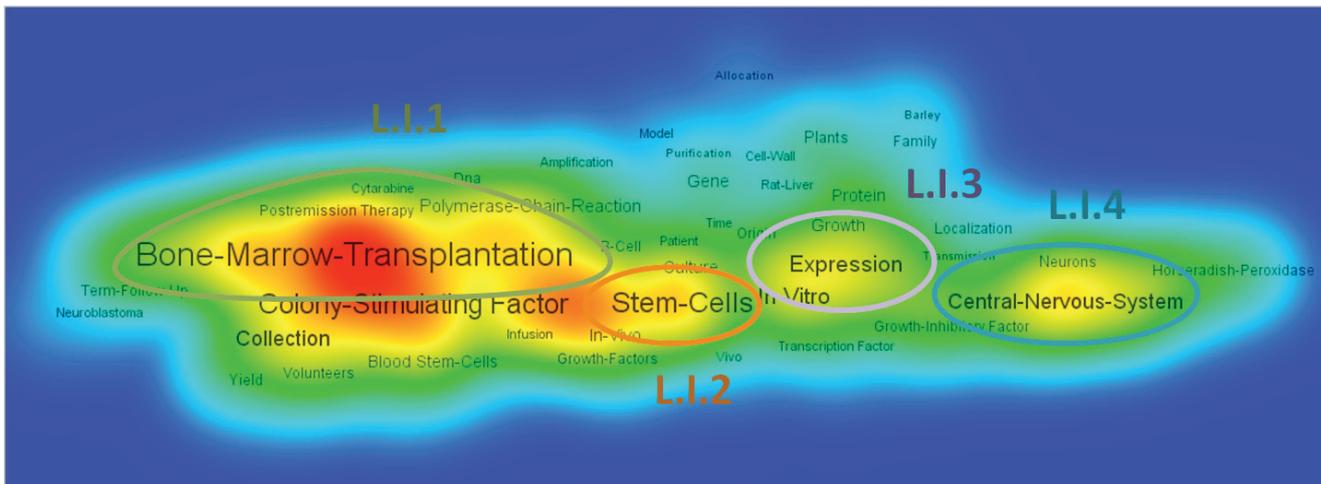


Figura 6. Primer período (1997-2001). Mapa de calor representado con VOSviewer.

3.2. Primer período (1997-2001)

El número de documentos es de 724 y el de descriptores de 2.080 KW+ únicos (tabla 4).

La visualización del dominio ha cambiado sustancialmente. A pesar de ello, la distinción entre la investigación clínica y la básica sigue siendo muy clara, incluso más, que con respecto al período completo.

En la figura 4 se aprecia un grupo muy numeroso de KW+ relacionados con la investigación clínica. La composición de los descriptores de esta línea (L.I.1) es muy similar a la que se ha descrito en el período completo. Su clara definición refleja que estos estudios están muy consolidados.

El resto de las líneas que integraban la investigación de carácter más básico (L.I.2, L.I.3 y L.I.4) no parecen estar tan definidas. Según la figura 4, es posible detectar la mayor parte de los descriptores presentes en el período completo, incluso los que ejercían mayor influencia dentro de cada línea de investigación. A pesar de ello, el comportamiento relacional entre ellos es distinto en este primer período. La mayoría de relaciones parten del descriptor Stem Cells, constituyendo el nodo principal. También se aprecia en la figura 5, donde constituye uno de los cuatro clusters que mejor se diferencian en el mapa. De manera más específica y poco definida, se detectan dos agrupaciones que mantienen cierta identidad:

- procesos que intervienen en la proliferación de líneas celulares (figura 5, clúster verde);
- relacionados con células madre neurales (figura 5, clúster rojo).

Lo mismo se puede localizar

Tabla 5. Datos del segundo período (2002-2006)

Segundo período (2002-2006)			
Documentos	Totales	1.250	
	Indizados con KW+	979 (78,32%)	
KW+	Totales	8.581	
	Únicos	Sin normalización	3.966
		Con normalización	3.852
	Por documento	8,76	
Frecuencia media de aparición	2,23		

en el mapa de densidades de la figura 6. Estas débiles agrupaciones forman líneas de investigación emergentes, como se muestra en los siguientes períodos.

3.3. Segundo período (2002-2006)

El número de documentos asciende a 1.250. El número de KW+ únicos también se incrementa, hasta 3.852 descriptores.

Se mantiene la distinción entre investigación clínica y básica. En la figura 9 se aprecia que esos son los dos focos más importantes, aunque la densidad es mucho mayor en la investigación clínica.

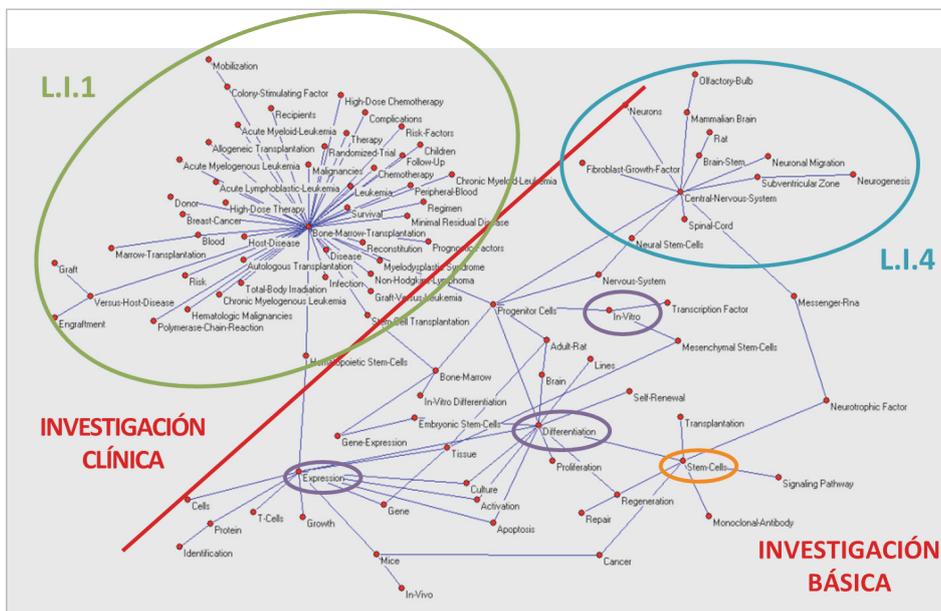


Figura 7. Segundo período (2002-2006). Representación en Pajek de los 96 KW+ más frecuentes.

en enfermedades hematológicas (L.I.1) y con la investigación en células madre embrionarias (L.I.3). En este período aparecen mejor definidos los estudios sobre procesos de proliferación y diferenciación celular y los relacionados con los procesos y las biotecnologías necesarias para localizar, producir, crecer y sobre todo, analizar in-vitro las células madre embrionarias (figura 10, zona inferior; figura 11, clúster azul oscuro).

La última línea de investigación (L.I.4) la conforman los estudios relacionados con

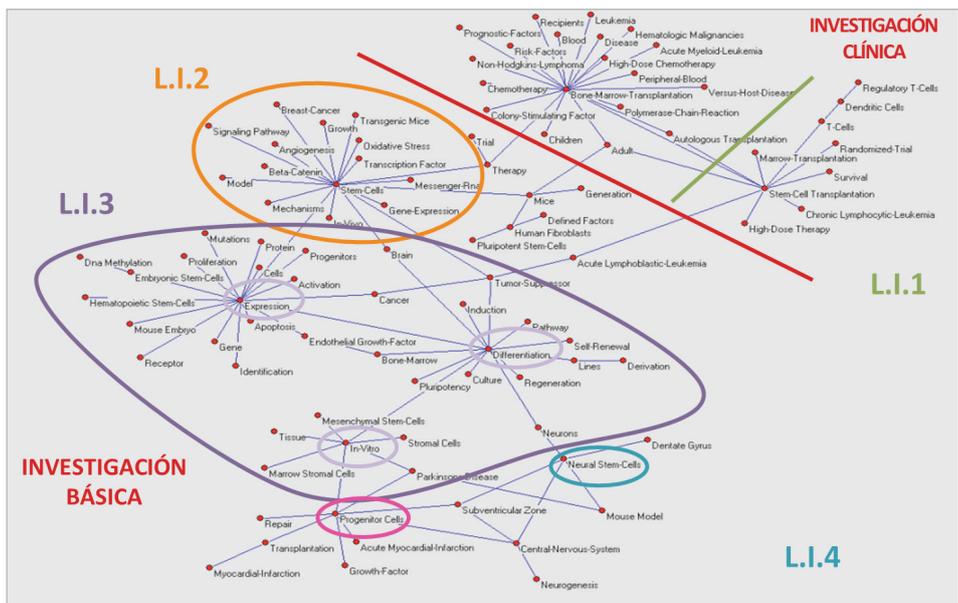


Figura 10. Tercer período (2007-2010). Representación en Pajek de los 98 KW+ más frecuentes.

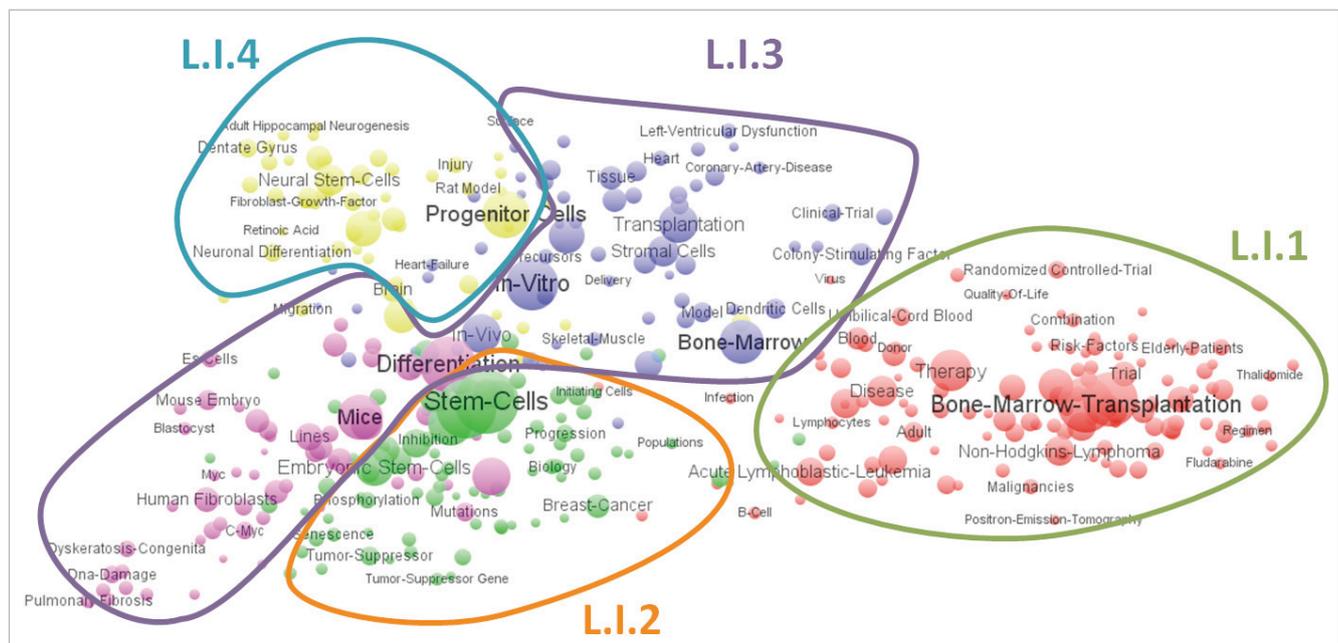


Figura 11. Tercer período (2007-2010). Mapa de clusters representado con VOSviewer (número de nodos: 390).

las células madre neurales (figura 11, clúster amarillo). En este período los KW+ han perdido definición y ya no constituyen un grupo tan homogéneo. Incluso se han desagregado en dos aspectos de la investigación: el desarrollo e identificación de las células madre neurales y el infarto de miocardio empleando progenitores celulares (figura 10). Es posible que esta pérdida de definición se deba a



Figura 12. Tercer período (2007-2010). Mapa de calor representado con VOSviewer.

una nueva línea de investigación emergente que pueda consolidarse con los años.

El estudio del período completo identifica las 4 líneas de investigación más consolidadas: uso terapéutico de células madre en enfermedades hematológicas; células madre hematopoyéticas; embrionarias; y neurales

4. Conclusiones

La evolución de la investigación española y su crecimiento están en consonancia con la que se ha llevado a cabo a nivel internacional, tanto en el crecimiento de la literatura a lo largo del período, como su fuerte expansión en el último período analizado. También las líneas de investigación detectadas en las visualizaciones obtenidas en este estudio están relacionadas con la evolución científica internacional, especialmente en la investigación en células madre embrionarias (*EuroStemCell et al.*, 2013).

El estudio del período completo muestra la estructura estática del dominio con la identificación de las 4 líneas de investigación más consolidadas:

- uso terapéutico de células madre en enfermedades oncohematológicas (L.I.1);
- células madre hematopoyéticas (L.I.2);
- células madre embrionarias (L.I.3);
- células madre neuronales (L.I.4).

A través del análisis evolutivo se aprecia cómo se han configurado estas líneas a partir de la consolidación de algunas emergentes y la pérdida de protagonismo de otras. En el primer período hay una fuerte presencia de la hematología y la oncología. El resto de las líneas no comenzaron a detectarse hasta el segundo período con la investigación sobre células madre hematopoyéticas. Finalmente los resultados han mostrado que el peso de estas últimas agrupaciones tuvo una presencia mucho más evidente en el último período, en detrimento de las líneas de investigación relacionadas con la aplicación y la clínica, a diferencia de lo que ocurría en el primer y segundo períodos.

Con estos resultados no es posible concluir que exista una relación directa entre los hitos científicos y normativos acontecidos en España entre los años 1997 y 2010. Sin embargo, parece que las tendencias de investigación están en sintonía con la evolución del marco jurídico-legislativo español y con las principales acciones de incentiación científica en el campo de las células madre. Aunque no es el objeto de estudio, sí parece interesante plantear como estudios futuros un análisis que examine si hay una relación causa-efecto entre los hitos normativos y los científicos.

La propuesta metodológica basada en la visualización de las relaciones entre los KW+ ha permitido obtener dos imágenes complementarias de la investigación española sobre cé-

lulas madre, proyectando tanto líneas de investigación claramente consolidadas como emergentes. Esto permite intuir metodológicamente la capacidad de predicción de este tipo de estudios en el desarrollo de un dominio temático a lo largo del tiempo. Su aplicabilidad trasciende a otros dominios científicos. La rigidez de las clasificaciones documentales utilizadas por las bases de datos, fundamentalmente de revistas científicas, que proporcionan un nivel de agregación de los dominios científicos excesivamente amplios y poco definidos se vería complementada por este tipo de análisis al permitir detectar un mayor grado de profundidad de los diferentes dominios científicos.

En cuanto a los programas de visualización utilizados, las imágenes ofrecidas por *Pajek* y *VOSviewer* coinciden en las líneas de investigación y en la identificación de los descriptores más influyentes de la red. Además se han complementado bien: *Pajek* en combinación con *PfNET* ha permitido sintetizar y resaltar la estructura relacional, mientras que *VOSviewer* ha hecho posible la identificación de los grupos y focos de investigación de una manera más plástica. Las metodologías de visualización empleadas en este estudio presentan visualizaciones estáticas e independientes de cada período.

Tanto en las visualizaciones del período completo como en la de los sub-períodos se distingue claramente la investigación básica y la clínica

De cara a estudios futuros sería interesante explorar las últimas funciones incorporadas a los dos programas: mapas superpuestos (*overlay maps*) que permiten fijar la posición de los nodos para hacer estudios evolutivos o comparativos (**Rafols; Porter; Leydesdorff, 2010; Leydesdorff; Rafols; Chen, 2013**), así como el uso de otros softwares de visualización como *Vison* utilizado en estudios similares (**Leydesdorff et al., 2008; Leydesdorff; Schank, 2008; Baur; Schank, 2008**). También sería interesante contrastar esta propuesta metodológica empleando otras unidades de análisis distintas a los KW+ con el objetivo de identificar el mejor sistema para definir las líneas de investigación presentes en cada dominio.

5. Notas

1. Un avance de los resultados de este trabajo se presentó en el *VI Encontro Ibérico Edicic 2013; Globalização, ciência, informação*, celebrado en Oporto (Portugal), 4-6 nov. 2013.

2. *KeyWords Plus (KW+)*
<http://interest.science.thomsonreuters.com/content/WOKUserTips-201010-SEA>
<http://wokinfor.com/media/pdf/qrc/wosqrc.pdf>

3. Los *Author keywords* son el otro tipo que utiliza *SCI* para indizar los documentos. Éstos se descartaron como unidades de análisis para el estudio, por su baja representatividad a partir de los resultados del estudio que se menciona.

6. Agradecimientos

Este trabajo ha sido posible gracias a la financiación pública obtenida mediante la beca pre doctoral del programa JAE-Predoc concedida por el Consejo Superior de Investigaciones Científicas (CSIC) y desarrollada en el Instituto de Políticas y Bienes Públicos (IPP).

7. Bibliografía

Batagelj, Vladimir; Mrvar, Andrej (2010). *Pajek 2.0: package for large network analysis*.

<http://vlado.fmf.uni-lj.si/pub/networks/pajek>

Baur, Michael; Schank, Thomas (2008). *Dynamic graph drawing in visone*. Karlsruhe: Technical University Karlsruhe <http://i11www.itl.uni-karlsruhe.de/extra/publications/bs-dgdv-08.pdf>

Börner, Katy; Chen, Chaomei; Boyack, Kevin W. (2003). "Visualizing knowledge domains". *Annual review of information science and technology*, v. 37, n. 1, pp. 179–255.

<http://nwb.cns.iu.edu/papers/arist02.pdf>

<http://dx.doi.org/10.1002/aris.1440370106>

Boyack, Kevin W.; Börner, Katy; Klavans, Richard (2009). "Mapping the structure and evolution of chemistry research". *Scientometrics*, v. 79, n. 1, pp. 45-60.

<http://cns.iu.edu/images/pub/2007-boyack-mapchem.pdf>

<http://dx.doi.org/10.1007/s11192-009-0403-5>

Boyack, Kevin W.; Klavans, Richard (2010). "Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?". *Journal of the American Society for Information Science and Technology*, v. 61, n. 12, pp. 2389-2404.

<http://dx.doi.org/10.1002/asi.21419>

Braam, Robert R.; Moed, Henk F.; Van-Raan, Anthony F. J. (1991). "Mapping of science by combined co-citation and word analysis. II: Dynamical aspects". *Journal of the American Society for Information Science*, v. 42, n. 4, pp. 252-266.

<http://www.cwts.nl/TvR/documents/AvR-CoCit-Word-II.pdf>

[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4%3C252::AID-ASIZ3E3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1097-4571(199105)42:4%3C252::AID-ASIZ3E3.0.CO;2-G)

Callon, Michell; Courtail, Jean-Pierre; Turner, William A.; Bauin, Serge (1983). "From translations to problematic networks: an introduction to co-word analysis". *Social science information sur les sciences sociales*, v. 22, n. 2, pp. 191-235.

<http://dx.doi.org/10.1177/053901883022002003>

Cambrosio, Alberto; Limoges, Camille; Courtail, Jean-Pierre; Laville, Françoise (1993). "Historical scientometrics? Mapping over 70 years of biological safety research with cword analysis". *Scientometrics*, v. 27, n. 2, pp. 119-143.

<http://dx.doi.org/10.1007/BF02016546>

Cantos-Mateos, Gisela; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Zulueta, María-Ángeles (2012). "Stem cell research: bibliometric analysis of main research areas through KeyWords plus". *Aslib proceedings*, v. 64, n. 6, pp. 561-590.

<http://digital.csic.es/handle/10261/62888?locale=es>

<http://dx.doi.org/10.1108/00012531211281698>

Cantos-Mateos, Gisela; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Zulueta, María-Ángeles (2013).

"Estudio comparativo sobre la visualización de redes de co-words a través de los descriptores del *Science citation index* y *Medline*". En: *I Congresso ISKO Espanha e Portugal / XI Congresso ISKO Espanha*, Oporto (Portugal), 7-9 nov. 2013, pp. 173-189.

<http://digital.csic.es/handle/10261/89290>

<http://dx.doi.org/10.1080/00387019808003304>

Chen, Chaomei; Ibekwe-SanJuan, Fidelia; Hou, Jianhua (2010). "The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis". *Journal of the American Society for Information Science and Technology*, v. 61, n. 7, pp. 1386-1409.

<http://arxiv.org/pdf/1002.1985.pdf>

<http://dx.doi.org/10.1002/asi.21309>

Chen, Dar-Zen; Huang, Mu-Hsuan; Hsieh, Hui-Chen; Lin, Chang-Pin (2011). "Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology". *Journal of informetrics*, v. 5, n. 3, pp. 400-412.

<http://dx.doi.org/10.1016/j.joi.2011.02.005>

De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Gómez-Crisóstomo, Rocío; González-Molina, Antonio; Muñoz-Fernández, Francisco-José; Vargas-Quesada, Benjamín (2007a). *Indicadores bibliométricos de la actividad científica española: 1990-2004*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt). ISBN: 84 690 5059 0

<http://eprints.rclis.org/16646>

De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Gómez-Crisóstomo, Rocío; Hassan-Montero, Yusef; Vargas-Quesada, Benjamín (2009). *Indicadores bibliométricos de la actividad científica española: 2007*. Madrid: Fundación Española para la Ciencia y la Tecnología (Fecyt).

http://icono.fecyt.es/informesypublicaciones/Documents/indicadores_bibliometricos_2007.pdf

De-Moya-Anegón, Félix; Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco José; Herrero-Solana, Víctor (2007b).

"Visualizing the marrow of science". *Journal of the American Society for Information Science and Technology*, v. 58, n. 14, pp. 2167-79.

<http://digital.csic.es/handle/10261/78646>

<http://dx.doi.org/10.1002/asi.20683>

De-Moya-Anegón, Félix; Vargas-Quesada, Benjamín; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco-José; Herrero-Solana, Víctor; González-Molina, Antonio; Chinchilla-Rodríguez, Zaida (2006). "Visualización y análisis de la estructura científica española: *ISI Web of Science* 1990-2005". *El profesional de la información*, v. 15, n. 4, pp. 258-269.

<http://eprints.rclis.org/8632>

Ding, Ying; Chowdhury, Gobinda G.; Foo, Schubert (2001). "Bibliometric cartography of information retrieval research by using co-word analysis". *Information processing & management*, v. 37, n. 6, pp. 67-78.

<http://core.kmi.open.ac.uk/display/9015701>

[http://dx.doi.org/10.1016/S0306-4573\(00\)00051-0](http://dx.doi.org/10.1016/S0306-4573(00)00051-0)

EuroStemCell, iCeMS, Elsevier (2013). "Stem cell research. Trends and perspectives on the evolving international landscape". http://www.eurostemcell.org/files/Stem-Cell-Report-Trends-and-Perspectives-on-the-Evolving-International-Landscape_Dec2013.pdf

Gábor, Csárdi (2006). "Dynamics of citation networks". En: *Artificial neural networks - Icnan 2006, 16th Intl conf*, Athens, Greece, Sept. 10-14, pp. 698-709.

Garfield, Eugene (1990). "Keywords Plus-ISI's breakthrough retrieval method. Part 1. Expanding your searching power on *Current contents* on diskette". *Current contents*, v. I, n. 32, pp. 5-9. <http://www.garfield.library.upenn.edu/essays/v13p295y1990.pdf>
<http://dx.doi.org/10.1007/s00338-008-0431-2>

Garfield, Eugene; Sher, Irving H. (1993). "Keywords Plus™ – Algorithmic derivative indexing". *Journal of the American Society for Information Science*, v. 44, n. 5, pp. 298-299. [http://www.garfield.library.upenn.edu/papers/jasis44\(5\)p298y1993.html](http://www.garfield.library.upenn.edu/papers/jasis44(5)p298y1993.html)
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199306\)44:5%3C298::AID-ASI5%3E3.0.CO;2-A](http://dx.doi.org/10.1002/(SICI)1097-4571(199306)44:5%3C298::AID-ASI5%3E3.0.CO;2-A)

Glänzel, Wolfgang; Thijs, Bart (2011). "Using 'core documents' for the representation of clusters and topics". *Scientometrics*, July, v. 88, n. 1, pp. 297-309. <http://dx.doi.org/10.1007/s11192-011-0347-4>

Hammond, John (1949). "Recovery and culture of tubal mouse ova". *Nature*, v. 163, n. 4131, pp. 28-29. <http://dx.doi.org/10.1038/163028b0>

Kamada, Tomihisa; Kawai, Satoru (1989). "An algorithm for drawing general undirected graphs". *Information processing letters*, v. 31, n. 1, April, 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)

Kessler, Meyer-Mike (1963). "Bibliographic coupling between scientific papers". *American documentation*, v. 14, n. 1, pp. 10-25. <http://dx.doi.org/10.1002/asi.5090140103>

Lee, Woo-Hyoung (2008). "How to identify emerging research fields using scientometrics: an example in the field of information security". *Scientometrics*, v. 76, n. 3, pp. 503-525. <http://dx.doi.org/10.1007/s11192-007-1898-2>

Leydesdorff, Loet; Rafols, Ismael (2009). "A global map of science based on the *ISI* subject categories". *Journal of the American Society for Information Science and Technology*, v. 60, n. 2, pp. 348-362. <http://arxiv.org/pdf/0911.1057.pdf>
<http://dx.doi.org/10.1002/asi.20967>

Leydesdorff, Loet; Rafols, Ismael; Chen, Chaomei (2013). "Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations". *Journal of the American Society for Information Science and Technology*, v. 64, n. 12, pp. 2573-2586. <http://arxiv.org/pdf/1301.1013.pdf>
<http://dx.doi.org/10.1002/asi.22946>

Leydesdorff, Loet; Schank, Thomas (2008). "Dynamic animations of journal maps: indicators of structural change and interdisciplinary developments". *Journal of the American Society for Information Science and Technology*, v. 59, n. 11, pp. 1810-1818. <http://arxiv.org/pdf/0911.1437v1.pdf>

Leydesdorff, Loet; Schank, Thomas; Scharnhorst, Andrea; De Nooy, Wouter (2008). "Animating the development of social networks over time using a dynamic extension of multidimensional scaling". *El profesional de la información*, nov.-dic., v. 17, n. 6, pp. 611-626. <http://www.elprofesionalde lainformacion.com/contenidos/2008/noviembre/04.pdf>
<http://dx.doi.org/10.3145/epi.2008.nov.04>

Leydesdorff, Loet; Welbers, Kasper (2011). "The semantic mapping of words and co-words in contexts". *Journal of informetrics*, July, v. 5, n. 3, pp. 469-475. <http://arxiv.org/pdf/1011.5209v2.pdf>
<http://dx.doi.org/10.1016/j.joi.2011.01.008>

Rafols, Ismael; Porter, Alan L.; Leydesdorff, Loet (2010). "Science overlay maps: a new tool for research policy and library management". *Journal of the American Society for Information Science and Technology*, v. 61, n. 9, pp. 1871-1887. <http://www.leydesdorff.net/overlaytoolkit/overlaytoolkit.pdf>
<http://dx.doi.org/10.1002/asi.21368>

Rip, Arie; Courtial, Jean-Pierre (1984). "Co-word maps of biotechnology: An example of cognitive scientometrics". *Scientometrics*, v. 6, n. 6, pp. 381-400. <http://doc.utwente.nl/57102/1/co-word.pdf>
<http://dx.doi.org/10.1007/BF02025827>

Romo-Fernández, Luz M.; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix (2013). "Co-word based thematic analysis of renewable energy (1990-2010)". *Scientometrics*, v. 97, n. 3, pp. 743-765. <http://dx.doi.org/10.1007/s11192-013-1009-5>

Small, Henry; Upham, Phineas (2009). "Citation structure of an emerging research area on the verge of application". *Scientometrics*, v. 79, n. 2, pp. 365-375. <http://dx.doi.org/10.1007/s11192-009-0424-0>

Soos, Sandor; Kampis, George; Gulyás, László (2013). "Large-scale temporal analysis of computer and information science". *The European physical journal special topics*, v. 222, n. 6, pp. 1441-1465. <http://dx.doi.org/10.1140/epjst/e2013-01936-6>

Takahashi, Kazutoshi; Tanabe, Koji; Ohnuki, Mari; Narita, Megumi; Ichisaka, Tomoko; Tomoda, Kiichiro; Yamanaka, Shinya (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors". *Cell*, v. 131 n. 5, pp. 861-872. <http://images.cell.com/images/Edimages/Cell/IEPs/3661.pdf>
<http://dx.doi.org/10.1016/j.cell.2007.11.019>

Takahashi, Kazutoshi; Yamanaka, Shinya (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors". *Cell*, v. 126, n. 4, pp. 663-676. [http://www.cell.com/cell/fulltext/S0092-8674\(06\)00976-7](http://www.cell.com/cell/fulltext/S0092-8674(06)00976-7)

<http://dx.doi.org/10.1016/j.cell.2006.07.024>

Van-Eck, Nees-Jan; Waltman, Ludo (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, v. 84, n. 2, pp. 523-38.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2883932>
<http://dx.doi.org/10.1007/s11192-009-0146-3>

Van-Raan, Anthony F. J.; Tijssen, Robert J. W. (1993). "The neural net of neural network research. An exercise in bibliometric mapping". *Scientometrics*, v. 26, n. 1, pp. 169-192.
<http://dx.doi.org/10.1007/BF02016799>

Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; González-Molina, Antonio; De-Moya-Anegón, Félix (2010). "Showing the essential science structure of a scientific domain and its evolution". *Information visualization*, v. 9, n. 4, pp. 288-300.
<http://www.ugr.es/~benjamin/showing.pdf>
<http://dx.doi.org/10.1057/ivs.2009.33>

Vargas-Quesada, Benjamín; De-Moya-Anegón, Félix; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Guerrero-Bote, Vicente (2008). "Development of the Spanish scientific landscape: ISI Web of science 1990-2005". *El profesional de la información*, v. 17, n. 1, pp. 22-37.
http://eprints.rclis.org/16683/1/evolucion_estructura_cientifica_espa%C3%B1ola.pdf
<http://dx.doi.org/10.3145/epi.2008.ene.03>

Wasserman, Stanley; Faust, Katherine (1998). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press. ISBN: 97 805 1181 5478.
<http://dx.doi.org/10.1017/CBO9780511815478>

White, Howard D. (2003). "Pathfinder networks and author co-citation analysis: a remapping of paradigmatic information scientists". *Journal of the American Society for Information Science and Technology*, v. 54, n. 5, pp. 423-34.
<http://dx.doi.org/10.1002/asi.10228>

White, Howard D.; Griffith, Belder C. (1981). "Author co-citation: a literature measure of intellectual structure". *Journal of the American Society for Information Science*, mayo,

v. 32, n.3, pp. 163-171.

<http://dx.doi.org/10.1002/asi.4630320302>

White, Howard D.; McCain, Katherine W. (1998). "Visualizing a discipline: an author co-citation analysis of information science, 1972-1995". *Journal of the American Society for Information Science*, v. 49, n. 4, pp. 327-375.
<http://comminfo.rutgers.edu/~kantor/601/Readings2004/Week2/w2r2.PDF>
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(19980401\)49:4%3C327::AID-ASIA4%3E3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1097-4571(19980401)49:4%3C327::AID-ASIA4%3E3.0.CO;2-4)

Wilmut, Ian; Schnieke, Angelika E.; McWhir, Jim; Kind, Alexander J.; Campbell, Keith H. S. (1997). "Viable offspring derived from fetal and adult mammalian cells". *Nature*, v. 385, n. 6619, pp. 810-813.
<http://dx.doi.org/10.1038/385810a0>

Zhao, Dangzhi; Strotmann, Andreas (2011). "Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field". *Scientometrics*, v. 87, n. 1, pp. 15-131.
<http://dx.doi.org/10.1007/s11192-010-0317-2>

Zitt, Michel; Bassecouard, Elise (1996). "Reassessment of co-citation methods for science indicators: effect of methods improving recall rates". *Scientometrics*, v. 37, n. 2, pp. 223-244.
<http://dx.doi.org/10.1007/BF02093622>

Zitt, Michel; Lelu, Alain; Bassecouard, Elise (2011). "Hybrid citation-word representations in science mapping: Portolan charts or research fields?" *Journal of the American Society for Information Science and Technology*, v. 62, n. 1, pp. 19-39.
<http://dx.doi.org/10.1002/asi.21440>

Zulueta, María-Ángeles; Cantos-Mateos, Gisela; Sánchez, Carmen; Vargas-Quesada, Benjamín (2011). "Research involving women and health in the Medline database, 1965-2005. Co-term analysis and visualization of main lines of research". *Scientometrics*, v. 88, n. 3, pp. 679-706.
<http://dx.doi.org/10.1007/s11192-011-0455-1>

Máster Universitario en
Gestión de Información en las Organizaciones
Facultad de Comunicación y Documentación de la Universidad de Murcia
Grupo semipresencial en la Universidad Autónoma Metropolitana de México Unidad Xochimilco



Preinscripción 2014/2015:
• 1ª fase: 18 de febrero - 2 de mayo de 2014
• 2ª fase: 1 - 18 de julio de 2014
• 3ª fase: 15 - 26 de septiembre de 2014

Más información:
<http://www.um.es/web/comunicacion>

UNIVERSIDAD DE MURCIA



GRAPHICAL INTERFACE OF THE SCIMAGO JOURNAL AND COUNTRY RANK: AN INTERACTIVE APPROACH TO ACCESSING BIBLIOMETRIC INFORMATION

Yusef Hassan-Montero, Vicente P. Guerrero-Bote y Félix De-Moya-Anegón

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



Yusef Hassan-Montero has a PhD in Documentation from the *University of Granada* and works for *SCImago Lab* in the design and development of interfaces for scientific visualization. He directs *nosolousabilidad.com* online magazine, is co-organizer of *UXSpain*, the most important national event on user experience in Spain, and teaches about interaction design in graduate courses of the *Open University of Catalonia (UOC)* and *Pompeu Fabra University*.
<http://orcid.org/0000-0002-4314-1175>

Churriana de la Vega, 18194 Granada, Spain
yusefhassan@gmail.com



Vicente P. Guerrero-Bote received a Ph.D. in Information Sciences in the *University of Granada*; a B.Sc. in Physics in the *Complutense University of Madrid*; is university professor of Research Evaluation and Network Analysis at the *University of Extremadura*, where he is the head of the *Department of Information and Communication*. He is a member of *SCImago*, *Research Group* dedicated to Scientometrics, being one of the developers of the *SCImago Journal Rank*.
<http://orcid.org/0000-0003-4821-9768>

Facultad de Ciencias de la Documentación y la Comunicación, Universidad de Extremadura
Pl. Ibn Marwan, s/n. 06001 Badajoz, Spain
guerrero@unex.es



Félix De-Moya-Anegón, Ph.D. in History from the *University of Granada*, is a research professor in the *Institute of Public Goods and Policies* of the *Higher Council for Scientific Research (CSIC)* in Madrid, and principal researcher of the associated unit *SCImago Research Group*. He has directed the R&D projects *SCImago Journal & Country Rank (SJR)*, *SCImago Institutions Rankings (SIR)*, *Atlas of Science*, etc.
<http://orcid.org/0000-0002-0255-8628>

Centro de Ciencias Humanas y Sociales (CCHS), CSIC
Albasanz, 26-28, 28037 Madrid, Spain
SCImago Research Group
felix.moya@scimago.es

Abstract

A graphical interface designed to access the bibliometric indicators database of the *SCImago Journal & Country Rank* portal (<http://scimagojr.com/shapeofscience>) is described. The map was generated from the relational matrix based on the citation, co-citation, and bibliographic coupling formed by the nearly 20,000 publications (journals and conference proceedings) registered in *Scopus*. The map layout is based on a variant of a force-directed algorithm using Noack's proposal of an edge-repulsion energy model. The interface allows the publications' bibliometric indicators and the cluster structures that they form to be displayed based on their shared use by the authors of the documents. To facilitate navigation, the interface automatically positions the reference areas and subject categories which are viewable via zoom-and-pan. The interface may be found to constitute a useful tool for analyses of the *Scopus* publications' presence in different scientific domains, and of the global distribution of the publishing capacity of different countries and regions. It uses the method of overlaying maps to locate subsets of selected publications in the context of the global publication structure.

Keywords

Maps of scientific publications, Graphical interfaces, Bibliometric information, Science maps, Science analysis tools, Citations, *Scopus*, *SCImago*.

Título: Interfaz gráfica del SCImago Journal & Country Rank: Un enfoque interactivo para acceder a la información bibliométrica

Article received on: 20-05-2014

Accepted on: 28-05-2014

Resumen

Diseño de una interfaz visual para el acceso a la base de datos de indicadores bibliométricos del portal *SCImago Journal & Country Rank*. El mapa se genera a partir de la matriz relacional basada en citación, cocitación y *bibliographic coupling* que forman las casi 20.000 publicaciones (revistas y actas de congreso) registradas en *Scopus*. El *layout* del mapa se basa en una variante de un algoritmo *force-directed* mediante el modelo de energía *edge-repulsion* propuesto por Noack. La interfaz permite la visualización de los indicadores bibliométricos de las publicaciones y la estructura de clusters que forman en función de su uso compartido por los autores de los trabajos. Para facilitar la navegación la interfaz incluye la referencia de áreas y categorías temáticas mediante posicionamiento automático y visibles mediante *zoom+pan*. La interfaz puede ser una herramienta útil para analizar la presencia de las publicaciones *Scopus* en los diferentes dominios científicos, así como la distribución mundial de la capacidad editorial de los diferentes países o regiones. La interfaz utiliza la metodología de mapas *overlay* para situar los subconjuntos de publicaciones seleccionadas en el contexto de la estructura global de publicaciones.

Palabras clave

Mapas de las publicaciones científicas, Mapas de la ciencia, Datos bibliométricos, Citas, Interfaces gráficas, Representación de la ciencia, Herramientas para analizar la ciencia, *Scopus*, *SCImago*.

Hassan-Montero, Yusef; Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix (2014). "Graphical interface of the *SCImago Journal and Country Rank*: An interactive approach to accessing bibliometric information". *El profesional de la información*, May-June, v. 23, n. 3, pp. 272-278.

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

Introduction

As noted by Van-Eck (2011), bibliometric maps representing the relationships among nodes can mainly take two forms: as graphical links (such as lines or arcs), or, without such links, by positioning the nodes to represent their mutual proximity or distance. For large networks with a high density of links, as in the case of the present study, the latter option, in which the degree of similarity or clustering among nodes is reflected by their spatial proximity, is the only viable option to ensure that the map is legible and understandable.

In the construction of distance-based maps, the algorithmic technique most widely used is known as MDS (multi-dimensional scaling). This technique takes as input a similarity or distance matrix in an N-dimensional space that is reduced to two or three dimensions (coordinates), trying to make the distances between nodes in the final representation approximate as closely as possible the distance they had in the N-dimensional space. The main drawback of MDS, however, is its tendency to produce circular maps where the nodes of higher degree are located near the centre of the map (Van-Eck; Waltman, Dekker; Van-den-Berg, 2010), an artefact of the logic of its algorithmic procedure that distorts the structural realities of the network.

An alternative to MDS which does not impose this artificial structure on the final layout, and in which the clusters of nodes are more easily recognized, is VOS (visualization of similarities) (Van-Eck; Waltman, 2007), used successfully in constructing various bibliometric maps. The two techniques are closely related. Indeed, VOS may be considered a weighted type of MDS in which proximities and weights are determined in one particular way (Van Eck; Waltman; Dekker; Van-den-Berg, 2010).

Another alternative is to use algorithms of a force-directed or energy-based type, which iteratively apply forces of attraction and repulsion among the nodes to reach a state

(i.e., layout) of minimum energy. Many of the more popular force-directed algorithms (Eades, 1984; Davidson; Harel, 1989; Fruchterman; Reingold, 1991) are unsuitable for constructing distance-based bibliometric maps since their final layout mainly favours aesthetic criteria such as a uniform distribution of the nodes, the optimal use of space, or the avoidance of crossing between links. I.e., these algorithms do not seek to reflect the degree of similarity among nodes through their position, but to obtain readable layouts in representations based on graphical links. While the algorithm of Kamada & Kawai (1989) does reflect certain properties of the network in the layout, such as the geodetic distance between the nodes of the graph, it too imposes artificial aesthetic criteria such as uniform link length, and therefore equal distances between nodes.

Force-directed methods have two components: an energy model that defines which layout to compute, and an energy minimization algorithm which defines how to compute it

In contrast, a proposal that is especially well-suited to the construction of distance-based bibliometric maps is that of Noack's force-directed approach (2004, 2007). Force-directed methods have two components: an energy model that defines which layout to compute, and an energy minimization algorithm which defines how to compute it. Noack's proposal focuses on the former, using two energy models termed node-repulsion LinLog and edge-repulsion LinLog. The main quality of these models, especially the edge-repulsion model, lies in their ability to yield layouts in which the position of the nodes not only reveals similarity relations, but also the community structure (Noack, 2009).

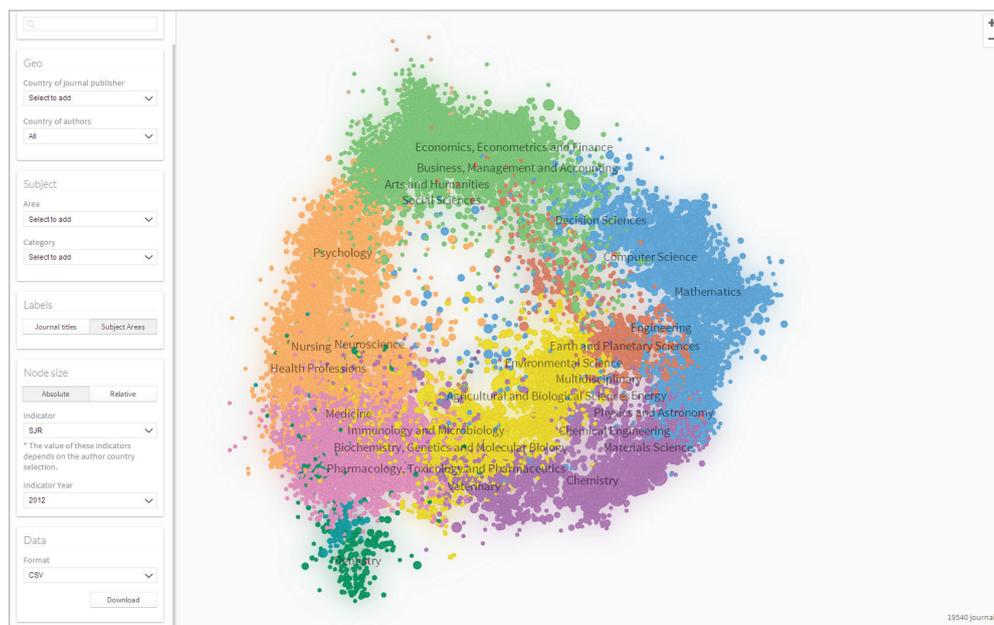


Figure 1. General view of the layout with the interface. The Subject Areas were selected as labels, and the SJR-2012 values as the size of the nodes.

Regarding the quality of the final layout, **Van-Eck (2001)** states *“In the case of maps with lots of objects, the accuracy of the VOS mapping technique at the local level can be somewhat disappointing”*, and adds *“An interesting mapping technique that seems to yield accurate results both at the local and at the global level is the LinLog technique”*. Finally, as will be discussed below, the force-directed type of algorithms have the advantage of being easily tunable through the application of well-known optimization strategies.

The objective of the present study was to design a graphical interface for SJ&CR based on the relationships connecting scientific journals, so that use of an overlay method would permit the academic production of different geographic and thematic domains to be represented on the overall base map.

Methods and data

Bibliometric maps display the structure of a domain through a graphical representation of the units of analysis and of their relationships. In the present study, the units of analysis are international scientific journals. To calculate the degree to which these journals are related, we applied the mean of the normalized values of three measures: citation, co-citation, and bibliographic coupling. When there only existed bibliographic coupling between two journals, the link is discarded due to its relative insignificance.

To facilitate subsequent processing, the directed graph formed by the journals is transformed into an undirected graph in which the weight of each link between two nodes is the sum of the weights of their directed links. The resulting network is composed of 19,540 vertices and 3,612,061 links.

The network was generated from *Scopus* data (2003-2012) and from the thematic classification of journals used in SJ&CR. This latter consists of Subject Areas (26 plus the General class) and Specific Subject Areas or Categories (309).

We started by implementing the force-directed algorithm

using the edge-repulsion energy model proposed by **Noack (2007)**.

Given the size of the network in this study, in order to process its layout with reasonable efficiency, we used a multi-scale strategy (**Hadany; Harel, 2001**). Such strategies are based on computing an abstraction of the graph on which a general layout is calculated, to which details are then added gradually, editing and fine-tuning the layout. The procedure used in the present study consisted of two phases: coarsening and un-coarsening.

The coarsening phase takes as input the original graph G_0 . In this graph, a downwards traverse is made of all its links ordered by their density, which is defined as

$$d_{ij} = w_{ij} / wD_i \bullet wD_j$$

where w_{ij} is the weight of the link, and wD_i and wD_j are the weighted degrees of the nodes (sum of the weights of all their links). For each link, if neither of its nodes has previously been collapsed, they are then collapsed to form a single node in graph G_1 , whose links will be the result of aggregating the links of the original two nodes. This process is repeated until reaching a graph G_n , the graph of greatest abstraction, in which it is no longer possible to continue collapsing nodes.

“ The results obtained in the present study suggest that the use of multi-scale strategies in combination with LinLog energy models, is a very efficient method to achieve high quality layouts for large networks ”

The un-coarsening phase starts from graph G_n , to which the force-directed positioning algorithm is applied using a very small number of iterations (in fact, we only used two iterations, although this is a criterion that depends on the topology and size of the network). After applying the algorithm, the positions (coordinates) of each node are inherited by the two nodes it corresponded to in graph G_{n-1} . This process of positioning or refining the layout is repeated until the original graph G_0 is reached again.

Although **Noack (2007)** shows a certain reluctance to regard as appropriate the use of multi-scale strategies in combination with LinLog energy models, the results obtained in the

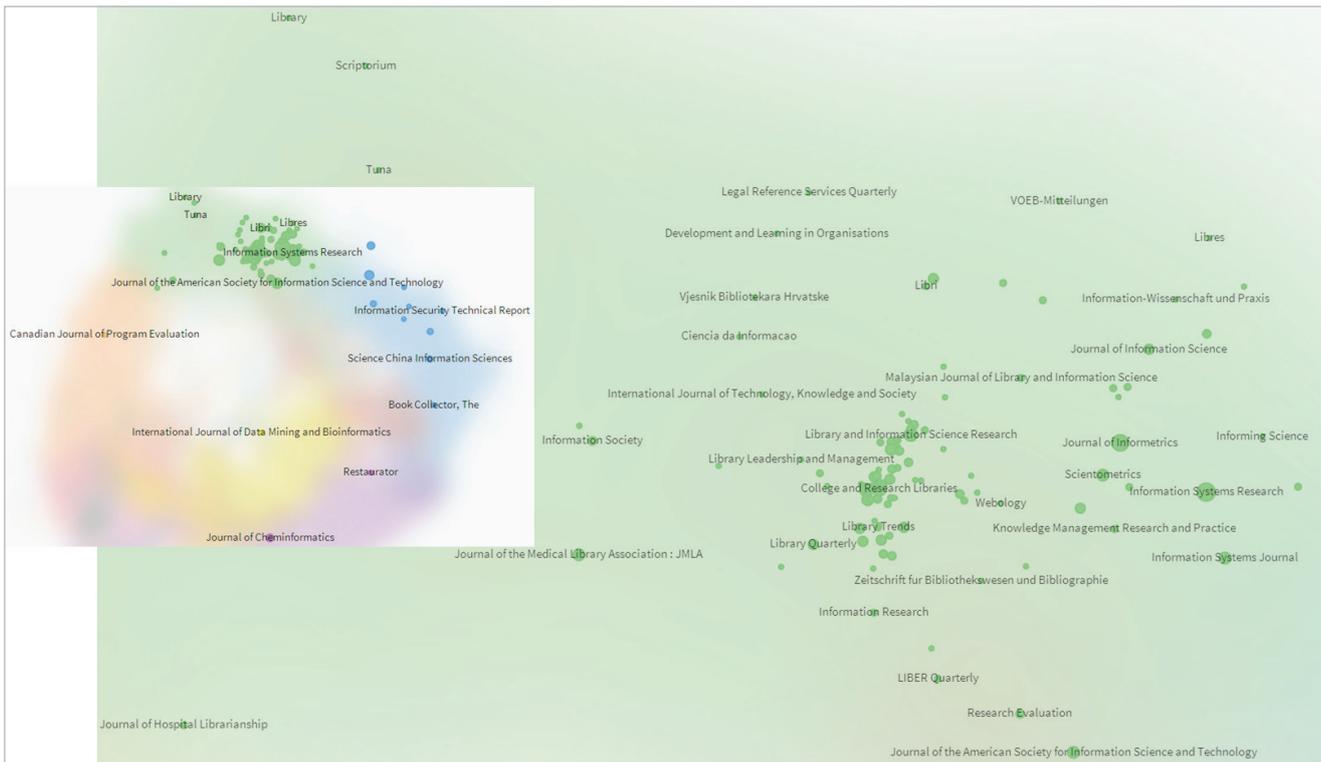


Figure 2. Library & Information Science worldwide with the *SJR-2012* as the absolute size of the nodes.

present study suggest that the procedure is indeed a very efficient method with which to achieve high quality layouts for large networks.

The nodes are represented graphically by discs whose size (their area) encodes the value of whichever bibliometric indicator that the user has selected (*SJR*, documents, references, citations...). The colour of the node identifies the clus-

ter to which the journal belongs. For the calculation of the different clusters, we used the algorithm implemented by **Noack** (2007) based on **Newman** and **Girvan's** measure of modularity (**Newman**, 2004; **Newman**; **Girvan**, 2004).

Once the overall map has been generated for the corresponding period, to select a domain, the interface provides the user with different options: the editor's country, the

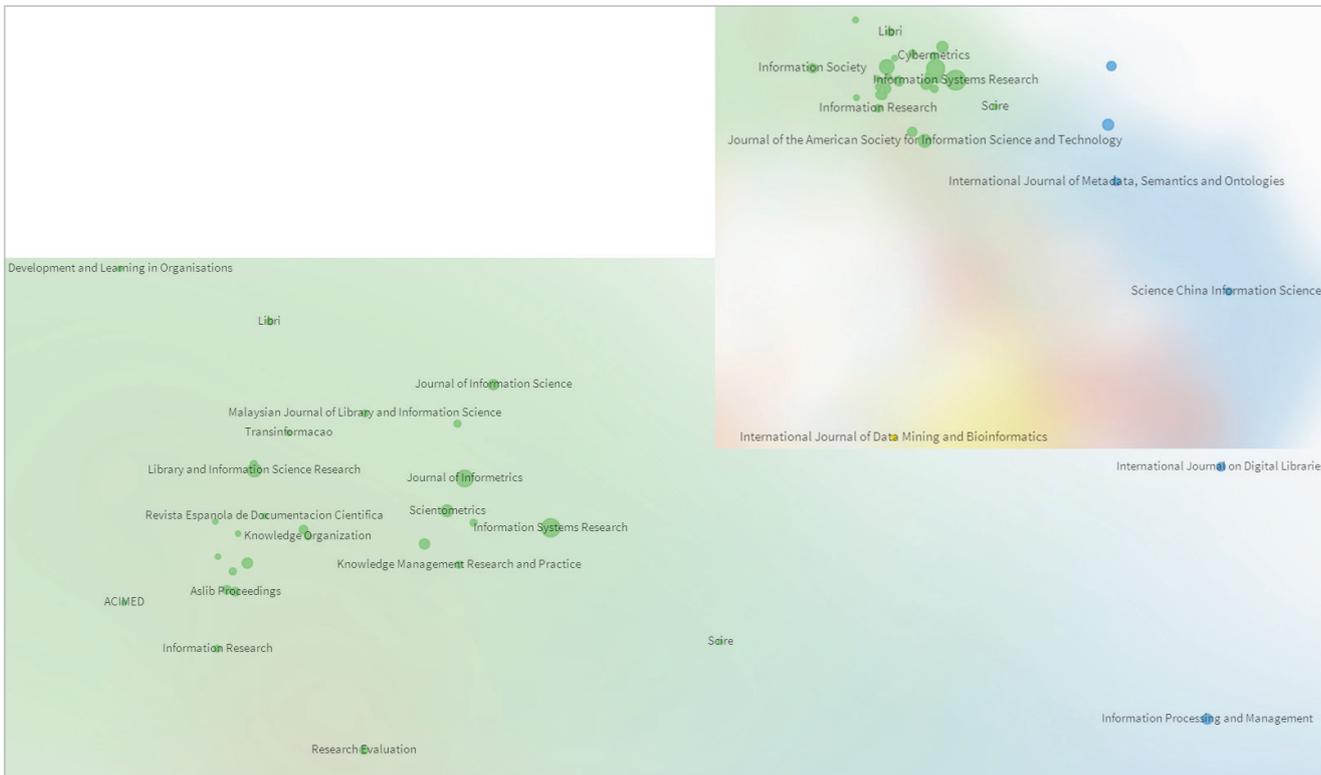


Figure 3. Library & Information Science in Spain with *SJR-2012* as the absolute size of the nodes.

authors' country, the area, and the subject category. The results of the selected domain are superimposed using the overlay technique (Leydesdorff; De-Moya-Aneón; Guerrero-Bote, 2010; 2014). In this case, as well as generating the overall map with different procedures and data, with everything being done online, the clusters are not only coded by the colour of the nodes, but also by the contour formed by the nodes of each cluster, combining Gaussian blur with transparencies.

To enable detailed exploration of the map, an interactive zoom-and-pan mechanism is included, as well as a journal search on title.

To facilitate the interpretation of the map, the interface allows one to select showing either the journal titles or the area and subject category labels. Each of these labels is positioned at the centroid of all the journals covering that area or category. Overlap between labels is avoided using the mechanism described in Koh, Lee, Kim & Seo (2010).

Results and discussion

Figure 1 presents an overview of the interface with the general layout. It includes all the journals (19,540), represented by nodes of sizes proportional to *SJR-2012*. As can be seen in the panel on the left, one can choose the indicator that will be rendered with the size of the nodes. This may be general for the journal, as is the case with *SJR*, or it may depend on the domain that has been selected, as might be the

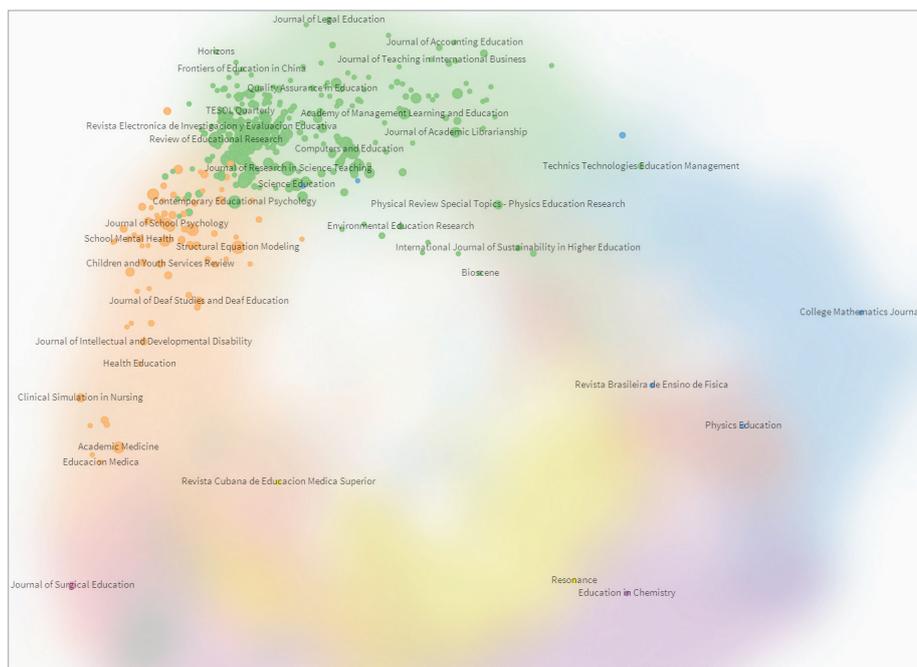


Figure 4. Overlay of the journals of the Category of Education with the *SJR-2012* as node size

output of the domain in the journal in question. This value can also be set to either absolute or relative.

The labels can also be set to either Subject Areas or Journal Titles.

The nine colours of the nodes correspond to the nine clusters or communities detected by the algorithm. The green cluster at the top is devoted mainly to Social Sciences and Humanities. The orange cluster on the left corresponds to Psychology, Neuroscience, Nursing, and Health Professions. Below it is a dark pink cluster devoted to Medicine which merges at the bottom with a blue-green cluster devoted to Ophthalmology and Optometry and then a teal-coloured cluster dedicated to Dentistry. More to the right, there are two clusters, one yellow dedicated to Life Sciences and the other purple devoted to Chemistry, Materials Science, and Chemical Engineering. Up and to the right, these two clusters fuse with another two, one reddish brown of Earth and Planetary Sciences, and the other blue, and rather elongated, starting from the bottom with Physics and Astronomy, then passing up through Mathematics, then Computer Science, to end with Decision Sciences when it merges with the first cluster.

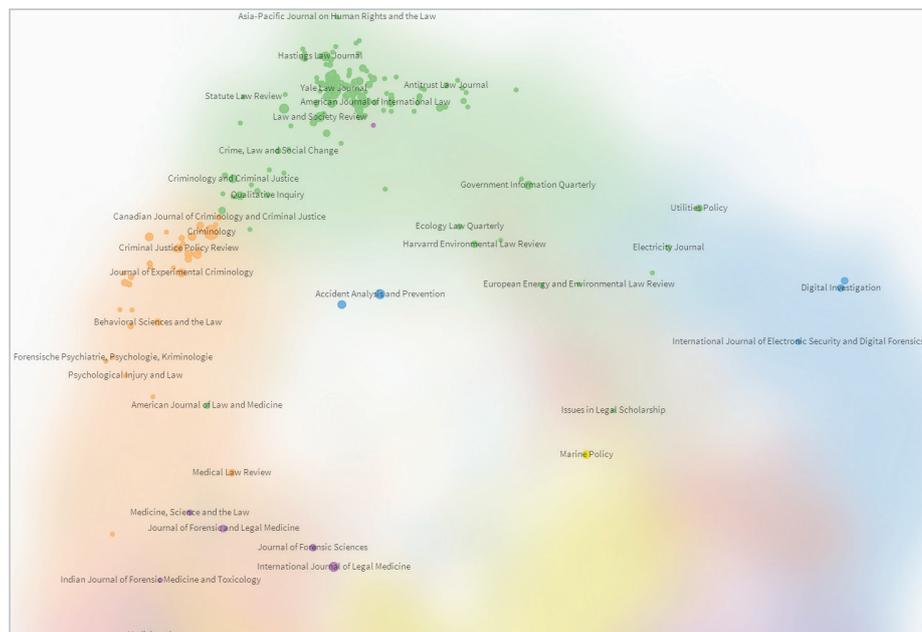


Figure 5. Overlay of the journals of the Category of Law with the *SJR-2012* as node size.

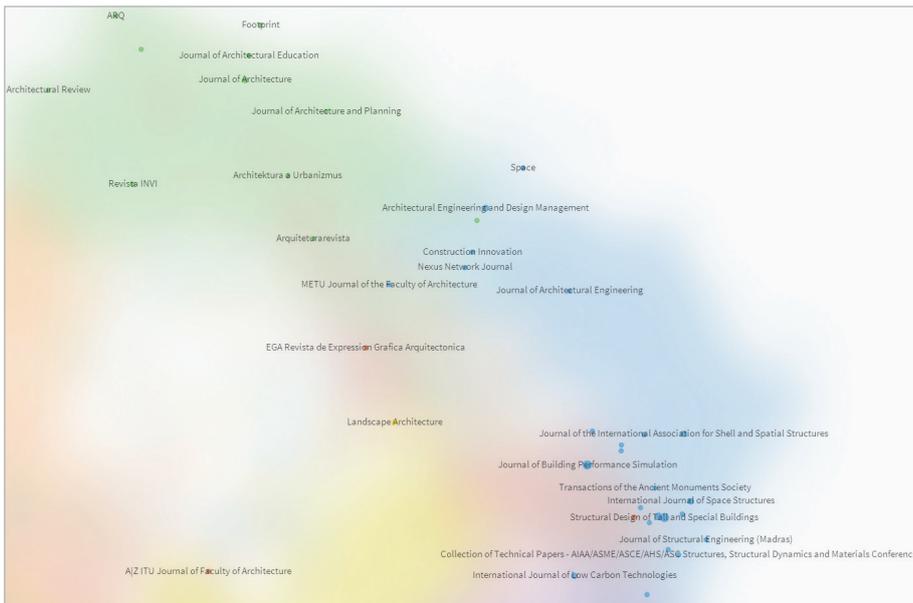


Figure 6. Overlay of the journals of the Category of Architecture with the *SJR-2012* as node size

the layout is evident in the very logic of the placement of the Subject Areas. Also worth mentioning is the clear coordination between the clustering algorithm and the layout since by far the greater part of the journal clusters (identified by their colours) are also grouped together visually.

If we zoom in on any of the parts with the Subject Areas as labels, those labels get larger, but they also blur out while the Specific Subject Area or Category labels start to appear.

Figure 2 shows the overlay corresponding to journals of the Specific Subject Area of Library & Information Science worldwide. The figure includes both a general view and a zoom-in on the most prominent grouping in which one can see a group of journals with Humanistic, Economic, and Scientometric content. Figure 3 shows the same Category but now delimited to the domain of Spain. There are now fewer journals since those displayed include only journals in which Spanish authors published in 2012. With fewer journals, some Spanish-language journals, such as *Scire* and the *Revista Española de Documentación Científica*, have become visible, and others can be located even though their title is not displayed because of the overlap, examples being the *Revista General de Información y Documentación* (slightly below and to the left of the *Revista Española de Documentación Científica*), *Investigación Bibliotecológica* (a little further down), *El Profesional de la Información* (and a little more further down), or *Cybermetrics* (above the *Journal of Informetrics*).

Figure 4 shows the overlay corresponding to the journals of the Category of Education worldwide. One can see the main group-

ing in the border zone between Psychology and Social Sciences. There also appear some scattered journals that correspond to educational journals in various specific disciplines.

Something similar is the case in figure 5 for Law. One sees the main grouping of journals in Humanities. There is another group of journals on Criminology found within Psychology, and, at the bottom, some journals devoted to Forensic Science and others that have published papers devoted to the different branches of Law.

In Architecture (figure 6), one sees two groups of journals, one very technical, and close to Engineering and the Hard Sciences, and another dedicated to architectural art.

Today it is necessary to combine different techniques, and especially to seek applications of the representations generated that go beyond what is obvious analytically

The case of Biomedical Engineering (figure 7) is somewhat special. Although *Scopus* includes it within Engineering, its journals do not focus on this area. Indeed, the label of the Category at the centroid of the positions of its journals appears at the bottom. Many journals are in a purple zone, corresponding to Chemistry or Materials Science; others in a yellow zone, corresponding to Biomedicine; and others in a pink zone, more within Medicine. This may be evidence

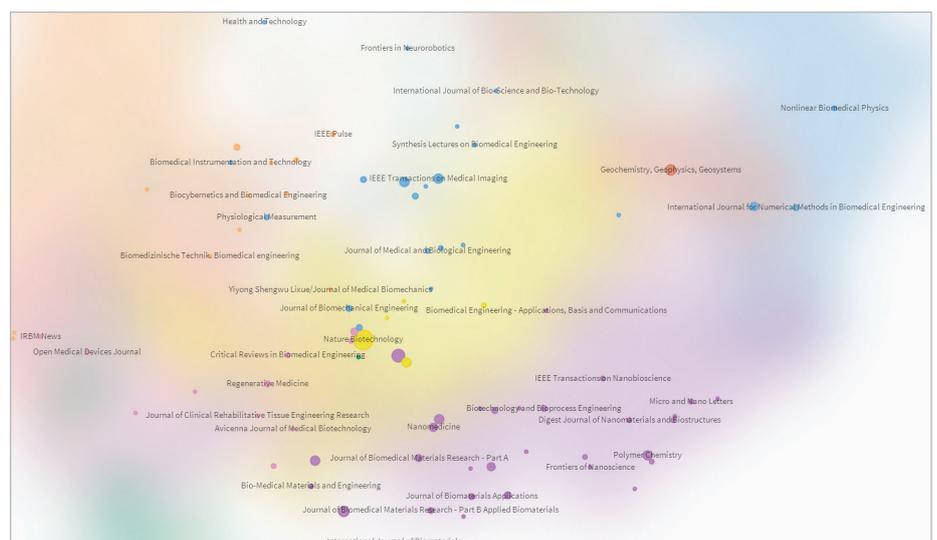


Figure 7. Overlay of the journals of the Category of Biomedical Engineering with the *SJR-2012* as node size

of a bad placement of the Category within the Engineering category in *Scopus*.

Conclusions

We have described a combination of force-directed, clustering, and overlay techniques to generate interactive maps capable of being used as efficient interfaces to access bibliometric information. Gone are the days when the mastery of a technique of mapping some particular form of bibliographic information was a novelty in the field of research. Today it is necessary to combine different techniques, and especially to seek applications of the representations generated that go beyond what is obvious analytically. Nonetheless, our principal objective with the present development was to provide users with a tool that would allow them to interact visually with all the information accessible through the *SCImago Journal & Country Rank* portal. The aim was, therefore, to offer an alternative to the traditional form-based access routes that are usual in bibliographic databases. This objective could not be attained without solving a host of human-computer interaction (HCI) problems posed by the simple volume of information being handled, and by the need to work at different levels of zoom with classifications represented graphically according to the user's requested context.

The solutions that we have described take as starting point the existence of a database of bibliometric indicators grouped by geographic domains, scientific fields, and publications which can now also be analysed visually in the form of graphical displays. This graphical approach is intended to constitute just another way of analysing that part of reality which the indicators can shed light on. In no way does it pretend to replace the tabular, georeferenced, or any other type of approach that the user can take. The methodological principle that motivates the need to generate multiple representations of the same social reality is that of their analytical complementarity.

Acknowledgments

This work was financed by the *Junta de Extremadura, Consejería de Empleo, Empresa e Innovación* and by the *European Social Fund* as part of the research group grant GR10019.

References

- Davidson, Ron; Harel, David** (1989). *Drawing graphs nicely using simulated annealing*. Technical Report CS89-13, Department of Applied Mathematics and Computer Science, The Weizmann Institute, Rehovot, Israel.
- Eades, Peter** (1984). "A heuristic for graph drawing". *Congressus numerantium*, v. 42, n. 11, pp. 149-160.
- Fruchterman, Thomas M. J.; Reingold, Edward M.** (1991). "Graph drawing by force-directed placement". *Software - practice & experience* (Wiley), v. 21, n. 11, pp. 1129-1164. <http://dx.doi.org/10.1002/spe.4380211102>
- Hadany, Ronny; Harel, David** (2001). "A multi-scale algorithm for drawing graphs nicely". *Discrete applied mathematics*, v. 113, n. 1, pp. 3-21. [http://dx.doi.org/10.1016/S0166-218X\(00\)00389-9](http://dx.doi.org/10.1016/S0166-218X(00)00389-9)
- Kamada, Tomihisa; Kawai, Satoru** (1989). "An algorithm for

drawing general undirected graphs". *Information processing letters*, n. 31, 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)

Klavans, Richard; Boyack, Kevin W. (2009). "Towards a consensus map of science". *Journal of the American Society for Information Science and Technology*, v. 60, n. 3, pp. 455-476. <http://dx.doi.org/10.1002/asi.20991>

Koh, Kyle; Lee, Bongshin; Kim, Bohyoung; Seo, Jinwook (2010). "ManiWordle: Providing flexible control over Wordle". *IEEE Trans vis comput graph.*, v. 16, n. 6, pp. 1190-1197. <http://dx.doi.org/10.1109/TVCG.2010.175>

Leydesdorff, Loet; De-Moya-Anegón, Félix; Guerrero-Bote, Vicente P. (2010). "Journal maps on the basis of *Scopus* data: A comparison with the *Journal Citation Reports* of the *ISI*". *Journal of the American Society for Information Science and Technology*, v. 61, n. 2, pp. 352-369. <http://dx.doi.org/10.1002/asi.21250>

Leydesdorff, Loet; De-Moya-Anegón, Félix; Guerrero-Bote, Vicente P. (2014). "Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of *Scopus* data (1996-2012)". *Journal of the American Society for Information Science and Technology* (en prensa), Wiley. <http://www.arxiv.org/pdf/1310.4966>

Newman, Mark E. J. (2004). "Analysis of weighted networks". *Physical review E*, n. 70, 056131. <http://dx.doi.org/10.1103/PhysRevE.70.056131>

Newman, Mark E. J.; Girvan, Michelle (2004). "Finding and evaluating community structure in networks". *Phys. rev. E*, n. 69, 026113. <http://dx.doi.org/10.1103/PhysRevE.69.026113>

Noack, Andreas (2004). *Visual clustering of graphs with nonuniform degrees*. Technical report 02/04, Brandenburg University of Technology at Cottbus (BTU).

Noack, Andreas (2007). "Energy models for graph clustering". *Journal of graph algorithms and applications*, v. 11, n. 2, pp. 453-480. <http://dx.doi.org/10.7155/jgaa.00154>

Noack, Andreas (2009). "Modularity clustering is force-directed layout". *Physical review E*, n. 79, 026102. <http://dx.doi.org/10.1103/PhysRevE.79.026102>

Van-Eck, Nees-Jan (2011). *Methodological advances in bibliometric mapping of science*. ERIM PhD Series research in management, 247. ISBN: 978 90 5892 291 5

Van-Eck, Nees-Jan; Waltman, Ludo (2007). "VOS: a new method for visualizing similarities between objects". In: H.-J. Lenz; R. Decker (eds.), *Advances in data analysis: Proceedings of the 30th annual conference of the German Classification Society* (pp. 299-306).

Van-Eck, Nees-Jan; Waltman, Ludo; Dekker, Rommert; Vandenberg, Jan (2010). "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS". *Journal of the American Society for Information Science and Technology*, v. 61, n. 12, pp. 2405-2416. <http://dx.doi.org/10.1002/asi.21421>



ADWORDS, IMÁGENES Y CEGUERA A LOS BANNERS. UN ESTUDIO CON EYE TRACKING



Laura Ortiz-Chaves, Eduardo Martínez-de-Pisón, Germán Cancela-López-Carrión, Juliano Gonçalves-de-Vasconcellos, Cristòfol Rovira y Mari-Carmen Marcos



Laura Ortiz-Chaves es graduada en información y documentación por la *Universidad de Salamanca* (2013). Actualmente es estudiante del máster en gestión de contenidos digitales de la *Universitat de Barcelona* y la *Universitat Pompeu Fabra*.
<http://orcid.org/0000-0001-9538-8607>

Universitat de Barcelona
Melcior de Palau, 140. 08014 Barcelona, España
lauraortizchaves@gmail.com



Eduardo Martínez-de-Pisón es licenciado en periodismo por la *Universidad Autónoma de Barcelona* (2013). Actualmente combina su trabajo de traductor con el máster en gestión de contenidos digitales de la *Universitat de Barcelona* y la *Universitat Pompeu Fabra*.
<http://orcid.org/0000-0002-1517-1430>

Universitat de Barcelona
Melcior de Palau, 140. 08014 Barcelona, España
eduardomartinezdepison@gmail.com



Germán Cancela-López-Carrión es graduado en producción de audio (*BA Degree in audio production*) por la *Universidad de Middlesex* (2013). Actualmente es estudiante del máster en gestión de contenidos digitales de la *Universitat de Barcelona* y la *Universitat Pompeu Fabra*.
<http://orcid.org/0000-0001-8033-1050>

Universitat de Barcelona
Melcior de Palau, 140. 08014 Barcelona, España
<http://gcancelalopez.com>
profesional@gcancelalopez.com



Juliano Gonçalves-de-Vasconcellos es agente técnico webdesigner del *Ministerio Público del Estado de Amazonas* en Brasil y estudiante del máster en gestión de contenidos digitales de la *Universitat de Barcelona* y la *Universitat Pompeu Fabra*. Especialista en gráfica digital en la *Universidad Federal de Pelotas* (2007), licenciado en diseño gráfico por la misma universidad (2005) y técnico en diseño industrial por el *Centro Federal de Educación Tecnológica de Pelotas* (1999).
<http://orcid.org/0000-0003-1559-3438>

Universitat de Barcelona
Melcior de Palau, 140. 08014 Barcelona, España
<http://www.julianovasconcellos.com>
julianovasconcellos@gmail.com



Cristòfol Rovira es profesor titular de la *Universitat Pompeu Fabra (UPF)*. Imparte docencia en los grados de periodismo y de publicidad y relaciones públicas y en el master oficial en gestión de contenidos digitales (*UB* y *UPF*) Director del máster en documentación digital (*UPF*) y del máster en buscadores (*UPF*). Es investigador del grupo *DigiDoc* de la *UPF* y director de la revista académica *Hipertext.net*.
<http://orcid.org/0000-0002-6463-3216>

Universitat Pompeu Fabra
Roc Boronat, 138. 08018 Barcelona, España
cristofol.rovira@upf.edu



Mari-Carmen Marcos es profesora de documentación en el *Departamento de Comunicación* de la *Universitat Pompeu Fabra (UPF)*. Es subdirectora del máster en documentación digital (*UPF*), co-directora del máster en buscadores (*UPF*), y profesora en el máster en gestión de contenidos digitales (*UB* y *UPF*). Es investigadora de los grupos *DigiDoc* y *Web Research Group* de la *UPF* y coordinadora de la revista académica *Hipertext.net*. Su línea de investigación es la interacción persona-ordenador en sistemas de recuperación de información.
<http://orcid.org/0000-0002-9576-3830>

Universitat Pompeu Fabra
Roc Boronat, 138. 08018 Barcelona, España
<http://www.mcarcos.com>
mcarmen.marcos@upf.edu

Resumen

A diferencia de la mayoría de la publicidad online, *Google* opta por un modelo puramente textual en los anuncios, que muestra en sus páginas de resultados en su servicio *AdWords*. Se realiza un experimento en el que se han modificado los *AdWords* originales de la zona lateral derecha para incorporar elementos gráficos, concretamente logotipos de los anunciantes y fotografías de los productos anunciados. Un grupo de usuarios es expuesto a estas páginas y mediante la técnica de *eye tracking* se analiza su comportamiento visual. Los resultados indican que la aparición de imágenes en estos anuncios laterales no supone un cambio en la interacción de los usuarios con los anuncios, lo que se achaca al llamado efecto de “ceguera a los banners”.

Palabras clave

AdWords, Publicidad, Buscadores, *Google*, *Eye tracking*, Test de usuarios.

Title: *AdWords*, images, and banner blindness: an eye-tracking study

Abstrac

Unlike most online advertising, *Google* opts for a purely textual model in the ads showed in its results pages. This paper presents an eye-tracking study conducted to determine if right-side graphic elements in *Google AdWords* affect users' visual behavior. *AdWords* have been modified to contain logos and pictures. The analysis concludes that the appearance of images does not change user interaction with ads, which is attributed to the “banner blindness” effect.

Keywords

AdWords, Advertisement, Search engines, *Google*, Eye tracking, User testing.

Ortiz-Chaves, Laura; Martínez-de-Pisón, Eduardo; Cancela-López-de-Carrión, Germán; Gonçalves-de-Vasconcellos, Juliano; Rovira, Cristòfol; Marcos, Mari-Carmen (2014). “AdWords, imágenes y ceguera a los banners. Un estudio con eye tracking”. *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 279-287.

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

1. Introducción

La publicidad es la base de los ingresos de muchos negocios de internet. *Google* ha sido consciente de ello desde sus inicios, y ofrece el servicio *AdWords* con el que los anunciantes muestran sus productos en la página de resultados del buscador (SERP, *search engine results page*) y el servicio *AdSense*, en el que los anuncios aparecen en la red de contenido formada por webs asociadas a *Google*.

Cuando los anuncios aparecen en el listado de resultados se suele aplicar el modelo de “pago por clic”, en el que el precio viene establecido por un sistema de subasta. El anunciante establece para qué búsquedas (para qué términos) quiere que su anuncio se muestre y el precio máximo que está dispuesto a pagar por cada clic que la gente haga en su anuncio y vaya a parar a su web. La posición del anuncio en relación con los otros anuncios que se disparan con la misma búsqueda viene determinada por diversos factores, en especial por la puja realizada: el precio máximo por clic que se ha indicado.

De forma complementaria, en esa posición influyen otros factores como la calidad del anuncio, que se calcula en función de la coherencia entre las palabras del anuncio, las palabras de las búsquedas y el texto de la página de aterrizado (la página enlazada desde el anuncio). El anunciante pagará exclusivamente cada vez que su anuncio sea clicado.

La calidad de los *AdWords* es importante porque a mayor relevancia es mayor la probabilidad de recibir clics, y por tanto de ingresar dinero: con un anunciante que pague 1€ por una palabra clave y cuyo anuncio consiga clics la mitad de las veces que aparezca, *Google* ingresará 0,5€ por impresión (un “impresión” es cada vez que aparece el anuncio); en cambio con un anunciante que pague 1,5€ pero consiga clics una quinta parte de las veces que se publique, *Google* ganará 0,3€ por impresión. Los datos del tercer trimestre de 2013 indican que *Google* reportó ingresos por publicidad consolidados de 14.890 millones de US\$, a los que habría que deducir los costes de adquisición de tráfico que fueron del 24% de los ingresos por publicidad¹.

La SERP de *Google* se divide en dos tipos de resultados: los orgánicos, obtenidos por el algoritmo del buscador y se sitúan en la parte central de la página; y los patrocinados, es decir, los anuncios de *AdWords*, que pueden aparecer en dos lugares: la parte superior de la SERP o el lateral derecho. A diferencia de los banners publicitarios, los anuncios en *Google* no incluyen elementos gráficos, lo que les da gran homogeneidad. Algunos sitios web como *Facebook* han seguido el formato de anuncio de *Google* pero sí incorporan imágenes. En vista de que el modelo de anuncio textual ha empezado a romperse en plataformas tan exitosas como *Facebook*, planteamos qué pasaría si se añadiesen elementos gráficos a los *AdWords* de *Google*, tratando de saber si ese cambio haría que los anuncios captasen más la atención de los usuarios.

Planteamos qué pasaría si se añadiesen elementos gráficos a los *AdWords* de *Google*, tratando de saber si ese cambio haría que los anuncios captasen más la atención de los usuarios

Tras una revisión bibliográfica sobre la atención visual en los *AdWords*, se llevó a cabo un experimento de laboratorio en el que cuatro grupos de usuarios realizaron un test bajo cuatro situaciones respectivamente:

- SERPs sin anuncios en la zona lateral;
- con anuncios en dicha zona;
- con los mismos anuncios pero incluyendo el logotipo del anunciante;
- con esos anuncios pero añadiendo una fotografía del objeto anunciado.

Un análisis estadístico de las diferencias en el comportamiento visual y de clic de los usuarios en las 4 versiones debía revelar si existen diferencias significativas que puedan apoyar una propuesta de introducción de elementos gráficos en los *AdWords* de la zona lateral.

2. Trabajos previos

Aunque no disponemos de cifras de *Google*, diversos estudios confirman que el número de clics en anuncios es notoriamente menor que en resultados orgánicos. El estudio realizado por **Jansen y Resnick** (2006) concluye que el 70% de las veces se opta por un resultado orgánico. También **Nielsen** (2007) lo corrobora, en este caso a través de un estudio en el que analiza la mirada de un grupo de usuarios en páginas de resultados de buscadores. **Nielsen** detecta que el comportamiento visual en las páginas de resultados de los buscadores varía en función del interés por el contenido: algunas personas hacen un escaneo rápido, otras realizan una lectura parcial y un tercer grupo realizan una lectura en profundidad; en todos los casos encuentra una escasa fijación de la mirada en los anuncios.

Otros trabajos confirman la relación entre la atención prestada a los anuncios y la intención de búsqueda. **Manning, Raghavan y Schütze** (2008) diferencian cuatro tipos de intención: informacional, de navegación, transaccional y de

conectividad. De ellas, sólo en los casos en que los usuarios entran en el buscador con la intención de realizar una transacción es probable que se fijen en los anuncios, y por tanto las posibilidades de conversión de los *AdWords* son mucho mayores que en los otros tipos de búsquedas. Un estudio previo (**Jansen; Brown; Resnick**, 2007) indicaba además que los anuncios son más relevantes que los resultados orgánicos en las consultas sobre comercio-e, lo que apoya también los resultados sobre la mayor conversión en búsquedas transaccionales.

El motivo de la poca atención a los anuncios puede estar relacionado con el descubrimiento de **Benway y Lane** (1998) acerca de la “ceguera a los banners”. Estos investigadores colocaron en una web educativa un vistoso letrado con gran tipografía y color rojo que anunciaba nuevos cursos, y observaron el comportamiento de los usuarios. El resultado fue que los usuarios mayoritariamente omitían la presencia del banner y clicaban en un pequeño link que decía “courses”, con lo que se demostró que los anuncios visualmente llamativos no son necesariamente más visibles ni reciben mayor número de clics. Este fenómeno ha seguido siendo estudiado, y en el caso de las SERPs se ha demostrado que la “ceguera” se acentúa cuando se trata de los anuncios situados en la zona lateral derecha con respecto a los de la zona superior de la página (**Owens; Chaparro; Palmer**, 2011).

Dada la importancia que tiene para los anunciantes recibir clics en los *AdWords*, diversos estudios han tratado de descubrir qué motivos influyen en los usuarios a la hora de fijar su atención en estos anuncios y en su decisión de clicarlos. Son bastantes las investigaciones publicadas sobre el comportamiento visual de las personas en las SERPs.

Granka, Feusner y Lorigo (2008) presentan un completo estado de la cuestión sobre la aplicación de la técnica de seguimiento de la mirada (*eye tracking*) al estudio de SERPs y detallan las distintas métricas que pueden utilizarse. Distintos trabajos usan el *eye tracking* y centran de nuevo su atención en la relación que se da entre la atención y la intención del usuario.

Los anuncios visualmente llamativos no son necesariamente más visibles ni reciben mayor número de clics

Buscher, Dumais y Cutrell (2010) estudian lo que ocurre en consultas informacionales (el usuario busca información sobre un tema) y navegacionales (busca una página web en particular), y usan en las SERPs testeadas anuncios tanto relevantes como irrelevantes a las consultas. Los autores concluyen que el comportamiento visual de las personas en los resultados orgánicos está influenciado por el tipo de intención de su consulta pero también por la relevancia de los anuncios mostrados; en cambio, la atención que dedican a los anuncios sólo depende de la relevancia, y no de la intención.

En esta misma línea, **González-Caro y Marcos** (2011) comprobaron que se registran pocas miradas en los anuncios, y

que los usuarios sólo centran su mirada en ellos cuando se trata de consultas con carácter comercial e intención transaccional; de esas miradas, el 70% recae en la zona de anuncios superior y sólo el 30% en la zona lateral, lo que concuerda con los resultados de **Shrestha, Owens y Chaparro** (2009), que demuestran cómo la ubicación de los anuncios tiene incluso más peso que la relevancia. Además, del poco tiempo que se leen los anuncios, casi la mitad se dedica a la lectura del título del anuncio (**Marcos; González-Caro**, 2010).

Otros estudios se centran en analizar la ratio de clics (CTR, *click-through ratio*) de los *AdWords*. **Richardson, Dominowska y Ragno** (2007) abordan el tema de la predicción del CTR proponiendo un algoritmo de estimación para los anuncios nuevos. **Yang y Ghose** (2010) realizan un experimento sobre un anunciante en particular, y monitorizan la actividad de los usuarios desde las SERPs para buscar relaciones entre la presencia de resultados orgánicos de esa empresa y el CTR de sus anuncios, y viceversa. Descubren que la relación es positiva en ambos sentidos: la presencia de anuncios favorece el CTR en los resultados orgánicos, y también -aunque en menor medida- la presencia de resultados orgánicos favorece el CTR en los *AdWords*.

Otros trabajos se centran en estudiar la forma en que las personas procesamos las imágenes cuando acompañan a los textos. El estudio de **Aggarwal y Van-Oostendorp** (2012) demuestra que las imágenes se procesan en los primeros momentos de ver una página web, concretamente en la primera décima parte del tiempo total dedicado a esa página. Si las imágenes son relevantes al texto, los usuarios se detienen en ellas intentando relacionarlas con la lectura; en cambio, si se trata de publicidad, no se fijan durante tanto tiempo en ellas, pero sí las miran varias veces y les distraen de la lectura, según el estudio de **Beymer, Orton y Russell** (2007).

El presente artículo tiene como objeto de estudio el formato de presentación de los *AdWords*, concretamente los que aparecen en la zona lateral derecha de la SERP por ser los que menos captan la atención de los usuarios. La pregunta de investigación que nos hacemos es si la incorporación de elementos gráficos en estos anuncios influye en el comportamiento -visual y de clic- de los usuarios en la SERP. Tras una revisión bibliográfica se pone de manifiesto que no hay estudios previos que aborden esta pregunta, por lo que sería el primer trabajo sobre incorporación de elementos gráficos en anuncios en buscadores. En la siguiente sección se detallan la metodología aplicada y el diseño experimental realizado, y se continúa con los resultados y las conclusiones.

3. Metodología

Eye tracking

Este estudio sigue un método de investigación experimental basado en un test con usuarios aplicando la técnica de *eye tracking*, que consiste en el seguimiento y grabación del movimiento ocular de las personas. Por medio de un sistema de alta precisión basado en el reflejo de rayos infrarrojos en la pupila de los usuarios, el dispositivo capta los movimientos oculares (sacadas) y el detenimiento de las pupilas

en determinados puntos (fijaciones). De esa forma detecta dónde se ha centrado la mirada y durante cuánto tiempo. El aparato graba al mismo tiempo en qué puntos se ha efectuado un clic.

Muestra

Las personas seleccionadas son usuarias habituales de *Google*. Se contó con la participación de 40 usuarios, reclutados de diversas facultades de la *Universitat Pompeu Fabra* (Barcelona), lugar donde se llevó a cabo el experimento. La proporción de hombres y mujeres era del 50%, todos entre 17 y 20 años. El tamaño de la muestra es similar e incluso mayor al usado en los trabajos previos mencionados en los que se ha aplicado la técnica de *eye tracking*.

Variables

La variable independiente del experimento es el tipo de anuncio mostrado. Se plantearon 4 variantes, cuya diferencia estriba únicamente en la presentación de los anuncios de la zona lateral por tratarse del área con menos miradas y clics según los trabajos previos consultados:

- SERP original obtenida de *Google*;
- variante de la anterior sin anuncios en el lateral;
- versión que incluía en los anuncios laterales el logo de las empresas anunciantes;
- versión que incluía en los anuncios laterales una foto del producto que se anunciaba.

La variable dependiente es el comportamiento -visual y de clic- de los usuarios, y se concreta en dos grupos de métricas, escogidas tras la revisión de bibliografía (**Granka; Feuser; Lorigo**, 2008):

1) Métricas aplicadas a las 3 áreas de la interfaz: zona lateral de anuncios, anuncios superiores y la de resultados orgánicos. Para el análisis los datos de porcentaje se han convertido a números absolutos:

- *Percentage fixated* (PF): porcentaje de casos en los que se registran miradas en cada zona.
- *Percentage clicked* (PC): porcentaje de casos en los que se registran clics en cada zona.

2) Métricas aplicadas al área objeto de estudio, la zona lateral de anuncios:

- *Time to first fixation* (TFF): tiempo promedio que pasa desde que se abre la SERP hasta que los usuarios fijan por primera vez su mirada en esta zona.
- *Total fixation duration* (TFD): tiempo promedio que dedican los usuarios a mirar esta zona.

Diseño experimental

Consistió en la preparación de las SERPs y del test en el software *Tobii studio*. Para comenzar se realizaron 4 búsquedas de tipo transaccional y comerciales, que tenían como escenario la preparación de un viaje de amigos a Rovaniemi (Finlandia). Así se obtuvieron 4 SERP para estas consultas: vuelos a Rovaniemi, hotel, alquiler de coche, y compra de ropa de nieve de segunda mano.

A partir de la captura de pantalla de las 4 páginas de resultados originales, todas con anuncios laterales, se crearon

las otras 3 versiones para cada una de ellas en un editor de imágenes. Todas las SERPs del estudio cuentan con 3 anuncios AdWords en la parte superior (los mismos en cada búsqueda en sus 4 versiones), 10 resultados orgánicos (siempre los mismos en cada SERP) y 5 AdWords laterales en los 3 casos en los que se previó que hubiera anuncios en esa zona, en un caso son los originales, en otro se han añadido los logotipos de las empresas anunciantes, y en el último se han incorporado las fotografías de los productos anunciados.

Se realizó una captura de pantalla de cada SERP (16 en total, puesto que son 4 búsquedas y cada una tiene 4 versiones) y se insertaron en documentos html para simular la apariencia de página web. Al clicar sobre cualquier resultado, aparecía un mensaje pidiendo al usuario que continuara con la siguiente tarea del test, así hasta completar las 4 búsquedas.

Se crearon 4 grupos, cada uno sería expuesto a una versión de las SERPs, y de forma aleatoria cada usuario fue asignado a uno de ellos, de manera que la muestra inicial se compuso de 4 grupos de 10 usuarios en cada uno.

Ejecución

Para la grabación de datos se usó el modelo de *eye tracker Tobii 1750* integrado en un monitor TFT de 17" con una resolución de 1280x1024 píxeles. Se calibró el equipo para cada grabación realizada, y el foco de luz estuvo siempre en la misma ubicación. La distancia entre el usuario y el monitor fue siempre de aproximadamente 60 cm, lo que se controló mediante el uso de una silla fija.

Los usuarios seguían las instrucciones que el propio test les daba, que consistía en revisar cada SERP y escoger mediante un clic el resultado que consideraban que les sería más útil para responder a la consulta planteada. Los usuarios no sabían que había distintas versiones del test ni que existía una manipulación de la zona de AdWords laterales.

Análisis

Una vez realizadas las grabaciones, se marcaron tres áreas en cada SERP:

- la zona de los resultados orgánicos;
- la de los anuncios en la zona superior;
- a de los anuncios en la zona lateral.

Esta marcación, realizada con el software *Tobii studio*, permite extraer las métricas de las miradas y de los clics de los usuarios en cada zona de la pantalla. A pesar de que sólo la zona lateral es objeto de este estudio, se marcaron también las otras dos zonas para poder saber en cuántos casos se registraban miradas en cada área, pues se quería contrastar lo leído en la bibliografía acerca de que la zona lateral es la que menos atención recibe de las tres.

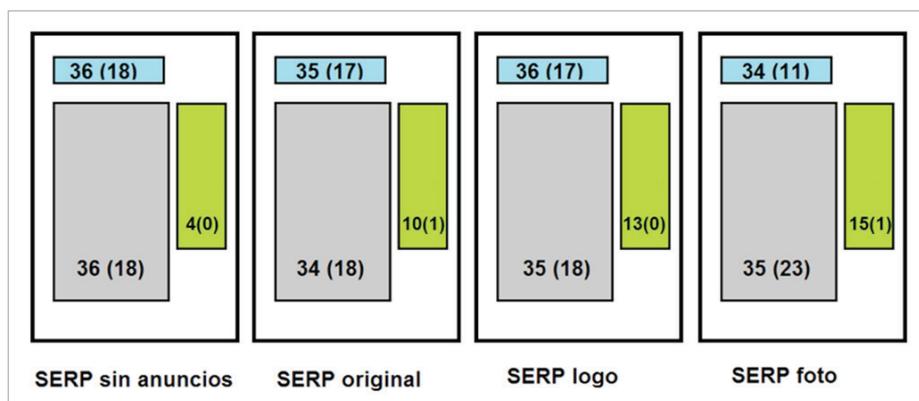


Figura 1. Esquema de las 4 versiones de SERP con el marcado de las 3 áreas: zona de anuncios en la parte superior, resultados orgánicos y zona de anuncios en el lateral. En cada área se ha indicado el número de casos (sumatorio de las SERPs de las 4 búsquedas efectuadas) en los que se registra alguna fijación, seguido entre paréntesis del número de casos en los que se registra un clic en esa área

En primer lugar se contaron los casos en los que cada zona había recibido miradas. A continuación y por medio de pruebas estadísticas, se compararon los datos registrados de las miradas en la zona de anuncios laterales de la SERP "original" con los datos de esa misma área para las otras versiones de las SERPs, denominadas "logo", "foto" y "sin anuncios". Los cálculos aplicados se han realizado teniendo en cuenta:

- Datos del tiempo absoluto, es decir, considerando todo el tiempo de permanencia en la SERP; y datos de tiempo relativo, considerando el porcentaje de tiempo dedicado a la zona lateral de anuncios con relación al tiempo total de cada usuario en la SERP.
- Datos de todos los usuarios de los que se han obtenido grabaciones correctas, y datos exclusivamente de aquellos que miraron el área de los anuncios laterales en cada SERP.

Para el análisis estadístico se usó el software *IBM SPSS* (versión *Statistics 22*). Los tests que se han utilizado son los propios de la estadística descriptiva, el *test Kolmogorov-Smirnov* para conocer la normalidad de los datos, el de *Levene* para identificar la igualdad o no de varianzas, y el *test Mann-Whitney* para comparación de muestras de medidas independientes para datos no paramétricos.

Tiempo de realización del estudio

El estudio se realizó con el calendario siguiente:

- octubre 2013: diseño y planificación;
- noviembre 2013: realización del experimento con el *eye tracker*;
- diciembre 2013: análisis estadístico;
- enero 2014: conclusiones y redacción del artículo.

4. Resultados

Cuatro de las 40 grabaciones fueron defectuosas y no se pudieron utilizar, de manera que para los cálculos se ha contado con 36 usuarios, 9 para cada versión. Como cada usuario ha sido expuesto a 4 SERP (porque había 4 búsquedas), la muestra total es de 36 casos de estudio para cada versión de SERP, 144 casos en total.

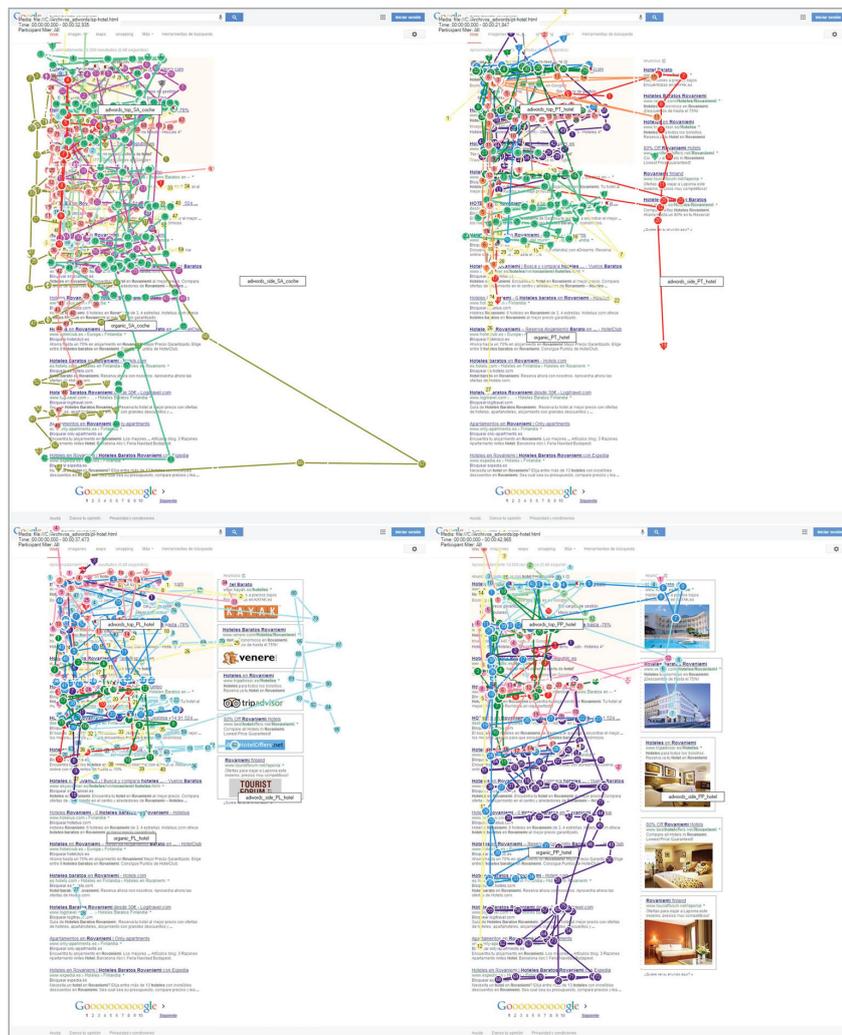


Figura 2. Mapa de recorrido de la mirada (*gazeplot*) de 9 usuarios en las 4 versiones de SERP para la búsqueda de hoteles en Rovaniemi, de izquierda a derecha y de arriba abajo: sin anuncios, original, con logotipo y con fotografías. El mapa ilustra cómo las fijaciones –representadas por pequeños círculos– y las sacadas –recorrido de la vista entre dos fijaciones– se concentran en las áreas de anuncios superiores y resultados orgánicos, y apenas se dan casos en los que las miradas vayan a la zona de anuncios lateral.

Fijaciones y clics por área

Los primeros resultados provienen de las métricas *percentage fixated* y *percentage clicked*. Se ha trasladado el porcentaje a números absolutos y muestran respectivamente el número de casos en que se fija la mirada en cada zona de la pantalla, y entre paréntesis el número de clics registrados en cada una (figura 1). Los usuarios miraron prácticamente siempre -entre 34 y 36 casos de los 36 registrados- tanto la zona de resultados orgánicos como la zona de *AdWords* en la parte superior, y el número de clics recibidos por ambas zonas es muy similar en todas las versiones excepto en la SERP con fotos, donde los orgánicos reciben bastantes más clics que los anuncios de la zona superior. En cambio, lo que en este trabajo nos interesa, que es el número de casos en los que se mira la zona de los anuncios laterales, es siempre muy bajo: 10 casos en la versión original, 13 en la versión con logotipos y 15 en la versión con fotografías. Mención aparte merecen los 4 casos en los que se mira esa zona lateral en la versión sin anuncios, esto es debido a que es una zona “de paso” para ir a la barra de *scroll*. Estas primeras observaciones van en línea con la bibliografía consultada,

donde se afirmaba que la zona lateral de anuncios recibe poca atención por parte de los usuarios.

También corresponde a lo apuntado por autores anteriores lo que se observó en la elección de los usuarios a la hora de escoger un resultado: mientras los *AdWords* de la zona superior son mirados y clicados en aproximadamente la mitad de los casos, los laterales presentan muchos menos casos en que son mirados y tan sólo en dos ocasiones se convierten en clics, una vez en la SERP con diseño original, y otra en la SERP con diseño de *AdWords* con fotografías.

De este esquema (figura 1) se deduce un patrón: el número de usuarios que miran la zona lateral de anuncios aumenta según añadimos riqueza gráfica en esa zona. Por lo tanto, estos primeros datos nos llevan a pensar que la riqueza gráfica de los *AdWords* puede ser un punto diferenciador a la hora de atraer la mirada de las personas. Los mapas de recorrido que proporciona el software *Tobii studio* ilustran muy bien esta observación (figura 2).

Comportamiento visual en la zona lateral

Con la finalidad de conocer si la riqueza gráfica supone una diferencia en el comportamiento de las personas, se ha realizado un análisis estadístico de las dos métricas relativas a la mirada:

- tiempo que tardan los usuarios en mirar los *AdWords* laterales (*TFF, time to first fixation*);

- tiempo que dedican los usuarios a mirar esos *AdWords* (*TFD, total fixation duration*).

Como se indicaba en el apartado de metodología, los cálculos se han realizado tanto considerando los valores absolutos como los relativos, e incluyendo a todos los usuarios en unos casos y sólo aquellos con fijaciones en la zona estudiada en otros.

En este análisis, los resultados del *test Mann-Whitney* indican que sólo en 4 de las 18 comparaciones realizadas se ha encontrado una diferencia estadísticamente significativa (p -valor $< 0,05$) al contrastar el comportamiento visual en la zona de anuncios laterales en el diseño original de la SERP con el comportamiento en las otras 3 versiones manipuladas (tabla 1).

Se han encontrado diferencias significativas en el comportamiento visual en estas situaciones:

- 1) Los usuarios pasan más tiempo mirando la zona de anuncios laterales de la SERP original que en la versión sin anuncios, algo que era previsible ya que si no hay contenidos es de esperar que no fijen en ella su mirada, salvo que

Tabla 1. Promedios (prom) y significancia estadística (p) obtenidos del *test Mann-Whitney* para muestras independientes de datos no paramétricos. Las métricas analizadas son TFF: tiempo que tardan los usuarios en mirar los AdWords laterales; TFD: tiempo que dedican los usuarios a mirar los AdWords laterales.

	Original vs Logo	Original vs Foto	Original vs Sin anuncios
(1) TFF, valores absolutos para todos los usuarios	Prom orig.: 1,650 seg	Prom orig.: 1,650 seg	Prom orig.: 1,650 seg
	Prom logo: 2,818 seg	Prom foto: 2,527 seg	Prom sin: 1,673 seg
	p = 0,403	p = 0,295	p = 0,085
(2) TFF, valores absolutos considerando sólo usuarios que miraron	Prom orig.: 5,940 seg	Prom orig.: 5,940 seg	Prom orig.: 5,940 seg
	Prom logo: 7,803 seg	Prom foto: 6,064 seg	Prom sin: 15,055 seg
	p = 0,642	p = 0,542	p = 0,723
(3) TFD, valores absolutos para todos los usuarios	Prom orig.: 1,078 seg	Prom orig.: 1,078 seg	Prom orig.: 1,078 seg
	Prom logo: 0,334 seg	Prom foto: 0,926 seg	Prom sin: 0,034 seg
	p = 0,860	p = 0,333	p = 0,037
(4) TFD, valores absolutos considerando sólo usuarios que miraron	Prom orig.: 3,880 seg	Prom orig.: 3,880 seg	Prom orig.: 3,880 seg
	Prom logo: 0,925 seg	Prom foto: 2,223 seg	Prom sin: 0,302 seg
	p = 0,110	p = 0,346	p = 0,005
(5) TFD, valores relativos para todos los usuarios	Prom orig: 0,054%	Prom orig: 0,054%	Prom orig: 0,054%
	Prom logo: 0,023 %	Prom foto: 0,059%	Prom sin: 0,003%
	p = 0,994	p = 0,438	p = 0,380
(6) TFD, valores relativos sólo para usuarios que miraron	Prom orig: 0,197%	Prom orig: 0,197%	Prom orig: 0,197%
	Prom logo: 0,066%	Prom foto: 0,141%	Prom sin: 0,026%
	p = 0,003	p = 0,166	p = 0,007

se dirijan a la barra de *scroll* y su vista pase por esa zona (tabla 1: métricas (3), (4) y (6)).

- 2) Los usuarios pasan más tiempo mirando la zona lateral de anuncios en la versión original que en la versión con logotipos (tabla 1: métrica (6)). Lo podemos interpretar si pensamos que con los logotipos todavía es más claro que esta zona contiene publicidad, de ahí que los usuarios sean más reacios a buscar la información allí.

La incorporación de imágenes en los anuncios laterales de las SERPs en Google no provoca un cambio significativo en el comportamiento visual de los usuarios que examinan la página web

5. Discusión

Si bien un primer análisis de los datos muestra que el número de casos en que los usuarios dirigen su mirada a la zona lateral aumenta según se incorporan elementos gráficos a los anuncios laterales (métrica *percentage fixated*) y que los clics se concentran en los anuncios superiores y los resultados orgánicos (métrica *percentage clicked*), al realizar pruebas estadísticas en las métricas de *time to first fixation* y *total fixation duration* los resultados muestran que la incorporación de imágenes en los anuncios laterales de las SERPs en Google (logotipos de los anunciantes o fotografías de los productos) no provoca un cambio significativo en el comportamiento visual de los usuarios que examinan la página web.

Un posible motivo por el que no se han podido encontrar diferencias estadísticamente significativas en esas dos métricas relativas al comportamiento visual de los usuarios en

las distintas versiones de las SERPs es el tamaño de muestra en la zona de anuncios laterales. Al tratarse de una zona que se mira muy poco, el número de casos con datos en la zona lateral es muy reducido. Dado que se trata de un experimento llevado a cabo necesariamente en laboratorio debido al uso de un *eye tracker*, trabajar con muestras de cientos o miles de usuarios -como ocurre en estudios de analítica web- es inviable. La homogeneidad en el perfil de los usuarios, todos ellos nativos digitales, usuarios de Google y en gran parte habituados a las compras online, puede haber introducido un sesgo en su comportamiento, por lo que en el futuro se ampliará la muestra con perfiles de otras

edades y hábitos en internet para comprobar si el patrón detectado se mantiene.

Ateniéndonos a los resultados obtenidos y para interpretar por qué el tipo de anuncio lateral no ha influido en el comportamiento visual en cuanto a la atracción visual y a la duración de las miradas, creemos que el motivo es el ya mencionado efecto de “ceguera a los banners”. Esa ceguera no significa literalmente que no los veamos, de hecho las personas tenemos un radio de visión de más de 180 grados, dentro del cual vemos con mayor nitidez aquellos objetos en lo que centramos la atención; esta visión nítida es la llamada mirada foveal y se llama así porque se registra en la fovea, la zona de la retina con más conos. La tecnología de *eye tracking* deja constancia de lo que las personas captamos con la fovea, y no con la mirada parafoveal y con la periférica, donde no ponemos la atención y la visión no es tan nítida. El efecto de ceguera se refiere a que si bien vemos los banners, en este caso los anuncios laterales, no lo hacemos poniendo atención en ellos, y el motivo puede ser que asumimos que no nos interesan e inconscientemente evitamos mirarlos con atención.

Los anuncios mostrados en la parte superior son mucho más efectivos en su versión textual que los anuncios laterales en cualquiera de las versiones testeadas

La diferencia sí que es significativa al comparar los anuncios originales con los que llevan logotipos; en este caso las miradas todavía son más esquivas cuando se presentan las marcas de forma gráfica. El efecto de ceguera, en cambio, no se produce en los anuncios de la zona superior; casi todos

los usuarios testeados en este experimento fijan en algún momento su mirada en estos anuncios, y en muchos casos los clican, por lo que de cara a los anunciantes, los anuncios mostrados en esta ubicación son mucho más efectivos en su versión textual que los anuncios laterales en cualquiera de las versiones testeadas

Al iniciar esta investigación conocíamos el efecto de ceguera a los banners, pero quisimos poner a prueba un diseño gráfico de anuncio para saber si al introducir elementos de este tipo se conseguía mayor atención visual en los anuncios. Los resultados que muestran el número de casos en los que se miran los anuncios van en esa dirección, pero la diferencia observada en cuanto al tiempo que tardan en verlos y el tiempo que dedican a mirarlos no es estadísticamente significativa, por lo que enriquecer gráficamente los AdWords no parece ser ahora mismo una vía rentable de negocio para Google ni para los anunciantes.

Enriquecer gráficamente los AdWords no parece ser ahora mismo una vía rentable de negocio para Google ni para los anunciantes

Los resultados obtenidos son novedosos y de interés para los equipos de ingeniería de los buscadores, y podrían dar lugar a nuevas investigaciones en las que las pruebas se realizaran sobre los AdWords de la zona superior de la SERP, ya que reciben muchas miradas y podría llegarse a conclusiones más contundentes, o a otras investigaciones encaminadas a evaluar la efectividad de los AdWords con distintas cualidades, como su extensión, lo que sería además de interés para los anunciantes.

Notas

1. Google: Q3 2013. *Quarterly earnings summary*. http://investor.google.com/pdf/2013Q3_google_earnings_slides.pdf
2. Este trabajo forma parte del proyecto de investigación "CSO2011-22691 - Nuevas estrategias de publicidad y promoción de las marcas turísticas españolas en la Web", Ministerio de Ciencia e Innovación, 2012-2014.

Bibliografía

- Aggarwal, Sonal; Van-Oostendorp, Herre** (2012). "When are pictures processed on a webpage?". En: *Intelligent human computer interaction (IHCI), 4th Intl conf*, IEEE, pp. 1-6. <http://dx.doi.org/10.1109/IHCI.2012.6481874>
- Benway, Jan P.; Lane, David M.** (1998). *Banner blindness: web searchers often miss "obvious" links*. Internet Technical Group. http://www.ruf.rice.edu/~lane/papers/banner_blindness.pdf
- Beymer, David; Orton, Peter; Russell, Daniel** (2007). "An eye tracking study of how pictures influence online reading". En: *Human-Computer Interaction - Interact 2007*. Springer Berlin Heidelberg, pp. 456-460. http://dx.doi.org/10.1007/978-3-540-74800-7_41

Buscher, Georg; Dumais, Susan T.; Cutrell, Edward (2010). "The good, the bad, and the random: an eye-tracking study of ad quality in web search". En: *Procs of the 33rd intl ACM Sigir conf on research and development in information retrieval*, July.

<http://research.microsoft.com/pubs/156119/SIGIR2010-Buscher-GoodBadRandom.pdf>
<http://dx.doi.org/10.1145/1835449.1835459>

Dumais, Susan T.; Buscher, Georg; Cutrell, Edward (2010). "Individual differences in gaze patterns for web search". En: *Procs of the 3rd Symposium on information interaction in context*. ACM.

http://research.microsoft.com/pubs/156118/IliX2010-DumaisEtal_Gaze_final.pdf
<http://dx.doi.org/10.1145/1840784.1840812>

González-Caro, Cristina; Marcos, Mari-Carmen (2011). "Different users and intents: an eye-tracking analysis of web search". En: King, Irwin; Nejdl, Wolfgang; Li, Hang. *WSDM'11 4th ACM Intl conf on web search and data mining*. <http://ciio.unab.edu.co/gti/images/documentos/umwa-8.pdf>

Granka, Laura; Feusner, Matthew; Lorigo, Lori (2008). "Eye monitoring in on-line search". En: Hammoud, Riad (ed.). *Passive eye monitoring: algorithms, applications and experiments*, pp. 347-372. ISBN: 978 3 540 75412 1
http://dx.doi.org/10.1007/978-3-540-75412-1_16

Jansen, Bernard J.; Brown, Anna; Resnick, Marc (2007). "Factors relating to the decision to click on a sponsored link". *Decision support systems*, Nov., v. 44, n. 1, pp. 46-59. <http://dx.doi.org/10.1016/j.dss.2007.02.009>

Jansen, Bernard J.; Resnick, Marc (2006). "An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce web searching". *Journal of the American Society for Information Science and Technology*, v. 57, n. 14. http://faculty.ist.psu.edu/jjansen/academic/pubs/sponsored_links_jansen.pdf
<http://dx.doi.org/10.1002/asi.20425>

Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hartmut (2008). *Introduction to information retrieval*. Cambridge University Press, ISBN: 9780511809071
<http://nlp.stanford.edu/IR-book>
<http://dx.doi.org/10.1017/CBO9780511809071>

Marcos, Mari-Carmen; González-Caro, Cristina (2010). "El comportamiento de los usuarios en la página de resultados de los buscadores: un estudio basado en la técnica de eye tracking". *El profesional de la información*, v. 19, n. 4, pp. 348-358. http://grupoweb.upf.es/WRG/dctos/marcos_gonzalez_2010.pdf
<http://dx.doi.org/10.3145/epi.2010.jul.03>

Nielsen, Jakob (2007). "Banner blindness: old and new findings". *Nielsen Norman Group: evidence based user experience research, training and consulting*, August 20. <http://www.nngroup.com/articles/banner-blindness-old-and-new-findings>

Owens, Justin W.; Chaparro, Barbara S.; Palmer, Evan M. (2011). "Text advertising blindness: the new banner blindness?". *Journal of usability studies*, v. 6, n. 3. http://usabilityprofessionals.org/upa_publications/jus/2011may/JUS_Owens_May_2011.pdf

Richardson, Matthew; Dominowska, Ewa; Ragno, Robert (2007). "Predicting clicks: estimating the click-through rate for new ads". En: *Procs of the 16th Intl conf on World Wide Web, WWW'07*, pp. 521-530. <http://www.conference.org/www2007/papers/paper784.pdf>

<http://dx.doi.org/10.1145/1242572.1242643>

Shrestha, Sav; Owens, Justinand; Chaparro, Barbara S. (2009). "The effect of location and congruency of text ads on information search". En: *Procs of the Human Factors and Ergonomics Society annual meeting*, October, v. 53, n. 18. <http://dx.doi.org/10.1177/154193120905301821>

Yang, Sha; Ghose, Anindya (2010). "Analyzing the relationship between organic and sponsored search advertising: positive, negative, or zero interdependence?". *Marketing science*, v. 29, n. 4, pp. 602-623. <http://dx.doi.org/10.1287/mksc.1090.0552>

ANUARIO THINKEPI

VERSIÓN PAPEL

Tarifas institucionales

<input type="checkbox"/> Anuario 2014	No disponible
<input type="checkbox"/> Anuario 2013	97 €
<input type="checkbox"/> Anuario 2012	50 €
<input type="checkbox"/> Anuario 2011	40 €
<input type="checkbox"/> Anuario 2010	35 €
<input type="checkbox"/> Anuario 2009	30 €
<input type="checkbox"/> Anuario 2008	30 €
<input type="checkbox"/> Anuario 2007	30 €

Tarifas individuales (particulares)

<input type="checkbox"/> Anuario 2014	No disponible
<input type="checkbox"/> Anuario 2013	43 €
<input type="checkbox"/> Anuario 2012	30 €
<input type="checkbox"/> Anuario 2011	23 €
<input type="checkbox"/> Anuario 2010	21 €
<input type="checkbox"/> Anuario 2009	21 €
<input type="checkbox"/> Anuario 2008	21 €
<input type="checkbox"/> Anuario 2007	21 €

VERSIÓN DIGITAL

Tarifas institucionales

<input type="checkbox"/> Anuario 2014	68 €
<input type="checkbox"/> Anuario 2013	65 €
<input type="checkbox"/> Anuario 2012	45 €
<input type="checkbox"/> Anuario 2011	35 €
<input type="checkbox"/> Anuario 2010	30 €
<input type="checkbox"/> Anuario 2009	30 €
<input type="checkbox"/> Anuario 2008	30 €
<input type="checkbox"/> Anuario 2007	30 €
<input type="checkbox"/> Anuario 2014+2013	120 €
<input type="checkbox"/> Anuario 2014+2013+2012	150 €
<input type="checkbox"/> Anuario 2014 a 2007	240 €

Tarifas individuales (particulares)

<input type="checkbox"/> Anuario 2014	35 €
<input type="checkbox"/> Anuario 2013	30 €
<input type="checkbox"/> Anuario 2012	25 €
<input type="checkbox"/> Anuario 2011	25 €
<input type="checkbox"/> Anuario 2010	20 €
<input type="checkbox"/> Anuario 2009	20 €
<input type="checkbox"/> Anuario 2008	20 €
<input type="checkbox"/> Anuario 2007	20 €
<input type="checkbox"/> Anuario 2014+2013	55 €
<input type="checkbox"/> Anuario 2014+2013+2012	65 €
<input type="checkbox"/> Anuario 2014 a 2007	120 €



Formulario de compra Anuario ThinkEPI

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

ALFABETIZACIÓN MEDIÁTICA Y CO-INNOVACIÓN EN LA MICROEMPRESA: PRIMERAS EVIDENCIAS PARA ESPAÑA

Juan-Francisco Martínez-Cerdá y Joan Torrent-Sellens



Juan-Francisco Martínez-Cerdá es PhD Candidate en el *eLearn Center – Universitat Oberta de Catalunya (UOC)*. Investiga aspectos implicados en el desarrollo de habilidades orientadas a la empleabilidad de los trabajadores, a través de los sistemas sociotécnicos (STS) relacionados con el *e-learning*. Investigador en la *Universitat Autònoma de Barcelona (UAB)*, fue jefe de estudios de la *Fundació Observatori per a la Societat de la Informació de Catalunya (FOBSIC)* y mantiene vinculación con el *Grupo Interdisciplinario de Investigación sobre las TIC (i2TIC)* de la UOC desde 2009. <http://orcid.org/0000-0002-6711-4956>

jmartinezcer@uoc.edu



Joan Torrent-Sellens es director de la *Business School - Universitat Oberta de Catalunya (UOC)*, director del *Grupo Interdisciplinario de Investigación sobre las TIC (i2TIC)* y profesor agregado de los *Estudios de Economía y Empresa (UOC)*. Sus líneas de investigación son la economía del conocimiento, el emprendimiento y la innovación, la empresa y el trabajo en red, las fuentes de productividad, competitividad y crecimiento económico, y la salud y el bienestar, ámbitos en los que ha publicado varios libros, así como artículos en revistas nacionales e internacionales. <http://orcid.org/0000-0002-6071-422X>

jtorrent@uoc.edu

Universitat Oberta de Catalunya. MediaTIC
Roc Boronat 117, Pl 6
08018 Barcelona, España

Resumen

El inexorable avance del uso de las TIC ha provocado cambios radicales a nivel social y empresarial. En este contexto, resulta clave establecer qué tipo de competencias personales, vinculadas con el uso eficiente de los medios de comunicación, pueden ser utilizadas desde una óptica empresarial, y cómo estas competencias mediáticas se pueden alinear con otros factores determinantes para la competitividad de la empresa como la innovación. Se analizan las habilidades mediáticas de los trabajadores y su conexión con el proceso de generación de innovaciones en la microempresa. Con este objetivo analítico, se ha realizado un trabajo de campo exploratorio a una muestra de 154 profesionales que desarrollan su actividad profesional en microempresas españolas.

Palabras clave

Alfabetización mediática, Fuentes co-innovadoras, Sociedad del conocimiento, Innovación, Gestión del conocimiento, Empresas.

Title: Media literacy and co-innovation in microenterprises: first evidence in Spain

Abstract

The inexorable progress of ICT has brought about radical changes in the social and business world. In this context, new patterns of communication and participation, especially since the advent of Web 2.0, and the widespread use of smartphones by the whole population, are particularly striking. In this situation, it is essential to establish what kind of personal skills related to the efficient use of media can be useful from a business perspective, and how these media literacy skills can be aligned with other factors, such as innovation, that determine the competitiveness of companies. In this paper, we analyze the media skills of workers and their connections to the innovation processes in microenterprises. With this analytical objective, exploratory fieldwork has been done in a sample of 154 professionals working in Spanish microenterprises.

Keywords

Media literacy, Co-innovative sources, Knowledge society, Innovation, Knowledge management, Enterprises.

Martínez-Cerdá, Juan-Francisco; Torrent-Sellens, Joan (2014). "Alfabetización mediática y co-innovación en la microempresa: primeras evidencias para España". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 288-299.

<http://dx.doi.org/10.3145/epi.2014.may.09>

Artículo recibido el 09-07-2013
Aceptación definitiva: 02-04-2014

1. Introducción

El análisis de la vinculación entre el uso de tecnologías de la información y la comunicación (TIC), y el uso consciente y crítico por parte de los ciudadanos de contenidos multimedia procedentes de medios de comunicación, nos remite a la denominada *alfabetización mediática* (Livingstone, 2003; European Commission, 2009; Celot; Pérez-Tornero, 2009; Koltay, 2011). Se trata de un área del conocimiento con un alcance más humanístico y de pensamiento crítico (Pérez-Tornero; Celot; Varis, 2007) que la denominada *alfabetización tecnológica o digital*, más orientada hacia las habilidades de acceso y uso de las TIC, o que la llamada *alfabetización informacional*, que considera una gran cantidad de tipos de fuentes de información, más allá de los medios de comunicación.

La alfabetización informacional se refiere a las competencias clave que “todas las personas precisan para su realización y desarrollo personales, así como para la ciudadanía activa, la inclusión social y el empleo”, como se señala en la *Recomendación del Parlamento Europeo* y del Consejo sobre las competencias clave para el aprendizaje permanente (European Commission, 2006). De este modo la alfabetización informacional se hace necesaria en aspectos relacionados con la vida profesional, social, cultural, personal o del ocio, como, por ejemplo, los relacionados con la comprensión de informaciones de carácter político, administrativo y laboral, así como en la educación para la salud y la educación para la información -conocimientos relacionados con el acceso, uso, lectura, producción, comunicación, evaluación y ética de la información- (Cuevas-Cerveró; García-Moreno, 2010), ámbito éste último de gran importancia durante las últimas décadas por su relevante relación con la economía de la información. Cabe señalar, por ejemplo, que la alfabetización informacional gozó del llamado “mes de concienciación nacional” en EUA, en octubre de 2009 (*The White House*, 2009). La economía de la información o informacional, desarrollada entre otros modelos a través del concepto de la empresa red (Castells, 1996), ha contribuido a la consolidación de la sociedad de la información a lo largo y ancho del planeta.

Existen relaciones de complementariedad entre usos intensivos de las TIC, formación y aprendizaje continuo de los trabajadores, y nuevas formas de organización en red del trabajo

Desde el punto de vista empresarial, la literatura ha constatado la existencia de nuevas fuentes co-innovadoras que explicarían la productividad de las empresas (Torrent-Sellens, 2010). Así, para diversas empresas de todo el mundo se ha constatado el establecimiento de relaciones de complementariedad en pos de la innovación, y a través de tres fuentes:

- 1) usos intensivos de las TIC;
- 2) formación y aprendizaje continuo de los trabajadores;
- 3) nuevas formas de organización en red del trabajo.

El análisis de la alfabetización mediática, entendida como uso eficiente y crítico a nivel empresarial de las noticias procedentes de los medios de comunicación, y el estudio de las fuentes co-innovadoras de la productividad empresarial, se han llevado a cabo en paralelo, pero sin líneas de conexión. Es interesante señalar la gran trascendencia que en el mundo de la empresa ha tenido la denominada inteligencia competitiva, asociada a procesos de innovación como los relacionados con la vigilancia tecnológica y la monitorización de patentes, y que mediante diferentes técnicas analíticas (como el planteamiento de escenarios, los estudios Delphi o los mapas tecnológicos) ha logrado convertirse en una de las principales herramientas estratégicas empresariales (Palo; Vicente, 1999; Aenor, 2011; Aldasoro-Alustiza; Cantonnet-Jordi; Cilleruelo-Carrasco, 2012).

Se empieza a establecer una primera conexión entre estas dos áreas del conocimiento, la inteligencia competitiva y la alfabetización mediática, aunque relegada sobre todo al desarrollo de los *social media* (Kubátová, 2011). La intención de este artículo es ampliar esta incipiente relación, conectando ambas líneas de análisis con el objetivo de ampliar la comprensión del proceso de transición hacia la empresa red y la sociedad del conocimiento (United Nations, 2005; Bindé, 2005).

En relación con los factores que intervienen en el desarrollo económico de la sociedad, resulta muy interesante tener en cuenta una aproximación que vaya más allá de la alfabetización digital e informacional, y que tenga en cuenta el ámbito de la alfabetización mediática, basada en factores que unan la utilización crítica, eficiente y provechosa por los trabajadores de las informaciones procedentes de los medios de comunicación.

Desde esta perspectiva, la presente investigación busca generar cierto entendimiento y comprensión acerca de la relación existente entre la alfabetización mediática y el contexto microempresarial de economía del conocimiento.

Evidentemente, si a nivel individual las personas poseen un conjunto de competencias que les permiten utilizar y aprovechar los contenidos que las TIC y los medios de comunicación les proporcionan, resulta especialmente interesante para el futuro de las empresas y del conjunto de la sociedad, conocer la influencia y el aprovechamiento de dichas habilidades en la empresa, sobre todo en el contexto de la innovación.

2. Marco analítico

El derecho a la comunicación, presente en el artículo 19 de la *Declaración de los Derechos Humanos de la ONU* (ONU, 1948), es la garantía para el acceso de los ciudadanos a los medios de comunicación. Se distinguen dos tipos de tendencias en el uso de la información procedente de los medios de comunicación:

- 1) la que tiene que ver con el mero consumismo informacional y comunicativo;
- 2) la relacionada con la ciudadanía activa, con una orientación crítica, política y democrática para el uso de la información accedida y/o producida (alfabetización mediática).

Es importante entender que ambas son necesarias y pue-

den ser desarrolladas mediante los programas educativos correspondientes.

La literatura científica ha identificado diversos usos de la alfabetización mediática para el desarrollo de la ciudadanía (Saleh, 2009), de los valores democráticos en contenidos curriculares (Burroughs *et al.*, 2009), y para la fusión de factores de las ciencias de la información y la comunicación dentro del desarrollo de políticas de alfabetización mediática (Lau; Cortés, 2009). La enseñanza de las capacidades para juzgar lo que se lee y se ve, y para producir y difundir aquello que se piensa, resulta especialmente clave en nuestra actual sociedad democrática (Moeller, 2009), que por otro lado utiliza la información como mercancía con la que se realizan intercambios, trueques, comercio y, en última instancia, economía.

La alfabetización mediática a nivel empresarial puede ser uno de los aspectos más importantes para el desarrollo de innovaciones empresariales

Es clave la vinculación que comienza a establecerse entre la alfabetización mediática y la gestión del conocimiento, como componente requerido en el capital humano y relacionado con la organización y la síntesis de información, en el camino hacia la sociedad del conocimiento (Unesco, 2013a). En este contexto de competencias tecnológicas para el acceso a la información (o alfabetización digital), cabe reseñar el aprendizaje autodidacta que los jóvenes -futuros trabajadores- están llevando a cabo, en cuanto a un uso de la información orientado hacia el ocio y la sociabilidad (Sánchez-Navarro; Aranda, 2011). Este aprendizaje puede ser aprovechado desde la óptica empresarial, puesto que conlleva habilidades y competencias personales muy valoradas en el contexto actual asociado a la innovación, como las relacionadas con el uso intensivo de las TIC, y el aprendizaje continuo de las herramientas informáticas y comunicativas existentes para dichos mecanismos de esparcimiento y compartición online de materiales audiovisuales multimedia.

También se observan los profundos cambios tecnológicos, económicos y sociales que las TIC están provocado, como el acceso a mercados a lo largo del planeta, y el cambio en los patrones de demanda de las familias y las empresas (Vilaseca-Requena *et al.*, 2003). Se detecta que los conocimientos científicos y tecnológicos se incorporan en las prácticas económicas (Torrent-Sellens, 2002) y que los factores técnicos y del campo de las habilidades personales tienen una influencia sobre los agentes económicos (McLuhan; Powers, 1989; McLuhan, 1996; Torrent-Sellens, 2008).

Además, durante la última década se ha observado una relación de reforzamiento entre factores que actúan de manera complementaria en pos de la innovación, y que se denominan fuentes co-innovadoras:

- el acceso a las TIC y el uso de las mismas;
- cualificación de los empleos y los trabajadores;
- reorganización de los entornos de trabajo.

Se distinguen diferentes impactos de estas variables en la productividad de las empresas (Torrent-Sellens; Ficapal-Cusí, 2010a; 2010b; 2011). Son claves tanto las competencias profesionales orientadas al uso de la información en los puestos de trabajo (Ortoll-Espinet, 2004) -que podríamos vincular con la alfabetización informacional, y que han sido llevadas a cabo tradicionalmente por las bibliotecas, impulsoras desde siempre de la misma-, como las que tienen que ver con la inteligencia competitiva (Jaguszewski; Williams, 2013). En nuestro campo de investigación, la Unesco establece la importancia de la alfabetización mediática para el diseño de planes formativos para trabajadores (Unesco, 2013b). Asimismo, la importancia de los medios de comunicación a nivel empresarial puede ser reconocida en la propuesta de Anderson, Olming y MacFarquhar (2008) para la Organización Internacional del Trabajo, que presenta tres tipos de roles de los medios de comunicación en el desarrollo de las empresas:

1. Cambio de actitudes y comportamientos -cultura empresarial- en los negocios y la sociedad;
2. Servicios e información sobre los *inputs* y *outputs* de mercado de las micro y pequeñas empresas;
3. Entornos de políticas y normativas para las pequeñas empresas.

Tener un óptimo uso o manejo crítico de los medios de comunicación -alfabetización mediática-, a nivel empresarial, puede ser uno de los aspectos más importantes para el desarrollo de innovaciones empresariales, la mejora de la productividad de la empresa y, en definitiva, el desarrollo de la sociedad del conocimiento.

En este artículo se plantean las siguientes preguntas de investigación:

- P1: ¿Comparten los trabajadores noticias procedentes de los medios de comunicación, en su contexto empresarial y de un modo cotidiano?
- P2: ¿Se aprovecha en el contexto empresarial la información que procede y se comparte de los medios de comunicación?

Asimismo, se tienen en cuenta las siguientes premisas:

- Pr1. Los trabajadores tienen acceso a medios de comunicación
- Pr2. Los trabajadores tienen conocimientos y son usuarios de las TIC
- Pr3. Las empresas disponen de infraestructuras TIC

Las preguntas de investigación guardan relación con las tres componentes propuestas por la Unesco para la evaluación de los niveles de alfabetización mediática e informacional a nivel global (Unesco, 2013b):

- la P1 se relaciona con las componentes 1 (acceso y recuperación) y 3 (creación y uso);
- la P2 mantiene una vinculación con la componente 2 (comprensión y evaluación).

También se pueden establecer relaciones entre las preguntas de investigación y los criterios y componentes del enfoque europeo para la medición de la alfabetización mediática

Tabla 1. Preguntas y respuestas del cuestionario. Encuesta online

Pregunta en el cuestionario	Valores (respuestas)
P1. ¿Qué medios de comunicación sueles utilizar normalmente, para acceder o enterarte de noticias?	Radio Televisión Prensa escrita Revistas periódicas Otros
	Prensa digital Redes sociales Otros medios digitales
P2. Si encuentras alguna información relevante para tu empresa, generalmente, ¿sueles compartirla con tus compañeros y/o jefes?	Sí
	No
P3. Si sueles compartir este tipo de información en tu empresa, ¿consideras que es bien aprovechada?	Sí
	No
P4. a) Aproximadamente, en tu equipo de trabajo más próximo, ¿cuántos compañeros suelen compartir alguna información relevante para la empresa, y que encuentran en los medios de comunicación?	Numérico
P4. b) ¿Cuántos compañeros tienes en tu equipo de trabajo más próximo?	Numérico
P5. Aproximadamente, ¿cada cuánto tiempo sueles tú -o tus compañeros- compartir información proveniente de los medios de comunicación, y que pueda ser útil para la empresa?	Diariamente
	Semanalmente
	Mensualmente
	Ocasionalmente
	Nunca Otros
P6. Independientemente de tus aportaciones, ¿son aprovechadas las noticias que aparecen en los medios de comunicación y pueden tener relación con tu empresa, como parte integrante de los procesos de innovación y/o toma de decisiones?	Sí, de un modo formal
	Sí, de un modo informal
	No
	Otros
P7. ¿En qué ámbitos de la empresa crees que la información existente en los medios de comunicación puede ser más aprovechable y útil para tu organización?	Administración/gestión/finanzas/RRHH
	Investigación + desarrollo + innovación (I+D+i)
	Infraestructural
	Compras/proveedores
	Producción de bienes y servicios
	Promoción/marketing
	Ventas/distribución/postventa
	Otros
P8. ¿Qué tareas relacionadas con internet sabes realizar?	Usar un buscador
	Enviar un email
	Hacer llamadas telefónicas o videoconferencias
	Compartir ficheros P2P (intercambio de música, películas, etc.)
	Descargar vídeos, películas, juegos, etc.
	Crear páginas web propias
	Subir contenidos a la web 2.0 (fotos y vídeos en <i>Youtube</i> , <i>Facebook</i> , etc.)
	Comprar productos y realizar pagos en tiendas online
	Comprar un dominio de internet propio
	Descargar software e instalarlo Otros
P9. ¿Qué usos de las tecnologías de la información y la comunicación (TIC) se llevan a cabo en el equipo de trabajo más próximo a ti, en tu empresa?	Uso de ordenadores
	Redes de ordenadores (LAN, con cables)
	Redes inalámbricas (wifi, sin cables)
	Conexión a internet
	Existencia de página web
	Existencia de intranet
	Integración de información: ERP, CRM, B2B, EDI, etc.
	Uso de software abierto (<i>open source</i>)
	Aplicaciones web 2.0 de forma empresarial (marketing vía <i>Facebook</i> , uso de <i>skype</i> para videoconferencias, uso de <i>Youtube</i> en la página web, etc.)
	Tienda electrónica (venta de los productos a través de internet)
	Telefonía móvil de empresa para uso empresarial Otros

a nivel de países (Celot; Pérez-Tornero, 2009).

Las preguntas planteadas están explícitamente relacionadas tanto con la definida alfabetización mediática como con el ámbito de la gestión del conocimiento. De cara a establecer una conexión entre la gestión del conocimiento y las fuentes co-innovadoras, hay que destacar que la gestión del conocimiento tiene un gran impacto sobre, como mínimo, dos de los factores que actúan de un modo complementario en pos de la innovación (fuentes co-innovadoras): la formación y el aprendizaje continuo de los trabajadores (Cohen; Levinthal, 1989), y las innovaciones no tecnológicas u organizativas (Fierro-Moreno; Mercado-Salgado; Cernas-Ortiz, 2013).

A partir de estas preguntas de investigación y autores reseñados, se establece una conexión entre las áreas de la gestión del conocimiento, la alfabetización mediática, y las fuentes co-innovadoras a nivel empresarial, relación que nos permite realizar un trabajo de campo con el que obtener datos que validen las cuestiones planteadas.

3. Metodología y materiales

El artículo pretende investigar las relaciones entre la alfabetización mediática, entendida como habilidades para el aprovechamiento de noticias procedentes de medios de comunicación por parte de los trabajadores de las empresas, y los diversos factores que intervienen, todos ellos bajo un esquema de complementariedad, en los procesos de innovación a nivel empresarial. Estos factores suelen denominarse fuentes co-innovadoras, y tienden a agruparse conceptualmente, como ya se ha señalado, en tres conjuntos: usos TIC; cualificación y nivel formativo de los trabajadores; y nuevas formas organizativas de los equipos de trabajo (innovación no tecnológica).

Para investigar las relaciones entre alfabetización mediática y co-innovación empresarial, se ha diseñado y puesto en marcha una encuesta online orientada a realizar una primera aproximación cuantitativa sobre el tema, siguiendo un enfoque basado en una metodología exploratoria con la que generar unas primeras hipótesis de trabajo que, posteriormente, puedan ser ampliadas de un modo mucho más profundo a través de, por ejemplo, un estudio descriptivo. Teniendo en cuenta el carácter novedoso de los objetivos de investigación, la realización de un estudio exploratorio resulta idónea (Babbie, 2007; Hernández-Sampieri; Fernández-Collado; Baptista-Lucio, 2010), puesto que nos permite adquirir una familiaridad con el tema, ayudándonos a encontrar unas primeras tendencias o relaciones entre las variables. Asimismo es una metodología esencial para abrir nuevos horizontes de investigación.

El proceso de captación utilizado es el conocido como técnica *snowball* o bola de nieve, que se basa en una muestra aleatoria no probabilística. Este tipo de muestreo, que se emplea por su rapidez y porque no supone un alto coste para la captación de datos, puede ser utilizado en estudios piloto de carácter cuantitativo (Statistics Canada, 2010), por lo que su aplicación en estudios exploratorios como éste es válida. Sus limitaciones, relacionadas con la imposibilidad de generalización de las conclusiones obtenidas al conjunto del universo estudiado, debido a la falta de representati-

dad probabilística de la muestra seleccionada, serán tenidas en cuenta en las conclusiones.

De este modo, la técnica *snowball* ha estado basada en una invitación a participar en la encuesta transmitida mediante e-mail (proceso iniciado el 28 de marzo de 2012), así como a través de las redes sociales *Facebook* y *Twitter*. En dicha propuesta de colaboración se comunicaban tanto los objetivos del estudio, como el url del cuestionario online autoadministrado (CAWI) a cumplimentar. También se señalaba lo siguiente, al objeto de poder tener en cuenta el hecho de que una persona estuviese trabajando o no actualmente, aspecto importante dado el contexto de crisis económica existente en España en la actualidad: "Algunas de las siguientes preguntas hacen referencia a la empresa en la que trabajas. Si tu situación actual es que no tienes trabajo, responde a ellas pensando en la última empresa para la que trabajaste". Finalmente, también se solicitaba la difusión de la encuesta, de tal manera que la técnica bola de nieve se pudiera desarrollar de modo incremental y continuo: "Te pido que lo difundas entre tus amigos y conocidos, para que pueda alcanzar un mínimo de respuestas válidas con el que poder realizar el análisis de datos".

La mayoría de preguntas del cuestionario se basan en investigaciones anteriores (INE, 2010, 2011; Torrent-Sellens et al., 2010a, 2010b; 2010c), y se muestran en la tabla 1.

Las variables de clasificación consideradas son:

- C1. Sexo
- C2. Edad
- C3. ¿Cuáles son tus estudios acabados de más alto grado?
- C4. Actividad principal desarrollada actualmente
- C5. ¿Cuál es la actividad principal de tu empresa?
- C6. Lugar de cumplimentación de la encuesta

La orientación hacia las denominadas fuentes co-innovadoras de las preguntas diseñadas, es la siguiente:

- usos TIC: P8, P9;
- capacitación y formación de los trabajadores: P8, P9, C3;
- innovación organizativa: P1, P2, P3, P4a), P4b), P5, P6, P7.

Asimismo, reseñamos la relación latente existente entre las variables analizadas y diversas actividades personales ligadas al entretenimiento y la sociabilidad online, como se destacaba en el marco analítico: P1, P2, P4a), P8.

En la tabla 2 se puede observar la ficha técnica de la encuesta realizada.

Antes del análisis de resultados, cabe reseñar que se ha efectuado un proceso de recodificación de las respuestas -tabla 3-, agrupándolas conforme a sus afinidades, y con vistas a la realización de un análisis de datos más optimizado al número de respuestas obtenidas.

Por otro lado, hay que hacer referencia al algoritmo utilizado para la realización de un análisis con el que comprobar qué reglas de asociación, entre las variables analizadas y los casos encontrados, aparecen como las más importantes y relevantes para la investigación, desde un punto de vista de complementar análisis descriptivos y mucho más típicos. El algoritmo es del tipo *Apriori*, que reduce iterativamente el

conjunto de casos necesarios para alcanzar el valor de confianza mínimo deseado (Agrawal; Srikant, 1994; Liu; Hsu; Ma, 1998).

Concretamente, el algoritmo busca relaciones entre los casos a analizar partiendo de todos los casos individuales posibles y contándolos, pasando después al cálculo del número de ocurrencias de subconjuntos -formados de un modo incremental: parejas, tríos, etc.- de la totalidad de casos posibles. El punto de interés del algoritmo se basa en que utiliza la información guardada de las ocurrencias de los subconjuntos ya calculados para eliminar, a priori y sin el pertinente conteo que haría falta, subconjuntos de casos que no cumplen los mínimos exigidos: si un subconjunto de casos no cumple los mínimos de ocurrencia deseada, cualquier subconjunto de casos más grande que lo contenga tampoco alcanzará dichos mínimos, por lo que no es tenido en cuenta para el conteo. Esta eliminación de subconjuntos de casos permite reducir el tiempo de cálculo, lo que acelera el proceso de obtención de los resultados.

4. Resultados

Los entrevistados suelen utilizar más medios de comunicación accesibles desde internet: un 87,0% utiliza medios de comunicación online, mientras que solamente un 13,0% utiliza medios tradicionales sin conexión a internet. Las características de la muestra, usuarios de internet, hace que esta utilización masiva de los medios de comunicación online sea claramente alta.

Las áreas de la empresa en las que se piensa que la información procedente de los medios de comunicación puede ser

más aprovechable y útil son dos: el ámbito de gestión, y el que tiene que ver con el I+D+i. Suman, entre ambos, más de tres cuartas partes (77,3%) del total de las respuestas.

Las áreas de la empresa en las que se piensa que la información de los medios de comunicación puede ser más útil son: gestión e I+D+i

Casi dos tercios (63,6%) de los encuestados hace un uso avanzado de las TIC. En cuanto a un uso medio de las TIC, la cifra se corresponde con el 31,8%. Además, las empresas de las personas encuestadas hacen un uso de las TIC mayoritariamente medio-avanzado: un 40,3% un uso avanzado, y casi la mitad (48,7%) un uso medio.

El análisis cruzado de las variables del cuestionario permite detectar una serie de tendencias:

- Prácticamente la totalidad de personas que utilizan medios de comunicación online suelen compartir las noticias que encuentran con sus compañeros de trabajo.
- Se observa una tendencia hacia el aprovechamiento efectivo de las noticias compartidas, tanto en personas que utilizan cualquier tipo de medios (online y offline), como en aquellas que solamente utilizan medios offline. No obstante, todavía existe campo para la mejora: 67 personas del total de la muestra de 154 piensan que la información compartida no es del todo aprovechada.
- Se detecta respecto a la frecuencia temporal con la que

Tabla 2. Ficha técnica y características sociodemográficas de los resultados. Encuesta online

Ficha técnica y características sociodemográficas de los resultados*	
Universo	Usuarios de internet
Selección de la muestra	Muestreo aleatorio no probabilístico: invitación mediante mensajes de correos electrónicos y de redes sociales (técnica bola de nieve).
Método de recolección de información	Cuestionario online con preguntas cerradas
Dimensión de la muestra	154 respuestas válidas
Error muestral	Margen de error del $\pm 7,90\%$ para los datos globales, con una población infinita, y en el caso de máxima indeterminación ($p=q=50\%$) y con un nivel de confianza del 95,5%
Período de trabajo de campo	29 de marzo de 2012 – 6 de mayo de 2012
Cuotas	Sin establecer
Ponderación	No aplicable
Respuesta por género	Mujeres: 61% Hombres: 39%
Edad media del encuestado	39 años
Estudios acabados de más alto grado	Universitarios: 84% Conforme a la distribución: - Educación universitaria: 35% - Educación universitaria postgrado y/o doctorado: 49%
Actividad principal desarrollada actualmente	Trabajador por cuenta ajena: 76%
Actividad principal de la empresa Basada en la Clasificación nacional de actividades económicas (CNAE-2009) (INE, 2007)	P. Educación: 29% Q. Actividades sanitarias y de servicios sociales: 14% O. Administración pública y defensa; Seguridad social obligatoria: 13% J. Información y comunicaciones: 8% M. Actividades profesionales, científicas y técnicas: 8% Otras: 29%
Lugar de cumplimentación de la encuesta	España: 87%

*Datos válidos a fecha 06-05-2012

- se comparten noticias de los medios de comunicación, que esta actividad es realizada de un modo casi cotidiano, ya sea diaria o semanalmente.
- Se observa una tendencia que busca el aprovechamiento de los medios de comunicación a nivel empresarial, aunque de un modo principalmente informal.
 - Tanto con medios de comunicación online como con tradicionales, se detecta una tendencia a su potencial aprovechamiento y utilidad en las áreas de gestión e I+D+i.
 - Como no podía ser de otro modo, la encuesta confirma que los usuarios con un nivel más avanzado de uso de las TIC suelen acceder a medios de comunicación online.
 - Los encuestados que acceden a medios de comunicación online suelen trabajar en empresas con un nivel medio-avanzado de uso de las TIC.

En la tabla 4 pueden verse todos estos aspectos de un modo sintético.

Con el objeto de complementar este análisis de datos, se ha utilizado el software *Weka Explorer* de la *University of Waikato* (Nueva Zelanda) (Hall et al., 2009a, 2009b) para el chequeo de las reglas de asociación más importantes, aplicando el algoritmo *Apriori*, comentado en el capítulo de metodología. Para el caso analizado en la investigación, los valores aplicados con dicho algoritmo han sido los siguientes: el número de casos mínimo=54, la confianza mínima=90%, y el número de iteraciones=13.

Concretamente, y aplicando este algoritmo, se puede afir-

Se puede generar un clima que ayude a compartir entre los empleados la información procedente de los medios de comunicación

mar que existe una tendencia con la que, si se dan los siguientes valores para las variables *input*, se generará en la empresa un clima que ayude a compartir entre los empleados información procedente de los medios de comunicación, y todo ello con una confianza superior al 96% (tabla 5). En la tabla 5 se han eliminado aquellas reglas que tenían un mayor grado de relación con el universo de la muestra, esto es, aquellas en las que dicho universo tenía una predisposición más alta para cumplir dichas reglas (un ejemplo: un universo formado por personas usuarias de internet hace que sean personas más usuarias de medios online, y que tengan un mayor nivel de conocimiento de las TIC).

Tabla 3. Recodificación de las preguntas y las respuestas

Preguntas	Valores	Nuevos valores
P1. Medios para acceder a las noticias	Radio Televisión Prensa escrita Revistas periódicas Otros	Tradicional
	Prensa digital Redes sociales Otros medios digitales	Online
P7. Ámbito más aprovechable y útil para el uso de la información procedente de los medios de comunicación	Administración/gestión/finanzas/RRHH Investigación + Desarrollo + Innovación	Gestión / I+D+i
	Infraestructural Compras/proveedores Producción de bienes y servicios Promoción/Marketing Ventas/distribución/postventa Otros	Otros
P8. Nivel de usuario TIC P9. Nivel de uso TIC en la empresa	Número de ítems señalados <=3	Básico
	3 < Número de ítems señalados <= 7	Medio
	Número de ítems señalados >7	Avanzado

De este modo, se detecta que los trabajadores tenderán a compartir las noticias que conocen, si tiene lugar al menos una de estas tres situaciones:

- los trabajadores perciben que la información compartida es aprovechada, tanto de un modo formal como informal;
- los ámbitos para su uso tienen que ver con la gestión y el I+D+i;
- la empresa tiene un nivel de uso de las TIC medio o avanzado.

5. Discusión y conclusiones

La investigación se propuso como objetivo conocer cómo afecta el grado de alfabetización mediática de los ciudadanos a las fuentes co-innovadoras de las empresas en las que trabajaban. Este objetivo ha sido verificado ya que se ha demostrado, mediante un estudio exploratorio, que un 96,1% de los trabajadores (sobre un total de 154 casos analizados mediante encuesta online) comparte en su lugar de trabajo informaciones y noticias procedentes de los medios de comunicación. Además, el 56,5% de estos consideran que esta información compartida ha sido aprovechada por parte de la empresa, aunque bajo una cultura empresarial que la ha sabido rentabilizar de un modo informal.

Se ha detectado una tendencia hacia la relación entre la alfabetización mediática y una de las tres fuentes co-innovadoras planteadas en el estudio: la relativa a la innovación organizacional que, asimismo, tiene mucha relación con parámetros relacionados con gestión del conocimiento a nivel empresarial.

Por otro lado, a nivel empresarial en España, y teniendo en cuenta las limitaciones del muestreo aleatorio no probabilístico utilizado, se ha detectado una tendencia consistente en que los trabajadores que más medios online consumen, tienden a compartir más el conocimiento que obtienen de dichos medios con sus compañeros de trabajo. Esto abre un

Tabla 4. Resultados del análisis de datos con las variables recodificadas

Análisis univariable de las variables básicas			
P1. Medios para acceder a las noticias	Online: 87,0% Tradicional: 13,0%		
P7. Ámbito más aprovechable y útil para el uso de la información procedente de los medios de comunicación	Gestión/I+D+i: 77,3% Otros: 22,7%		
P8. Nivel de usuario TIC	Avanzado: 63,6% Medio: 31,8% Básico: 4,6%		
P9. Nivel de uso TIC en la empresa	Avanzado: 40,3% Medio: 48,7% Básico: 11,0%		
Análisis cruzado de variables			
Se facilitan los datos en valores absolutos, ya que la dimensión de la muestra (154 respuestas válidas) podría generar malinterpretaciones, si se trabajase con datos porcentuales relativos.			
P1. Medios para acceder a las noticias cruzado con P2. Noticias compartidas	Número de casos	Online	Tradicional
	Sí comparten	132	16
	No comparten	2	4
P1. Medios para acceder a las noticias cruzado con P3. Información compartida aprovechada	Número de casos	Online	Tradicional
	Sí aprovechada	75	12
	No aprovechada	59	8
P1. Medios para acceder a las noticias cruzado con P5. Frecuencia de compartir	Número de casos	Online	Tradicional
	Diariamente	27	6
	Semanalmente	57	3
	Mensualmente	10	1
	Ocasionalmente	39	10
	Otros	1	0
P1. Medios para acceder a las noticias cruzado con P6. Cultura de aprovechar las noticias	Número de casos	Online	Tradicional
	Sí, formal	14	3
	Sí, informal	79	10
	No	38	6
	Otros	3	1
P1. Medios para acceder a las noticias cruzado con P7. Ámbito más aprovechable y útil para el uso de la información procedente de los medios de comunicación	Número de casos	Online	Tradicional
	Gestión/I+D+i	104	15
	Otros	30	5
P1. Medios para acceder a las noticias cruzado con P8. Nivel de usuario TIC	Número de casos	Online	Tradicional
	Avanzado	88	10
	Medio	42	7
	Básico	4	3
P1. Medios para acceder a las noticias cruzado con P9. Nivel de uso TIC en la empresa	Número de casos	Online	Tradicional
	Avanzado	55	7
	Medio	67	8
	Básico	12	5

campo en el que las empresas puedan obtener potenciales réditos, si saben aprovechar y modular formalmente estas conductas de sus trabajadores.

Un 77,3% del total de encuestados piensa que los medios de comunicación pueden ser utilizados de un modo aprovechable y útil en ámbitos empresariales asociados a la gestión y al I+D+i, por lo que dichas áreas empresariales se perfilan, bajo un esquema de tendencias, como las grandes beneficiarias de un aprovechamiento de los niveles de alfabetización mediática existente en los trabajadores.

La investigación también abre el camino para mostrar as-

pectos que van más allá del aprovechamiento de los niveles de alfabetización digital o informacional existente en los trabajadores de las empresas, mostrando que la alfabetización mediática, tradicionalmente desarrollada con objetivos sociales y democráticos, puede tener un gran campo de utilización a nivel empresarial. Los trabajadores que tengan mejores competencias mediáticas están en mejor situación para contribuir a una mejora de la gestión del conocimiento y, con ello, poder aportar y contribuir al desarrollo de la innovación en sus empresas, ayudando a su evolución, viabilidad y sostenibilidad, ámbitos necesarios para el desarrollo de la sociedad de la información.

Tabla 5. Reglas de asociación

Variable input (X)		Variable output (Y)		Casos input	Casos output	Confianza P(Y X)
P3. Información compartida aprovechada	Sí	P2. Noticias compartidas	Sí	87	86	0,99
P6. Cultura de aprovechar noticias	Sí, pero de una manera informal			89	88	0,99
P7. Ámbito más aprovechable y útil para el uso de la información procedente de los medios de comunicación	Gestión/I+D+i			119	115	0,97
P9. Nivel de uso TIC en la empresa	Avanzado			62	61	0,98
	Medio			75	72	0,96

Es especialmente interesante recordar (INE, 2013) que un 95,7% y un 37,9% de las empresas con menos de 10 empleados y con conexión a internet, utiliza la Red para buscar información y para observar el mercado, respectivamente, porcentajes que alcanzan el 96,8% y el 51,9%, respectivamente, para el resto de empresas con conexión a internet, pero con mayor número de trabajadores. Estos datos ayudan a contextualizar la importancia de las tendencias observadas mediante la presente investigación, puesto que los medios de comunicación ayudan a satisfacer dichas necesidades empresariales, y la alfabetización mediática de los trabajadores es una de las herramientas para resolverlas.

La alfabetización mediática puede tener un gran campo de utilización a nivel empresarial

Por otro lado, y teniendo en cuenta las relaciones establecidas entre las preguntas de investigación y las componentes propuestas por organismos internacionales como la Unesco y la Comisión Europea, para la medición de los niveles de alfabetización mediática e informacional a nivel de países, es muy interesante observar que los resultados presentan un gran y potencial uso de los medios de comunicación en la empresa. Este hecho legitima -dentro del contexto de haber realizado una investigación exploratoria- la necesidad de tener en cuenta a las organizaciones empresariales como sujeto de análisis para la medición de las componentes asociadas al acceso y uso a la información y los medios de comunicación, conforme al marco de trabajo planteado por la Unesco.

Respecto a la metodología utilizada, el sistema de indicadores planteado, basado en diversas investigaciones puestas anteriormente en marcha, puede ofrecer un buen punto de partida con el que poder trabajar la medición del grado de uso de los medios de comunicación en entornos empresariales, así como de los niveles de competencias mediáticas existentes en los trabajadores. Considerando el grado de encuesta experimental que hemos asignado al estudio, bajo una metodología exploratoria, la longitud y formato del cuestionario diseñado (9 preguntas realizadas mediante cuestionarios online) permite que pueda ser fácilmente aceptada por los trabajadores de las empresas. Por otro lado, las diferencias de valores para los resultados obtenidos, pese a poder variar en una horquilla de $\pm 7,90\%$ (error

para los datos globales, entendiendo que es una muestra no representativa del conjunto de trabajadores españoles), nos permiten detectar unas primeras tendencias válidas para poblaciones semejantes, esto es, trabajadores conectados a internet, de entre 30 y 40 años, y con una formación universitaria.

Finalmente, y a la vista de todo esto, aparece como clave el hecho de que las áreas empresariales orientadas a la gestión de la formación de los trabajadores tengan que tener en cuenta el ámbito de la alfabetización mediática -o educación en medios-, como uno de los campos en los que fortalecer las competencias de los trabajadores.

6. Bibliografía

Aenor (2011). *UNE 166006:2011. Gestión de la I+D+i: Sistema de vigilancia tecnológica e inteligencia competitiva.*

Agrawal, Rakesh; Srikant, Ramakrishnan (1994). "Fast algorithms for mining association rules in large databases". En: *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 478-499. <http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>

Aldasoro-Alustiza, Juan-Carlos; Cantonnet-Jordi, María Luisa; Cilleruelo-Carrasco, Ernesto (2012). "La vigilancia tecnológica y la inteligencia competitiva en los estándares de gestión de la calidad en I+D+i". En: *XVI Congreso de ingeniería de organización*. Vigo, 18 a 20 de julio de 2012, pp. 1162-1168. http://adingor.es/congresos/web/uploads/cio/cio2012/SP_04_Gestion_Innovacion_Tecnologica_y_Organizativa//1162-1168.pdf

Anderson, Gavin; Olming, Karl-Oskar; MacFarquhar, Nicolas (2008). *Integrating mass media in small enterprise development: current knowledge and good practices*. Geneva: International Labour Office. ISBN: 978 9221211426; 978 9221211433 (web pdf) http://www.ilo.org/wcmsp5/groups/public/@ed_emp/documents/publication/wcms_105072.pdf

Babbie, Earl (2007). *The practice of social research*. Eleventh edition. Belmont, California: Thomson Wadsworth. ISBN: 0 495 09325 4

Bindé, Jérôme (2005). *Towards knowledge societies*. Paris: UNESCO. ISBN: 978 92 3 104000 9; 92 3 204000 x (eng) <http://unesdoc.unesco.org/ulis/cgi-bin/ulis.pl?catno=141843>

- Burroughs, Susie; Brocato, Kay; Hopper, Peggy F.; Sanders, Angela** (2009). "Media literacy: a central component of democratic citizenship". *The educational forum*, v. 73, n. 2, pp. 154–167.
<http://dx.doi.org/10.1080/00131720902739627>
- Castells, Manuel** (1996). *The rise of the network society*. Oxford and Malden, Massachusetts: Blackwell Publishers. ISBN: 978 1557866172
- Celot, Paolo; Pérez-Tornero, José-Manuel** (2009). *Study on assessment criteria for media literacy levels*. Brussels: EAVI – European Association for Viewers' Interests.
http://www.eavi.eu/joomla/images/stories/Publications/Study2_Assessment/mlstudy2/eavi_study_on_assessment_criteria_for_media_literacy_levels_in_europe.pdf
- Cohen, Wesley M.; Levinthal, Daniel A.** (1989). "Innovation and learning: the two faces of R&D". *The Economic Journal*, 99, pp. 569–596.
<http://www.uzh.ch/iou/orga/ssl-dir/wiki/uploads/Main/v27.pdf>
- Cuevas-Cerveró, Aurora; García-Moreno, María-Antonia** (2010). "Ideas, un modelo de evaluación para inclusión digital y alfabetización informacional orientado a salud". *El profesional de la información*, v. 19, n. 3, pp. 240–245.
<http://dx.doi.org/10.3145/epi.2010.may.03>
- European Commission* (2006). "Recomendación del Parlamento Europeo y del Consejo de 18 de diciembre de 2006 sobre las competencias clave para el aprendizaje permanente (2006/962/CE)". *Diario oficial de la Unión Europea* 30-12-2006.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:ES:PDF>
- European Commission* (2009). *Commission Recommendation 2009/625/EC of 20 August 2009 on media literacy in the digital environment for a more competitive audiovisual and content industry and an inclusive knowledge society*.
<http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009H0625>
- Fierro-Moreno, Eréndira; Mercado-Salgado, Patricia; Cernas-Ortiz, Daniel-Arturo** (2013). "El efecto de la cultura centrada en el conocimiento y la interacción social en la innovación organizativa. El efecto mediador de la gestión del conocimiento". *Esic market economics and business journal*, mayo-agosto, v. 44, n. 2, pp. 87–108.
http://www.esic.edu/documentos/revistas/esicmk/130510_122815_E.pdf
- Hall, Mark; Frank, Eibe; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H.** (2009a). "The WEKA data mining software: an update". *ACM SIGKDD explorations newsletter*, v. 11, n. 1, pp. 10–18.
<http://dx.doi.org/10.1145/1656274.1656278>
- Hall, Mark; Frank, Eibe; Holmes, Geoffrey; Pfahringer, Bernhard; Reutemann, Peter; Witten, Ian H.** (2009b). *Weka 3 - Data mining with open source machine learning software in Java*.
<http://www.cs.waikato.ac.nz/ml/weka>
- Hernández-Sampieri, Roberto; Fernández-Collado, Carlos; Baptista-Lucio, María del Pilar** (2010). *Metodología de la investigación*. Quinta edición. México: McGraw-Hill. ISBN: 978 607 15 0291 9
https://www.u-cursos.cl/fau/2013/2/DGH-406/1/foro/r/Metodologia_de_la_investigacion,_5ta_Edicion_-_Sampieri.pdf
- Instituto Nacional de Estadística* (2007). CNAE 2009. *Clasificación nacional de actividades económicas*.
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t40/clasrev&file=inebase>
- Instituto Nacional de Estadística* (2010). *Encuesta sobre el uso de TIC y comercio electrónico en las empresas*. Cuestionarios.
<http://www.ine.es/metodologia/t09/t09e02cues.htm>
- Instituto Nacional de Estadística* (2011). *Encuesta de uso de TIC y comercio electrónico (CE) en las empresas 2010-2011*.
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft09%2Fe02&file=inebase&L=0>
- Instituto Nacional de Estadística* (2013). *Encuesta de uso de TIC y comercio electrónico (CE) en las empresas 2012-2013*.
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft09%2Fe02&file=inebase&L=0>
- Jaguszewski, Janice M.; Williams, Karen** (2013). *New roles for new times: transforming liaison roles in research libraries*. Washington: Association of Research Libraries.
<http://www.arl.org/storage/documents/publications/NRNT-Liaison-Roles-final.pdf>
- Koltay, Tibor** (2011). "The media and the literacies: media literacy, information literacy, digital literacy". *Media, culture & society*, v. 33, n. 2, pp. 211–221.
<http://dx.doi.org/10.1177/0163443710393382>
- Kubátová, Jaroslava** (2011). "Knowledge sharing culture for the net generation". *Journal of international scientific publications: economy & business*, v. 5, n. 2, pp. 181–194.
<http://www.scientific-publications.net/download/economy-and-business-2011-2.pdf>
- Lau, Jesús; Cortés, Jesús** (2009). "Habilidades informativas: convergencia entre ciencias de información y comunicación". *Comunicar*, v. XVI, n. 32, pp. 21–30.
<http://www.revistacomunicar.com/index.php?contenido=etalles&numero=32&articulo=32-2009-06>
<http://dx.doi.org/10.3916/c32-2009-02-001>
- Liu, Bing; Hsu, Wynne; Ma, Yiming** (1998). "Integrating classification and association rule mining". En: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 80–86.
http://www.cs.rit.edu/~dmrg/dm_winter/reading/integ_class_arm.pdf
- Livingstone, Sonia** (2003). "The changing nature and uses of media literacy". *MEDIA@LSE electronic working papers*, n. 4.
http://eprints.lse.ac.uk/13476/1/The_changing_nature_and_uses_of_media_literacy.pdf
- McLuhan, Marshall** (1996). *Comprender los medios de comunicación: las extensiones del ser humano*. Barcelona: Paidós. ISBN: 8449302404

http://cedoc.infod.edu.ar/upload/McLuhan_Marshall___Comprender_los_medios_de_comunicacion.pdf

McLuhan, Marshall; Powers, Bruce R. (1989). *La aldea global: transformaciones en la vida y los medios de comunicación mundiales en el siglo XXI*. Barcelona: Gedisa, ISBN: 978 8474324037

Moeller, Susan (2009). "Fomentar la libertad de expresión con la alfabetización mediática mundial". *Comunicar*, v. XVI, n. 32, pp. 65–72.

<http://rabida.uhu.es/dspace/bitstream/handle/10272/2812/b15488652.pdf?sequence=1>
<http://dx.doi.org/10.3916/c32-2009-02-005>

Organización de las Naciones Unidas (1948). Declaración universal de derechos humanos.

<http://www.un.org/es/documents/udhr>

Ortoll-Espinet, Eva (2004). "Competencias profesionales y uso de la información en el lugar de trabajo". *El profesional de la información*, 2004, septiembre-octubre, v. 13, n. 5, pp. 338-345.

<http://www.elprofesionalde lainformacion.com/contenidos/2004/septiembre/2.pdf>

Palop, Fernando; Vicente, José Miguel (1999). *Vigilancia tecnológica e inteligencia competitiva: su potencial para la empresa española*. Madrid: Fundación Cotec.

http://www.delfos.co.cu/boletines/bsa/pdf/potencial_vtec.pdf

Pérez-Tornero, José-Manuel; Celot, Paolo; Varis, Tapio (2007). *Current trends and approaches to media literacy in Europe*. Media Literacy Studies. Brussels: European Commission.

http://www.mediamillion.com/wp-content/uploads/2011/05/Estudio_Current-trends-and-approaches-ML-in-Europe.pdf

Saleh, Ibrahim (2009). "La alfabetización mediática en Medio Oriente y Norte de África: más allá del círculo vicioso del oxímoron". *Comunicar*, v. XVI, n. 32, pp. 119–129.

<http://www.redalyc.org/pdf/158/15812476015.pdf>
<http://dx.doi.org/10.3916/c32-2009-02-010>

Sánchez-Navarro, Jordi; Aranda, Daniel (2011). "Internet como fuente de información para la vida cotidiana de los jóvenes españoles". *El profesional de la información*, v. 20, n. 1, pp. 32-37.

<http://eprints.rclis.org/15332/1/032-037.pdf>
<http://dx.doi.org/10.3145/epi.2011.ene.04>

Statistics Canada (2010). *Survey methods and practices*. Ottawa, Canada: Minister of Industry. ISBN: 978 1 100 16410 6
<http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.pdf>

The White House (2009). *Presidential proclamation national information literacy awareness month*. The White House.

http://www.whitehouse.gov/the_press_office/Presidential-Proclamation-National-Information-Literacy-Awareness-Month

Torrent-Sellens, Joan (2002). "De la nueva economía a la economía del conocimiento: hacia la tercera revolución in-

dustrial". *Revista de economía mundial*, v. 7, pp. 39–68.

<http://rabida.uhu.es/dspace/bitstream/handle/10272/418/b1215529.pdf?sequence=1>

Torrent-Sellens, Joan (coord.) (2008). *La empresa red. Tecnologías de la información y la comunicación, productividad y competitividad*. Barcelona: Ariel. ISBN: 978-84-344-4277-1

Torrent-Sellens, Joan (2010). "Competitivitat internacional. Cap a les noves fonts coinnovadores del creixement industrial?". En: Albalade-del-Sol, Daniel, et al. *La indústria catalana després de la crisi* (Papers d'Economia Industrial, 32).

Barcelona, Departament d'Innovació, Universitats i Empresa (DIUE), pp. 177-194. ISBN: 978 8439386261

http://www.gencat.cat/diue/doc/doc_18907310_1.pdf

Torrent-Sellens, Joan; Díaz-Chao, Ángel; Ficapal Cusí, Pilar; Jiménez-Zarco, Ana Isabel; Viñas-Bardolet, Clara; Soler, Ivan; Ruiz, Mònica; Martínez-Cerdá, Juan-Francisco; Moré, Tomàs (2010a). *TIC Microempreses comerç. Estudi sobre l'adopció i l'ús de les TIC a les microempreses del sector del comerç, 2010. Estructura i elements de valor de la microempresa comercial a Catalunya*. Barcelona: FOBSIC (Fundació Observatori per a la Societat de la Informació de Catalunya).
http://i2tic.net/media/Informe_microempreses_comerc.pdf

Torrent-Sellens, Joan; Díaz-Chao, Ángel; Ficapal Cusí, Pilar; Jiménez-Zarco, Ana Isabel; Viñas-Bardolet, Clara; Soler, Ivan; Ruiz, Mònica; Martínez-Cerdá, Juan-Francisco; Moré, Tomàs (2010b). *TIC Microempreses Turisme. Estudi sobre l'adopció i l'ús de les TIC a les microempreses del sector del turisme, 2010. Estructura i elements de valor de la microempresa turística a Catalunya*. Barcelona: FOBSIC (Fundació Observatori per a la Societat de la Informació de Catalunya).
http://i2tic.net/media/Presentacio_microempreses_turisme.pdf

Torrent-Sellens, Joan; Díaz-Chao, Ángel; Jiménez-Zarco, Ana-Isabel; Ficapal-Cusí, Pilar; Soler, Ivan; Ruiz, Mònica; Martínez-Cerdá, Juan-Francisco; Rodon, Ana (2010c). *El sector empresarial de les TIC a Catalunya (II). Estructura i elements de valor de l'empresa TIC a Catalunya. Anàlisi de la font primària d'informació*. Barcelona: FOBSIC (Fundació Observatori per a la Societat de la Informació de Catalunya).
http://i2tic.net/media/Sector_TIC_2009_Segon_Lliurable.pdf

Torrent-Sellens, Joan; Ficapal-Cusí, Pilar (2010a). "TIC, co-innovación y productividad empresarial: evidencia empírica para Cataluña y comparación internacional de resultados". *Revista de economía mundial*, v. 26, pp. 203–233.
http://rabida.uhu.es/dspace/bitstream/handle/10272/4872/TIC_co-innovacion_productividad.pdf?sequence=2

Torrent-Sellens, Joan; Ficapal-Cusí, Pilar (2010b). "¿Nuevas fuentes co-innovadoras de la productividad empresarial?". *Innovar*, v. 20, n. 38, pp. 111–124.
<http://www.revistas.unal.edu.co/index.php/innovar/article/view/22295>

Torrent-Sellens, Joan; Ficapal-Cusí, Pilar (2011). "TIC, cualificación, organización y productividad del trabajo: un análisis empírico sobre las nuevas fuentes de la eficiencia empresarial en Cataluña". *Investigaciones Regionales*, 20, pp. 93-115.
<http://www.redalyc.org/articulo.oa?id=28920705005>

Unesco (2013a). *Conceptual relationship of information literacy and media literacy in knowledge societies*, Paris: Unesco. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/wsis/WSIS_10_Event/WSIS_-_Series_of_research_papers_-_Conceptual_Relationship_between_Information_Literacy_and_Media_Literacy.pdf

Unesco (2013b). *Global media and information literacy assessment framework: country readiness and competencies*, Paris: UNESCO. ISBN: 978 92 3 001221 2 <http://unesdoc.unesco.org/images/0022/002246/224655e.pdf>

United Nations (2005). *Understanding knowledge societies: in twenty questions and answers with the index of knowl-*

edge societies. New York: United Nations Publications. ISBN: 9211091454

<http://unpan1.un.org/intradoc/groups/public/documents/un/unpan020643.pdf>

Vilaseca-Requena, Jordi; Torrent-Sellens, Joan; Cabañero-Pisa, Carlos F.; Castillo-Merino, David; Colomé-Perales, Rosa; Díaz-Chao, Ángel; Ficapal-Cusí, Pilar; Jiménez-Zarco, Ana-Isabel; Lladós-Masllorens, Josep; Martínez-Argüelles, María-Jesús; Meseguer-Artola, Antoni; Plana-Erta, Dolors; Rodríguez-Ardur, Inma (2003). *Las TIC y las transformaciones de la empresa catalana. Informe de investigación II* (documento de síntesis). Barcelona: Universitat Oberta de Catalunya.

http://www.uoc.edu/in3/pic/esp/pdf/PIC_empresa_abs_esp.pdf

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>

UNIVERSIDAD DE
MURCIA



CENTUM
CIEN AÑOS DE LA UNIVERSIDAD DE MURCIA
1915 | 2015

El profesional de la
información

ANUARIO

Think
EPI

CREDIBILIDAD DE LOS MEDIOS DE COMUNICACIÓN: ANÁLISIS DE LA PRENSA DIARIA DESDE EL COMPORTAMIENTO DEL CONSUMIDOR

Cristina Calvo-Porrál, Valentín-Alejandro Martínez-Fernández y Óscar Juanatey-Boga



Cristina Calvo-Porrál es licenciada en administración y dirección de Empresas por el *Icade* de Madrid (*Universidad Pontificia de Comillas*), y licenciada en investigación y técnicas de mercado por el mismo centro. Es doctora en economía por la *Universidad de A Coruña* y cuenta con numerosas publicaciones en el área de marketing. Su carrera profesional se ha desarrollado durante más de diez años en la industria de la moda, como responsable de Expansión Internacional de *Carolina Herrera* y *Purificación García*. Actualmente es profesora del *Área de Comercialización e Investigación de Mercados* de *Facultad de Economía y Empresa* de la *Universidad de A Coruña*.
<http://orcid.org/0000-0002-5470-0729>

Universidad de A Coruña
Facultad de Economía y Empresa
Campus Elviña, s/n. 15071 A Coruña, España
ccalvo@udc.es



Valentín-Alejandro Martínez-Fernández es licenciado en ciencias de la información por la *Universidad Complutense de Madrid (UCM)*. Master MBA en dirección y administración de empresas por la *Universidad de A Coruña (UDC)*. Doctor en ciencias de la información por la UCM y profesor titular de universidad en la *Facultad de Ciencias de la Comunicación* y en la *Facultad de Economía y Empresa* de la UDC. Ha sido director del diario *El ideal gallego* así como directivo en otras empresas informativas. Es autor de artículos sobre medios de comunicación y estrategias de comunicación.
<http://orcid.org/0000-0003-0069-675X>

Universidad de A Coruña
Facultad de Ciencias de la Comunicación
Campus de Elviña, s/n. 15071 A Coruña, España
valejand@udc.es



Óscar Juanatey-Boga es licenciado y doctor en ciencias económicas y empresariales por la *Universidad de A Coruña (UDC)*. Master MBA en dirección y administración de empresas, master en dirección comercial y marketing, y master en comunicación empresarial por la UDC. Profesor del *Área de Comercialización e Investigación de Mercados* en la *Facultad de Economía y Empresa* de la misma universidad, es autor de artículos sobre medios de comunicación y estrategias de comunicación.
<http://orcid.org/0000-0001-5226-2748>

Universidad de A Coruña
Facultad de Economía y Empresa
Campus Elviña, s/n. 15071 A Coruña, España
oscarjb@udc.es

Resumen

Se analiza la credibilidad de los medios de comunicación, y en concreto de los diarios generalistas. Mediante el empleo de un modelo de ecuaciones estructurales, y con una muestra de 454 cuestionarios sobre los diarios *El país*, *El mundo*, *Abc*, *La razón* y *La voz de Galicia*, se estudian los antecedentes y consecuencias de su credibilidad. Los resultados muestran la importante influencia de la calidad percibida, la imagen y la notoriedad del medio en la actitud de los consumidores, así como la influencia de la actitud sobre la credibilidad de que goza el medio. Los resultados conllevan interesantes implicaciones para los gestores de medios de comunicación, que deben priorizar la imagen y notoriedad del medio, así como la calidad que perciben los consumidores de información en un entorno de competencia creciente.

Palabras clave

Credibilidad, Actitud, Medios de comunicación, Prensa escrita, Diarios, Consumidores, Usuarios, Estudios de usuarios, Cuestionarios.

Title: Media credibility: analysis of the daily press based on consumer behaviour

Artículo recibido el 01-10-2013
Aceptación definitiva: 28-01-2014

Abstract

We analysed mass media credibility, focusing particularly on generalist journals. We carried out structural equation modeling, based on a sample comprised of 454 questionnaires relative to the newspapers *El país*, *El mundo*, *Abc*, *La razón* and *La voz de Galicia*, in order to analyze the antecedents and consequences of credibility. The results obtained show the important influence of perceived quality, media image and awareness on consumers' attitudes, along with the influence of attitude on media credibility. Our findings provide useful insights for mass media managers that may enhance media image and awareness, along with consumers' perceptions of quality, in an increasingly competitive environment..

Keywords

Credibility, Attitude, Mass media, Newspapers, Consumers, Users, User studies, Questionnaires.

Calvo-Porrá, Cristina; Martínez-Fernández, Valentín-Alejandro; Juanatey-Boga, Óscar (2014). "Credibilidad de los medios de comunicación: análisis de la prensa diaria desde el comportamiento del consumidor". *El profesional de la información*, marzo-abril, v. 23, n. 3, pp. 300-309.

<http://dx.doi.org/10.3145/epi.2014.may.10>

1. Introducción

En las dos últimas décadas se advierte una gran proliferación de nuevos medios de comunicación social, así como una gran fragmentación de las audiencias, lo que ha cambiado significativamente el escenario de la industria de la comunicación en España (Chan-Olmsted; Yung-Wook, 2001). Se trata de un entorno altamente competitivo, en el cual la gestión de marca es para los medios una estrategia necesaria para crear y establecer su credibilidad (Vila-López; Küster-Boluda; Bigné, 2013).

Muchos consumidores no pueden juzgar si una información es verdadera o no, o si está manipulada o no, de modo que han de confiar en el medio de comunicación. La credibilidad se convierte así en uno de los criterios esenciales en la selección de un medio, lo que provoca que los poco creíbles sean rechazados o no tengan incidencia en la conformación de opinión. Que la credibilidad influya en el comportamiento del consumidor y sea un factor decisivo en el éxito periodístico y comercial de un medio de comunicación (Schweiger, 2000) lleva a plantearnos cuáles son sus variables determinantes, el proceso para evaluarla, y las características del medio que influyen en su valoración.

Hay pocos trabajos sobre credibilidad de los medios de comunicación desde la perspectiva del consumidor de información (Fombrun; Van-Riel, 2004; Rathnayake, 2008; Walsh et al., 2009). Y en este contexto se plantea el presente trabajo, con el objetivo de analizar los antecedentes y las consecuencias de la credibilidad de los medios de comunicación -más específicamente de los diarios generalistas de referencia-, desde la perspectiva del consumidor. Para ello se ha diseñado la investigación tomando como referencia los diarios *El país*, *El mundo*, *Abc*, *La razón* y *La voz de Galicia*.

2. Revisión de la literatura

2.1. Concepto de credibilidad de los medios

La credibilidad es un concepto multidimensional (Gaziano; McGrath, 1986) y ha sido analizado tanto por la psicología social como por las ciencias de la comunicación. Si bien para la primera, la credibilidad es una variable subjetiva, defini-

da por la percepción del receptor de la información (Tseng; Fogg, 1999), las ciencias de la comunicación señalan el carácter objetivo de la credibilidad, definida como el grado o nivel de precisión y exactitud de una información (Greer, 2003; Flanagin; Metzger, 2000). Hovland, Janis y Kelley (1953) propusieron el concepto de credibilidad de mayor influencia en la literatura (Erdem; Swait, 2004; Trimble; Rifon, 2006), señalando como sus dos componentes a la experiencia y la confianza. Definieron la credibilidad de una marca o producto, en este caso de un medio de comunicación, como la percepción del consumidor de su veracidad, confianza, buena voluntad y también de su experiencia.

“ La credibilidad es un criterio esencial en la selección de un medio, lo que provoca que los poco creíbles sean rechazados o no tengan incidencia en la conformación de opinión ”

La credibilidad del medio de comunicación se refiere al efecto que sus características concretas tienen sobre la credibilidad de sus mensajes, e igualmente a la carencia de un interés particular a la hora de transmitir información (Metzger et al., 2003). De acuerdo con Wathen y Burkell (2002), un medio de comunicación posee credibilidad si su contenido y sus mensajes son considerados válidos y de confianza. Por tanto su credibilidad se refiere a su buen hacer en términos de confiabilidad desde el punto de vista de los consumidores (Newell; Goldsmith, 2001). Autores como Nozato (2002) o Flanagin y Metzger (2007) definen los factores que permiten evaluar la credibilidad de un medio y entre ellos destacan experiencia, objetividad, dinamismo, equidad, ausencia de sesgo, integridad y exactitud, e igualmente fiabilidad y reputación.

En el caso concreto de la prensa, Hovland, Janis y Kelley (1953) identificaron dos factores principales de la credibilidad -confianza y experiencia-, además de señalar una serie de dimensiones subyacentes como son: la cualificación del medio, su dinamismo, su conocimiento y competencia, su

precisión e imparcialidad. Finalmente, **Gaziano y McGrath** (1986) indican que la credibilidad de los diarios está formada por la información fiable y veraz, la ausencia de sesgo, la separación clara de información y opinión, la independencia de los intereses y medios que financian el diario, así como ofrecer la información de forma completa, mostrando los hechos tal y como son –contando la historia completa–.

La credibilidad de los medios de comunicación es similar a la de las marcas de productos o servicios, dado que es muy sencillo para el consumidor atribuir credibilidad a la marca de un medio reconocido (**Schweiger**, 2000). El consumidor de información es quien determina, en su rol de consumidor de noticias, la credibilidad de un medio. De esta manera, **Liu** (2004) define la valoración de la credibilidad de un medio como el proceso cognitivo mediante el que la información es filtrada y seleccionada por el consumidor; proceso en el cual las creencias del individuo, así como la actualidad, veracidad y calidad de la información tienen un impacto positivo. Por consiguiente, numerosos autores afirman que los consumidores tienden a escoger aquellos medios que perciben como más creíbles (**Taylor**, 1986; **Rieh**; **Danielson**, 2007).

La credibilidad de un medio de comunicación la determina el receptor, esto es, el consumidor de información

Comportamiento del consumidor de información

La *Teoría del procesamiento de la información* (**Sughan**, 1980) afirma que los consumidores toman decisiones en un contexto de riesgo e incertidumbre, y que desean reducir el coste asociado a sus decisiones de consumo. En este contexto se justifica que los consumidores escojan marcas (cabezas) que gozan de credibilidad, para simplificar su proceso de compra y reducir el riesgo a equivocarse (**Erdem**; **Swait**, 1998). En el consumo de información, el consumidor tiene dos opciones: optar por creer lo que dicen los medios, sabiendo que tienen cierto sesgo informativo, en ocasiones neutralizado por la identificación con la línea editorial, o bien adoptar una postura crítica respecto a los mismos. Sin embargo, son pocos los consumidores que invierten tiempo en la formación de una opinión crítica, y por tanto tienden a asumir como cierto y creíble aquello que los medios de comunicación publican (**Piñeros-Piza et al.**, 2011). Por tanto las creencias de los consumidores y su percepción, son variables que influyen en la credibilidad otorgada a un medio de comunicación (**Blackwell**; **Engel**; **Miniard**, 2002).

De acuerdo a lo señalado, la credibilidad de un medio de comunicación la determina el receptor, esto es, el consumidor de información. Por ello el presente trabajo se ha planteado desde el enfoque del consumidor, con el objetivo de analizar los antecedentes –calidad percibida, conciencia social e imagen y notoriedad– y las consecuencias de la credibilidad de los medios de comunicación –fidelidad y aceptación de nuevos productos–, para un tipo determinado de medio como son los diarios generalistas de referencia.

2.2. Antecedentes y consecuencias de la credibilidad de los diarios

Calidad percibida

Se entiende como calidad percibida de un medio de comunicación a la capacidad del medio para atender a una audiencia diversa, ofrecer contenidos relevantes y de actualidad, promover valores, generar conciencia y espíritu crítico, respetar las leyes y la cultura, o estimular la creatividad (**Webster**, 1986). Siguiendo a **Wang y Soergel** (1998), puede definirse como la valoración de la bondad de un medio en términos de calidad periodística. Por otro lado se relaciona con la respuesta subjetiva del consumidor hacia las características de un producto o servicio, y de esta manera, referida a un medio, se puede definir como el criterio del consumidor en cuanto a la excelencia y la veracidad del mismo (**Taylor**, 1986).

En este sentido **Liu** (2004) estima que la calidad percibida del medio de comunicación tiene un impacto positivo en su credibilidad. De acuerdo con todo ello, se plantea la siguiente hipótesis:

H_0 : La calidad percibida del diario tiene una influencia positiva en la actitud hacia el mismo

Conciencia social

En la línea de **Hovland, Janis y Kelley** (1953) y la de **Gaziano y McGrath** (1986), puede considerarse que la conciencia social incluye factores como el comportamiento ético y moral del medio en el tratamiento de la información, su compromiso con el interés público, así como la consideración de la opinión y punto de vista de los lectores. Además, estudios previos demuestran una relación positiva entre la conciencia social del medio y la actitud mostrada por los consumidores (**Gaziano**; **McGrath**, 1986). Consecuentemente se plantea la siguiente hipótesis:

H_1 : La conciencia social del diario tiene una influencia positiva en la actitud hacia el mismo

Imagen y notoriedad

Los consumidores asocian la credibilidad de un medio con su imagen general (**Haley**, 1996), al entender por tal la imagen efectiva que tienen de un medio concreto y que va más allá de la información que proporciona (**Boulding**, 1969). De esta manera la imagen, las asociaciones y el simbolismo asociado a un medio de comunicación contribuyen a su credibilidad e influyen en la actitud de los consumidores, dado que transmiten información sobre el mismo (**Gotsi**; **Wilson**, 2001). La imagen incluye la estética del contenido y la información emitida por el medio y las acciones llevadas a cabo por cualquier área de la empresa de comunicación (**Carrillo**; **Castillo-Díaz**; **Tato-Jiménez**, 2008). Numerosos autores afirman que las imágenes y asociaciones influyen en la credibilidad percibida del medio, y que los consumidores toman decisiones sobre su credibilidad basándose en asociaciones, imágenes o señales (**Rieh**; **Danielson**, 2007). Por otro lado diversos autores coinciden en que la notoriedad de un medio de comunicación, su familiaridad para el consumidor o su capacidad para reconocerlo y distinguirlo, también influyen en la actitud hacia el mismo (**Haley**, 1996). De ahí que se

proponga la siguiente hipótesis:

H_2 : La imagen y notoriedad del diario tiene una influencia positiva en la actitud hacia el mismo

Actitud

Estudios realizados en diferentes sectores e industrias, prueban la existencia de una relación positiva y significativa entre la actitud del consumidor, su credibilidad percibida y su comportamiento (Goldsmith; Lafferty; Newell, 2000; Lafferty; Goldsmith, 2004). La credibilidad es juicio subjetivo respecto a los atributos de un servicio o producto, y depende de la actitud favorable o desfavorable hacia éstos (Piñeros-Piza et al., 2011). Consecuentemente la credibilidad de los medios de comunicación dependerá de la percepción que el consumidor tiene de una serie de variables –experiencia, confianza...-, así como de su actitud hacia el medio de comunicación en sí mismo (Fogg, 2003). Siguiendo a Meyer (1998), hay una clara correlación entre la credibilidad de los diarios y la actitud de la audiencia. Ello lleva a plantear la siguiente hipótesis:

H_3 : La actitud hacia el diario tiene una influencia positiva en su credibilidad

Fidelidad

Puede definirse como el resultado de un proceso por el cual el consumidor adquiere asociaciones y percepciones positivas sobre un producto o marca, y que finalmente generará un compromiso por parte del consumidor (Dick; Basu, 1994). De esta manera, la fidelidad refleja el grado de satisfacción del consumidor con un producto o servicio (Aaker, 1991). Estudios previos vinculan la actitud del consumidor con su comportamiento (Mueller, 1957; Woo; Cho; Kwon, 2008); y según Rust y Donthu (1988), existe una relación positiva entre la actitud del consumidor hacia un medio de comunicación y su fidelidad hacia el mismo.

La credibilidad del medio de comunicación –su confiabilidad, su honestidad y experiencia- influirá también en el comportamiento del consumidor, y concretamente en la fidelidad y la aceptación de nuevos productos. Estudios previos señalan que los consumidores tienden a escoger aquellos medios que perciben como más creíbles (Rieh; Danielson, 2007). Finalmente, diversos autores han demostrado la relación entre la credibilidad del medio de comunicación y un incremento de consumo del mismo (Van-Weezel, 2006); así como de la fidelidad de la audiencia (Meyer, 2004). Por ello, presentamos las siguientes hipótesis:

H_4 : La actitud hacia el diario tiene una influencia positiva en la fidelidad hacia el mismo

H_6 : La credibilidad del diario tiene una influencia positiva en la fidelidad hacia el mismo

Aceptación de nuevos productos

Una de las consecuencias que tienen en el comportamiento del consumidor la actitud y la credibilidad, es su predisposición hacia nuevos productos del medio. Diversos traba-

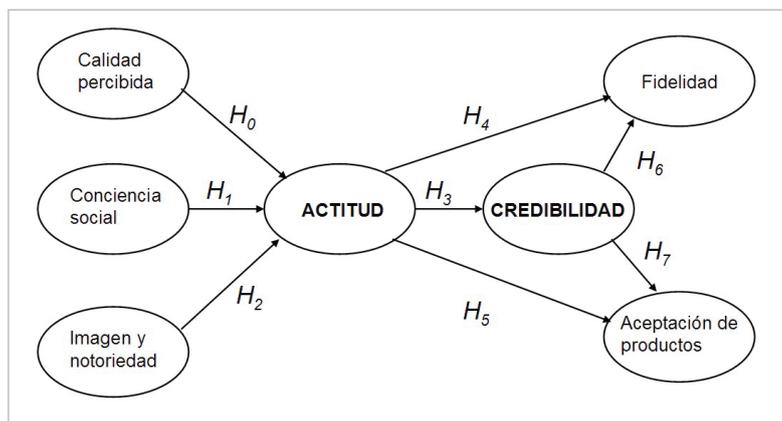


Figura 1. Modelo inicial propuesto

jos han demostrado la relación existente entre la credibilidad del medio y la aceptación hacia los nuevos productos que éste introduce (Bigné-Alcañiz, 1992; Amos; Holmes; Strutton, 2008). Así, se proponen las siguientes hipótesis:

H_5 : La actitud hacia el diario tiene una influencia positiva en la aceptación de nuevos formatos

H_7 : La credibilidad del diario tiene una influencia positiva en la aceptación de nuevos formatos

3. Metodología

3.1. Selección de la muestra

Se ha tenido en cuenta únicamente el canal o medio de información, en cuanto a variables de credibilidad, y se ha seleccionado exclusivamente prensa en soporte papel, prescindiendo de la digital (Pornpitakpan, 2004). Se siguieron dos criterios para la selección de diarios:

- tener amplia notoriedad y difusión; es decir, ser conocidos por la mayoría de los lectores y tener tirada y distribución nacional;
- presentar marcadas líneas editoriales.

Así, se realizó un primer estudio exploratorio para conocer los diarios de referencia en España, y conforme a estos resultados, se seleccionaron *El país*, *El mundo*, *Abc*, y *La razón*. También se decidió incluir a *La voz de Galicia* como cabecera representativa de los diarios de difusión autonómica. El método de muestreo fue aleatorio, enviando a cada participante un cuestionario correspondiente a uno de los diarios seleccionados. Los cuestionarios fueron enviados a personas residentes en el territorio español.

El método de recogida de información fue el cuestionario electrónico auto-administrado que agrupaba las distintas variables objeto de estudio mediante escalas clásicas tipo Likert de 5 puntos (siendo 1 totalmente en desacuerdo y 5 totalmente de acuerdo). Además se incluyeron variables socio demográficas como género, edad, nivel de renta o de estudios. El 44,5% de los participantes eran hombres, frente a un 55,5% de mujeres. El 19,48% tenía estudios primarios, mientras, el 47,09% poseía un título universitario, un 27,71% estudios secundarios y un 5,72% de formación profesional. Por último, y en relación al nivel de renta, un 35,17% ganaba entre 6.000 y 18.000 euros al año, un 36,48% entre 18.000 y 30.000, y un 16,5% declaraba unos ingresos superiores a los

30.000 euros por año. El trabajo de campo se llevó a cabo durante los meses de marzo y abril de 2013 y se obtuvieron 454 cuestionarios válidos: *El país* (123), *El mundo* (88), *Abc* (80), *La razón* (65) y *La voz de Galicia* (98). Por último, el error muestral fue del 4,7%, y el nivel de confianza del 95%.

3.2. Selección de las variables de medida

En relación con las variables empleadas en nuestro estudio, se han seleccionado de acuerdo a la literatura previa sobre el tema y mediante la elección de variables ya validadas en investigaciones anteriores.

- Calidad percibida. Se empleó una escala que trata de recoger el juicio global del consumidor acerca de la información o bien su superioridad frente a otros medios. La escala utilizada se basó en los trabajos de **Medina y Ojer** (2009) y de **Gutierrez-Cea** (2000).
- Conciencia social de los diarios. Se midió mediante los ítems propuestos por **Gaziano y McGrath** (1986).
- Notoriedad e imagen. Se emplearon los ítems propuestos por **Boulding** (1969) y por **Aaker** (1991), referidos al conocimiento y familiaridad que el consumidor tiene con el medio. La variable credibilidad se midió mediante ítems propuestos por **Gaziano y McGrath** (1986), así como por **Newell y Goldsmith** (2001).
- Actitud de los consumidores. Se midió a través de la escala planteada por **Bruner y Hensel** (1992).
- Fidelidad hacia el medio. Se utilizaron ítems propuestos por **Nguyen y Leblanc** (2001).
- Aceptación de nuevos productos. Se midió a partir de la escala propuesta por **Klink y Smith** (2001).

4. Discusión de resultados

4.1. Análisis del modelo de medida

En el presente estudio se ha llevado a cabo un modelo de ecuaciones estructurales, a través del programa *Amos 18.0*, considerando las variables como reflectivas. Antes de analizar el modelo propuesto, examinamos su estructura multidimensional con un análisis factorial confirmatorio, con el propósito de comprobar la fiabilidad y validez de la escala de medida. Un primer análisis reveló la necesidad de eliminar un ítem de la escala inicial propuesta, en concreto el relacionado con la variable imagen/notoriedad (*Aso3*). Una vez suprimido este indicador, los resultados muestran una especificación adecuada de la estructura factorial propuesta (tabla 1). Para el análisis de la fiabilidad y de la consistencia interna del modelo, se calcularon los valores *alpha de Cronbach*, los índices de fiabilidad compuesta y los valores de la varianza extraída. Siguiendo a **Anderson y Gerbing** (1988) y **Hair et al.** (1992), la fiabilidad se considera aceptable cuando los valores *alfa de Cronbach* superan el valor de 0,7, valor superado por todos los constructos en nuestro análisis. En cuanto a la fiabilidad compuesta -medida de consistencia interna de los constructos-, se confirmaría en nuestro modelo, dado que los valores de los coeficientes lambda estandarizados son superiores a 0,5 (**Bagozzi; Yi**, 1988), a excepción del ítem *Cal4*. En relación con la validez discriminante, los valores de la varianza extraída también se consideran aceptables, dado que superan el valor 0,5 (**Hair et al.**, 1999).

Para el ajuste estructural del modelo, se analizaron una serie de medidas. De acuerdo con los resultados, chi-cuadrado obtiene un valor de 871,523 (DF=386, $p < 0,001$), y se logra un índice de bondad del ajuste (GFI) de 0,878, pudiendo considerarse indicadores fiables del ajuste del modelo (**Bollen**, 1989; **Hair et al.** 1999). Asimismo, el índice de ajuste normalizado (NFI), alcanza un valor de 0,927, siendo los valores superiores a 0,9 generalmente aceptados como evidencia de un buen ajuste (**Browne; Cudeck**, 1993). El RMSEA (*root mean square error of approximation*) alcanza un valor de 0,053, evidenciando un ajuste aceptable del modelo general (**Bagozzi; Yi**, 1988).

Tabla 1. Cargas factoriales de las variables latentes, indicadores de fiabilidad y consistencia interna

Constructos	Indicadores	Lambda (Carga factorial)	Cronbach Alfa	CR (Índice fiabilidad compuesta)	AVE (Índice varianza extraída)
Calidad percibida	<i>Cal1</i>	0,857	0,725	0,740	0,546
	<i>Cal2</i>	0,728			
	<i>Cal3</i>	0,611			
	<i>Cal4</i>	0,335			
Conciencia social	<i>Cs1</i>	0,542	0,749	0,767	0,557
	<i>Cs2</i>	0,586			
	<i>Cs3</i>	0,759			
	<i>Cs4</i>	0,786			
Notoriedad /imagen	<i>Aso1</i>	0,882	0,818	0,823	0,700
	<i>Aso2</i>	0,789			
Credibilidad	<i>Cr1</i>	0,894	0,938	0,937	0,652
	<i>Cr2</i>	0,878			
	<i>Cr3</i>	0,887			
	<i>Cr4</i>	0,896			
	<i>Cr5</i>	0,681			
	<i>Cr6</i>	0,708			
	<i>Cr7</i>	0,665			
	<i>Cr8</i>	0,808			
Actitud	<i>Ac1</i>	0,886	0,949	0,940	0,797
	<i>Ac2</i>	0,911			
	<i>Ac3</i>	0,887			
	<i>Ac4</i>	0,887			
Fidelidad	<i>Fid1</i>	0,712	0,888	0,870	0,632
	<i>Fid2</i>	0,614			
	<i>Fid3</i>	0,924			
	<i>Fid4</i>	0,890			
Aceptación nuevos productos	<i>Ap1</i>	0,807	0,863	0,868	0,767
	<i>Ap2</i>	0,939			

Las medidas del ajuste incremental, así como las medidas de parsimonia también indican un buen ajuste, dado que las primeras alcanzan valores superiores a 0,9 o bien muy próximas, como es el caso del *índice de bondad del ajuste normalizado* o AGFI (Hair et al., 1999). Por último, las medidas de parsimonia también indican un ajuste apropiado del modelo, dado que el *índice de ajuste comparativo* (CFI), el *índice de ajuste normalizado* (IFI) y el *índice de Tucker-Lewis* (TLI) alcanzan valores superiores al 0,9. (Hair et al. 1999).

4.2. Análisis del modelo propuesto

Nuestros resultados ponen de relieve que todas las relaciones causales planteadas son positivas y estadísticamente significativas, a excepción de la influencia de la variable conciencia social sobre la actitud, dado que no se halla suficiente evidencia empírica para proponer una relación significativa entre ambas (figura 2). El análisis de los coeficientes estandarizados muestra que los antecedentes de la credibilidad del medio de comunicación, esto es, la calidad percibida ($\beta_{14}=0,491$), y la imagen y notoriedad ($\beta_{34}=0,661$), tienen una influencia positiva y significativa sobre la actitud, mostrando el mayor peso. Es decir, la variable imagen y notoriedad del medio tiene un efecto dominante en la actitud del consumidor. Los resultados sugieren que la notoriedad del medio, esto es, la familiaridad de éste para el consumidor y su capacidad para reconocerlo y recordarlo, así como una imagen positiva y favorable ejercen la mayor influencia en la actitud del consumidor de información.

La variable imagen y notoriedad del medio tiene un efecto dominante en la actitud del consumidor

Sin embargo, las relaciones causales descritas en este modelo no muestran evidencia significativa de la influencia de la variable conciencia social sobre la actitud del consumidor. Esto puede deberse a que en el entorno actual, caracterizado por una crisis institucional y económica, el conocimiento de numerosos casos de corrupción se debe a la labor de los medios de comunicación –entre ellos los diarios-, y por tanto el consumidor presupone y da por supuesta dicha conciencia social de la prensa. Y es que según autores como **Gaziano y McGrath** (1986), la conciencia social incluye factores como el comportamiento ético y moral en el tratamiento de la información y el compromiso con el interés público, valores que se ponen de relieve en el contexto actual.

Por otro lado el análisis de la variable actitud revela su influencia positiva y significativa sobre la credibilidad ($\beta_{45}=0,865$), la fidelidad hacia el medio ($\beta_{46}=0,941$) y la aceptación de nuevos productos ($\beta_{47}=0,889$). De hecho, la relación entre la actitud y la fidelidad hacia el

Tabla 2. Índices de ajuste del modelo estructural

Medidas absolutas de ajuste	
Chi-cuadrado	871,523
Grados de libertad	386
Nivel de significación	0,000
Índice de bondad del ajuste (GFI)	0,878
Root mean square error of approx (RMSEA)	0,053
Medidas de ajuste incremental	
Índice de bondad del ajuste normalizado (AGFI)	0,853
Índice de ajuste normalizado (NFI)	0,927
Índice del ajuste incremental (IFI)	0,958
Índice de Tucker-Lewis (TLI)	0,953
Índice de ajuste comparativo (CFI)	0,958
Medidas de parsimonia	
Chi cuadrado normalizada	2,258

medio es la más fuerte identificada en el modelo, poniendo de relieve la fuerte influencia que una mejora o cambio en la actitud puede ocasionar. Asimismo, la credibilidad también muestra una influencia positiva sobre la fidelidad hacia el medio ($\beta_{56}=0,527$) y la aceptación de nuevos productos ($\beta_{57}=0,464$), pero con un peso menor. Los resultados obtenidos están en la línea de investigaciones previas (Vila-López; Küster; Bigné, 2013) que señalan que la fidelidad hacia los diarios y la aceptación de productos nuevos se explica tanto por la actitud del consumidor, como por la credibilidad del medio.

El análisis del modelo propuesto sugiere que todas las hipótesis planteadas inicialmente se aceptan $-H_0, H_2, H_3, H_4, H_5, H_6, H_7$, a excepción de H_1 . Los resultados muestran que si un diario mejora su calidad percibida o su notoriedad e imagen, ello repercutirá en una mejor actitud de los consumidores (H_0 y H_2). Además, una mejora en la actitud hacia el diario conllevará una mayor fidelidad hacia el mismo (H_4), y una mejor predisposición hacia los nuevos formatos y productos que pueda introducir en el mercado (H_5); así como una mayor credibilidad (H_3). Finalmente, la credibilidad del medio también influye sobre ambas consecuencias, aceptándose igualmente H_6 y H_7 . Sin embargo, la hipótesis inicial H_1 : *La conciencia social del diario tiene una influencia positiva en la actitud hacia el mismo*, no puede ser aceptada, dado que no se halla suficiente evidencia empírica para proponer una relación significativa entre ambas variables.

Tabla 3. Relaciones causales entre variables y contraste de hipótesis (coeficientes estandarizados)

Relaciones causales	Coefficientes estandarizados	Valor t	Hipótesis
Calidad percibida \rightarrow Actitud	$\beta_{14}=0,491^*$	2,125	$H_0=$ Se acepta
Conciencia social \rightarrow Actitud	$\beta_{24}=0,084^{ns}$	0,508	$H_1=$ No se acepta
Imagen/notoriedad \rightarrow Actitud	$\beta_{34}=0,661^*$	3,594	$H_2=$ Se acepta
Actitud \rightarrow Credibilidad	$\beta_{45}=0,865^*$	6,741	$H_3=$ Se acepta
Actitud \rightarrow Fidelidad	$\beta_{46}=0,941^*$	6,856	$H_4=$ Se acepta
Actitud \rightarrow Aceptación nuevos productos	$\beta_{47}=0,889^*$	6,820	$H_5=$ Se acepta
Credibilidad \rightarrow Fidelidad	$\beta_{56}=0,527^*$	2,719	$H_6=$ Se acepta
Credibilidad \rightarrow Aceptación nuevos productos	$\beta_{57}=0,464^*$	1,967	$H_7=$ Se acepta
$\chi^2=871,523$ df=386 p=,000			
R^2 (Actitud)= 0,883 R^2 (Credibilidad)= 0,932			
ns=no significativo			
* significativo (p<0,05)			

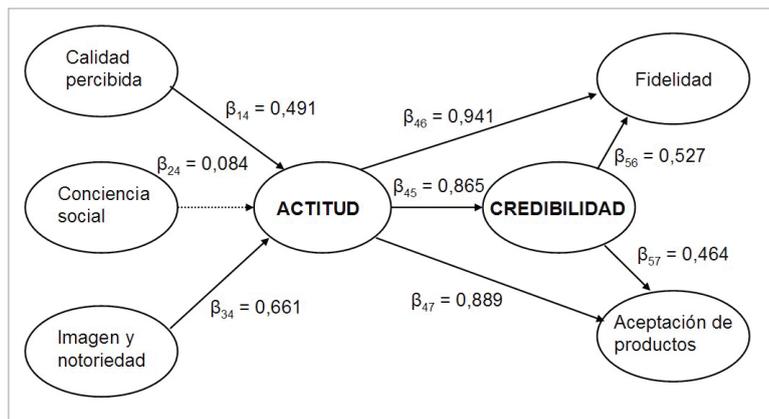


Figura 2. Relaciones causales

5. Conclusiones

En primer lugar, los resultados obtenidos señalan el carácter multidimensional del concepto de credibilidad de los medios de comunicación (Gaziano; McGrath, 1986), que se prueba empíricamente para el caso de los diarios generalistas y de referencia.

En segundo lugar, se confirma la existencia de un orden causal en el proceso de formación de la credibilidad del medio de comunicación, y el efecto positivo de dicha credibilidad en el comportamiento del consumidor, más concretamente en su fidelidad hacia el medio y su predisposición a aceptar nuevos formatos y productos.

Si un diario mejora su calidad percibida o su notoriedad e imagen, ello repercutirá en una mejor actitud de los consumidores

En cuanto a los antecedentes de la credibilidad, se confirma la existencia de una relación positiva y significativa entre la calidad percibida, la notoriedad e imagen y la actitud del consumidor de información. De manera que cuanto mayor sea la notoriedad del diario, su calidad percibida y más favorable sea su imagen de la marca, más favorable será la actitud de los consumidores y mejor será su valoración. Por tanto, para la construcción de un medio creíble, se debe tener en cuenta la calidad que el consumidor infiere a partir de su experiencia (Gotsi; Wilson, 2001). Asimismo, la creación de un medio con una imagen positiva, favorable y familiar entre los consumidores es uno de los primeros pasos que las empresas de la industria de la comunicación deben tener en cuenta para la creación de un medio creíble, dado que una imagen e identidad sólida contribuyen notablemente a incrementar la credibilidad. Esto es, una imagen atractiva y sólida genera la percepción de que el medio es creíble.

De la misma manera, se muestra la existencia de una relación positiva entre la actitud y la credibilidad del medio, así como de ambas variables sobre la fidelidad y la aceptación hacia nuevos productos por parte de los consumidores. Podemos concluir que la creación y el incremento de la fidelidad de los consumidores es especialmente importante para

los medios de comunicación en un entorno de gran competencia y fragmentación del sector; y por lo tanto, los gestores de los medios han de buscar y procurar una audiencia o un consumidor fiel, siendo crucial la credibilidad percibida del medio. Se puede afirmar que la creación de medios de comunicación creíbles se convierte en una prioridad para las empresas.

Nuestra investigación añade valor a la literatura existente, mediante una comprensión más profunda del proceso de creación de credibilidad de los medios de comunicación, para el caso de los diarios generalistas y de referencia, mostrando las relaciones entre las variables

anteriores, así como en las consecuencias que tiene en la respuesta por parte del consumidor de información.

A la luz de estos resultados cabe señalar una serie de implicaciones para los responsables de los medios de comunicación.

- Sugerimos una apuesta clara por incrementar y mejorar la credibilidad del medio, especialmente en un entorno tan competitivo, lo que hace necesario que los directivos de los medios analicen y conozcan la credibilidad que transmiten (Balmer; Greyser, 2003), entendiendo la credibilidad como un concepto multidimensional que recoge atributos como honestidad, habilidad, experiencia y confianza asociados al medio.
- Para conseguir la credibilidad del medio es fundamental mejorar la calidad percibida del mismo, así como reforzar e incrementar su notoriedad e imagen. La mejora de la calidad percibida debería pasar por ofrecer un contenido tanto de información como de opinión, dirigido a una audiencia heterogénea, con pluralidad de creencias y opiniones, proporcionando información de calidad, sin sesgos, contando la historia completa y contribuyendo a la formación de opinión pública (Ortega, 2008; Bigné-Alcañiz; Currás-Pérez; Sánchez-García, 2009).
- Los gestores deben mejorar la notoriedad e imagen de los diarios, comenzando con su identidad visual, actualizando su logotipo, tipografía, nombres y símbolos (Van-Weezel, 2006). Para ello las empresas periodísticas han de destinar recursos para el incremento y mejora de su notoriedad en el mercado, su familiaridad y capacidad de reconocimiento y recuerdo por parte de los consumidores; así como para crear una cabecera visible y diferenciada de otros diarios, mediante el empleo de una marca y un logotipo claramente identificables. En definitiva los diarios deben proveer valor a los consumidores de información, invitándoles a una experiencia placentera a través de sus contenidos (Erdem; Swait, 1998).

En cuanto a las principales limitaciones del estudio, se debe señalar lo siguiente:

- la investigación empírica se ha circunscrito a España, así como a un tipo específico de medio de comunicación –la prensa escrita generalista y de referencia–, lo que plantea serias dificultades a la hora de extrapolar los resultados al conjunto de medios de comunicación presentes en el

mercado, ya sean tradicionales o en soporte digital;

- nuestro trabajo ha analizado una serie de variables propuestas anteriormente por la literatura, obviando otras importantes en el estudio de la credibilidad o actitud hacia los medios, como la familiaridad con el medio (Vila-López; Küster; Bigné, 2013) o la reputación o prestigio (Newell; Goldsmith, 2001; Herrera-Damas, 2013). Por estos motivos, las relaciones que pone de manifiesto nuestra investigación deben ser tomadas con cautela.

Los gestores de los medios han de buscar y procurar una audiencia o un consumidor fiel, siendo crucial la credibilidad percibida del medio

6. Bibliografía

- Aaker, David A.** (1991). *Managing brand equity capitalizing on the value of brand name*. New York: The Free Press. ISBN: 0 02 900101 3
- Amos, Clinton; Holmes, Gary; Strutton, David** (2008). "Exploring the relationship between celebrity endorser effects and advertising effectiveness: a quantitative synthesis of effect size". *International journal of advertising*, v. 27, n. 2, pp. 209-234, ISSN: 0265 0487
<http://hull.aug.edu/hcbWeb/media/research/Amos-Holmes-Strutton-IJA-2008.pdf>
- Anderson, James C.; Gerbing, David W.** (1988). "Structural equation modeling in practice: A review and recommended two step approach". *Psychological bulletin*, v. 103, pp. 411-423.
<http://www.fdeuw.unimaas.nl/meteor/EDEN/Mike%20Brady/Anderson%20and%20Gerbing%201988.pdf>
- Bagozzi, Richard P.; Yi, Youjue** (1988). "On the evaluation of structural equation models". *Journal of the Academy of Marketing Science*, v. 16, n. 1, pp. 74-94.
<http://dx.doi.org/10.1007%2FBF02723327>
- Balmer, John; Greyser, Stephen** (2003). *Revealing the corporation. Perspectives on identity image, reputation, corporate branding and corporate-level marketing*. Routledge: London. ISBN: 0 415 28420 1
- Bollen, Kenneth** (1989). *Structural equations with latent variables. Wiley series in probability and mathematical statistic*. John Wiley and Sons, New York. ISBN: 0 471 01171 1
- Bigné-Alcañiz, José-Enrique** (1992). "El efecto persuasivo de la fuente: su extensión objetiva". *Alta dirección*, v. 27, n. 162, marzo-abril, pp. 75-83.
- Bigné-Alcañiz, José-Enrique; Currás-Pérez, Rafael; Sánchez-García, Isabel** (2009). "Brand credibility in cause-related marketing: the moderating role of consumer values". *Journal of product and brand management*, v. 18, n. 6, pp. 437-447.
<http://dx.doi.org/10.1108/10610420910989758>
- Blackwell, Roger; Engel, James; Miniard, Paul** (2002). *Comportamiento del consumidor*. México: Thompson. ISBN: 970 686 187 4
- Boulding, Kenneth E.** (1969). *The image: knowledge in life and society*. 7th edition. Michigan: Ann Arbor, The University of Michigan Press. ISBN: 0 472 06047 3
- Browne, Michael W.; Cudeck, Robert** (1993). *Alternative ways of assessing model fit*. In: Bollen, Kenneth A.; Long, J. Scott (eds.). *Testing structural equation models*. Sage Publications. Thousand Oaks, C. A. pp. 136-161. ISBN: 0 8039 4506 X
- Bruner, Gordon; Hensel, Paul** (1992). *Marketing scales handbook: a compilation of multi item measures*. Chicago: American Marketing Association. ISBN: 0 877 57226 7
- Carrillo, María-Victoria; Castillo-Díaz, Ana; Tato-Jiménez, Juan-Luis** (2008). "El valor de lo intangible: La gestión de la reputación corporativa. El caso de la marca Telefónica". *Observatorio journal*, v. 7, pp. 239-254.
<http://obs.obercom.pt/index.php/obs/article/download/123/215>
- Chan-Olmsted, Sylvia M.; Yungwook, Kim** (2001). "Perceptions of branding among television station managers: an exploratory study". *Journal of broadcasting & electronic media*, v. 45, n. 1, pp. 75-91.
http://dx.doi.org/10.1207%2Fs15506878jobem4501_6
- Dick, Alan; Basu, Kunal** (1994). "Customer loyalty: towards an integrated framework". *Journal of the academy of marketing science*, v. 22, n. 2, pp. 99-113.
<http://dx.doi.org/10.1177/0092070394222001>
- Erdem, Tülin; Swait, Joffre** (1998). "Brand equity as a signalling phenomenon". *Journal of consumer psychology*, v. 7, n. 2, pp. 131-157.
http://markenmanagement.files.wordpress.com/2012/01/erdem_brandequityasasignalinphenomenon.pdf
http://dx.doi.org/10.1207%2Fs15327663jcp0702_02
- Erdem, Tülin; Swait, Joffre** (2004). "Brand credibility, brand consideration and choice". *Journal of consumer research*, v. 31, n. 1, pp. 191-198.
<http://dx.doi.org/10.1086%2F383434>
- Flanagin, Andrew J.; Metzger, Miriam J.** (2000). "Perceptions of internet information credibility". *Journalism and mass communication quarterly*, v. 77, n.3, pp. 515-540.
<http://www.jasonmorrison.net/iakm/4006074.pdf>
<http://dx.doi.org/10.1177%2F107769900007700304>
- Fogg, Brian J.** (2003). "Prominence-interpretation theory: explaining how people assess credibility online". En: *Conference on Human Factors in Computing Systems*, pp. 722-723. Ft. Lauderdale: Florida (USA), ISBN: 1 58 113637 4
<http://credibility.stanford.edu/pdf/PITheory.pdf>
<http://dx.doi.org/10.1145%2F765948.765951>
- Fombrun, Charles J.; Van-Riel, Cees** (2004). *Fame and fortune: how successful companies build winning reputations*. Financial Times Prentice-Hall: Upper Saddle River, NJ. ISBN: 0 13 093737 1
- Gaziano, Cecile; McGrath, Kristin** (1986). "Measuring the concept of credibility". *Journalism quarterly*, autumn, pp.

451-462.

<http://www.aejmc.org/home/wp-content/uploads/2012/09/Journalism-Quarterly-1986-Gaziano-451-621.pdf>
<http://dx.doi.org/10.1177%2F107769908606300301>

Goldsmith, Ronald; Lafferty, Barbara; Newell, Stephen (2000). "The influence of corporate credibility on consumer attitudes and purchase intent". *Corporate reputation review*, v. 3, n. 4, pp. 304-318.

<http://dx.doi.org/10.1057%2Fpalgrave.crr.1540122>

Gotsi, Manto; Wilson, Alan (2001). "Corporate reputation: seeking a definition". *Corporate communications: an international journal*, v. 6, n. 1, pp. 24-30.

<http://dx.doi.org/10.1108/13563280110381189>

Greer, Jennifer (2003). "Evaluating the credibility of online information: A test of source and advertising influence". *Mass communication and society*, v. 6, n. 11, pp. 11-28.

http://dx.doi.org/10.1207%2FS15327825MCS0601_3

Gutiérrez-Cea, Charo (2000). "Televisión y calidad: perspectivas de investigación y criterios de evaluación". *Revista de estudios de comunicación*, v. 9, pp. 1-20.

Haley, Eric (1996). "Exploring the construct of organization as source: consumers' understandings of organizational sponsorship of advocacy advertising". *Journal of advertising*, v. 25, n. 2, pp. 19-3.

<http://dx.doi.org/10.1080/00913367.1996.10673497>

Hair, Joseph F.; Anderson, Rolph E.; Tatham, Ronald L.; Black, William (1992). *Multivariable data analysis*. 5th ed. Prentice Hall, Upper Saddle River, New Jersey. ISBN: 0 13 894858 5

Hair, Joseph F.; Anderson, Rolph E.; Tatham, Ronald L.; Black, William (1999). *Análisis Multivariante*. Prentice Hall, Madrid, ISBN: 84 8322 035 0

Herrera-Damas, Susana (2013). "Indicaciones recurrentes en las normativas para el uso periodístico de las redes sociales". *El profesional de la información*, enero-febrero, v. 22, n. 1, pp. 46-53.

<http://dx.doi.org/10.3145/epi.2013.ene.06>

Hovland, Carl; Janis, Irving; Kelley, Harold (1953). *Communication and persuasion. Psychological studies of opinion change*. Yale University Press, New Haven. ISBN: 0 30 000573 3

Klink, Richard R.; Smith, Daniel C. (2001). "Threats to the external validity of brand extension research". *Journal of marketing research*, v. 38, pp. 326-35.

<http://dx.doi.org/10.1509%2Fjmkr.38.3.326.18864>

Lafferty, Barbara; Goldsmith, Ronald (2004). "How influential are corporate credibility and endorser attractiveness when innovations react to advertisements for a new high-technology product". *Corporate reputation review*, n. 7, n. 1, pp. 24-36.

<http://dx.doi.org/10.1057%2Fpalgrave.crr.1540209>

Liu, Ziming (2004). "Perceptions of credibility of scholarly information on the web". *Information processing and management*, v. 40, n. 6, pp. 1027-1038.

[http://dx.doi.org/10.1016/S0306-4573\(03\)00064-5](http://dx.doi.org/10.1016/S0306-4573(03)00064-5)

Medina, Mercedes; Ojer, Teresa (2009). "Valoración del servicio público de televisión: comparación entre la BBC y TVE". *Revista latina de comunicación social*, v. 64, pp. 275-299.

<http://dx.doi.org/10.4185/RLCS-64-2009-823-275-299>

Metzger, Miriam J.; Flanagin, Andrew J.; Eyal, Karen; Lemus, Daisy R.; McCann, Robert M. (2003). "Credibility for the 21st century: integrating perspectives on source, message and media credibility in the contemporary media environment". *Communication yearbook*, v. 27, pp. 293-336.

http://dx.doi.org/10.1207/s15567419cy2701_10

Meyer, Philip (1998). "Imagining public journalism". En: Lambeth, Edmund; Meyer, Philip; Thorson, Esther (Eds.). *Assessing public journalism*. Columbia: University of Missouri Press, pp. 251-273. ISBN: 0 82 621158 5

Meyer, Philip (2004). *The vanishing newspaper: saving journalism in the information age*. Columbia: University of Missouri Press. ISBN: 0 82 621877 6

Mueller, Eva (1957). "Effects of consumer attitudes on purchases". *The American economic review*, pp. 946-965.

Newell, Stephen J.; Goldsmith, Ronald E. (2001). "The development of a scale to measure perceived corporate credibility". *Journal of business research*, v. 52, n. 3, pp. 235-247.

<http://dx.doi.org/10.1016%2FS0148-2963%2899%2900104-6>

Nguyen, Nha; Leblanc, Gaston (2001). "Corporate image and corporate reputation in customers retention decisions in services". *Journal of retailing and consumer services*, v. 8, n.4, pp. 227-236.

<http://dx.doi.org/10.1016%2FS0969-6989%2800%2900029-1>

Nozato, Yoshiko (2002). "Credibility of online newspapers". *Annual Convention of Association for Education in Journalism and Mass Communication*. Washington D.C., August.

Ortega, Félix (2008). "La encrucijada de la televisión pública en Chile". *Cuadernos de información*, v. 23, pp. 42-61.

<http://cuadernos.uc.cl/uc/index.php/CDI/article/view/95/87>

Piñeros-Piza, Catalina; Gutiérrez-Coba, Liliana-María; Gómez-Díaz, Javier-Andrés; Pérez-Acosta, Andrés M.; Salgado-Cardona, Andrea; Mora, María-Lucía; Grijalva, Nataly; Estrada, Juliana; Ramírez, María-José (2011). "Credibilidad percibida del periodismo a través de la Internet: una visión desde la psicología del consumidor". *Perspectivas psicológicas*, v. 7, n. 2, pp. 225-238.

http://www.usta.edu.co/otraspaginas/diversitas/doc_pdf/diversitas_11/vol.7no.2/02_credibilidad_percibida.pdf

Pornpitakpan, Chanthika (2004). "The persuasiveness of source credibility: a critical review of five decades's evidence". *Journal of applied social psychology*, v. 34, n. 2, pp. 243-281.

<http://dx.doi.org/10.1111/j.1559-1816.2004.tb02547.x>

Rathnayake, Chamil-Viranga (2008). "Brand personality and its impact on brand feelings: a study among young television viewers". *South Asian journal of management*, v. 15, n. 2, pp. 7-26, ISSN: 0971 5428

Rieh, Soo-Young; Danielson, David (2007). "Credibility: a multidisciplinary framework". *Annual review of information*

- science and technology*, v. 41, n. 1, pp. 307-364.
http://rieh.people.si.umich.edu/~rieh/papers/rieh_ARIST2007.pdf
<http://dx.doi.org/10.1002%2Faris.2007.1440410114>
- Rust, Roland T.; Donthu, Naveen** (1988). "A programming and positioning strategy for cable television". *Journal of advertising*, v. 17, n. 4, pp. 6-13.
<http://dx.doi.org/10.1080%2F00913367.1988.10673124>
- Schweiger, Wolfgang** (2000). "Media credibility: experience or image?". *European journal of communication*, v. 15, n. 37, pp. 37-59.
<http://dx.doi.org/10.1177/0267323100015001002>
- Sughan, Steven** (1980). "The cost of thinking". *Journal of consumer research*, v. 7, pp. 99-111.
- Taylor, Robert-Saxton** (1986). *Value-added processes in information systems*. Norwood, NJ: Ablex. ISBN: 978 0893912734
- Trimble, Carrie; Rifon, Nora** (2006). "Consumer perceptions of compatibility in cause-related marketing messages". *International journal of nonprofit and voluntary sector marketing*, v. 11, n. 1, pp. 29-47.
<http://dx.doi.org/10.1002%2Fnmvsm.42>
- Tseng, Shawn; Fogg, B. J.** (1999). "Credibility and computing technology". *Communications of the ACM*, v. 42, n. 5, pp. 39-44.
<http://www.mehmetgokturk.com/files/hcicourse/p39-tseng.pdf>
<http://dx.doi.org/10.1145%2F301353.301402>
- Van-Weezel, Aldo** (2006). "Contenidos de calidad: una apuesta lógica". *Cuadernos de gestión*, v. 19, pp. 62-67.
<http://www.redalyc.org/articulo.oa?id=97117399009>
- Vila-López, Natalia; Küster-Boluda, Inés; Bigné, Enrique** (2013). "Credibilidad corporativa en cadenas televisivas: antecedentes y efectos". *Revista española de investigación de marketing ESIC*, v. 17, n. 1, pp. 61-82.
http://www.esic.edu/documentos/revistas/reim/130320_161859_E.pdf
- Walsh, Gianfranco; Mitchell, Vincent; Jackson, Paul; Beatty, Sharon** (2009). "Examining the antecedents and consequences of corporate reputation: a customer perspective". *British journal of management*, v. 20, n. 2, pp. 187-203.
<http://dx.doi.org/10.1111%2Fj.1467-8551.2007.00557.x>
- Wang, Peiling; Soergel, Dagobert** (1998). "A cognitive model of document use during a research project Study 1: Document selection". *Journal of the American Society for Information Science*, v. 49, n. 2, pp. 115-133.
http://www.asis.org/Publications/JASIS/Best_Jasist/1999WangandSoergel.pdf
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199802\)49:2<115::AID-ASIS3>3.0.CO;2-T](http://dx.doi.org/10.1002/(SICI)1097-4571(199802)49:2<115::AID-ASIS3>3.0.CO;2-T)
- Wathen, Nadine; Burkell, Jacquelyn** (2002). "Believe it or not, factors influencing credibility on the web". *Journal of the American Society for Information Science and Technology*, v. 53, n. 2, pp. 134-144.
<http://dx.doi.org/10.1002/asi.10016>
- Webster, James G.** (1986) "Audience behavior in the new media environment". *Journal of communication*, v. 36, n. 6, pp. 77-94.
<http://dx.doi.org/10.1111/j.1460-2466.1986.tb01439.x>
- Woo, Jong; Cho, Chang-Hoan; Kwon, Joon**, (2008). "The role of affect and cognition in consumer evaluations of corporate visual identity: Perspectives from the United States and Korea". *Journal of brand management*, v. 15, n.6, pp. 382-398
<http://dx.doi.org/10.1057/bm.2008.11>

Próximos temas centrales

Julio 2014	Altmétricas
Septiembre 2014	Humanidades digitales
Noviembre 2014	Big data y analítica web
Enero 2015	Marketing de productos de información
Marzo 2015	Web semántica
Mayo 2015	Privacidad y seguridad de información
Julio 2015	Bases de datos bibliográficas

Los interesados por favor consulten detenidamente las Normas para autores:

<http://www.elprofesionaldeinformacion.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>

Suscríbete a

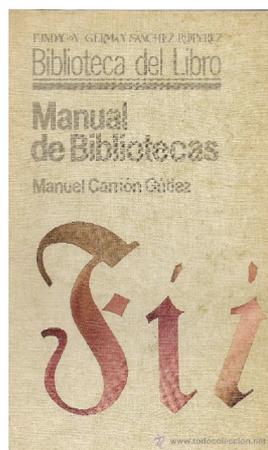
Mi Biblioteca

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

VISUALIZACIÓN DE LA INFORMACIÓN EN LA DEMOCRATIZACIÓN DE LOS DATOS: PROPUESTAS DESDE EL PERIODISMO Y LA NARRATIVIDAD

Clàudia Sánchez-Bonvehí y Mireia Ribera



Clàudia Sánchez-Bonvehí es estudiante de doctorado en el programa *Información y documentación en la sociedad del conocimiento* de la *Universitat de Barcelona (UB)*, licenciada en comunicación audiovisual por la *Universitat Pompeu Fabra (UPF)* y máster en gestión de contenidos digitales por la *UB* y *UPF*. Ha trabajado como periodista en el periódico *Regió 7*, ha colaborado con el departamento de comunicación y cultura del *Instituto Cervantes* de Bruselas y como coordinadora del programa *Col·laboratori interuniversitari de recursos d'aprenentatge en xarxa (Cirax)* en el *CBUC*.
<http://orcid.org/0000-0002-5053-7901>

claudiasbonvehi@gmail.com



Mireia Ribera es doctora en documentación e ingeniera informática. Es profesora en el *Departamento de Biblioteconomía y Documentación* de la *Universitat de Barcelona (UB)*. Trabaja en el área de la accesibilidad y usabilidad en lectura digital, y ha impartido la asignatura de visualización de la información en el máster en gestión de contenidos digitales en la *UB*.
<http://orcid.org/0000-0003-1455-1869>

ribera@ub.edu

Universitat de Barcelona

Melcior de Palau, 140. 08014 Barcelona, España

Resumen

Se revisa la visualización de la información desde una perspectiva comunicativa, considerándola como una nueva fórmula para la divulgación de datos complejos a un público heterogéneo y no versado en el análisis de datos cuantitativos. Se propone la incorporación de las técnicas de *storytelling* como herramienta para mantener la atención y conducir la lectura de las visualizaciones, y complementar las dos bases tradicionales de la disciplina: percepción e interacción. La reflexión se acompaña de visualizaciones publicadas en el marco de la comunicación de masas, desde propuestas presentadas por la prensa generalista anglosajona a aplicaciones en línea con el objetivo de aproximar los datos a los ciudadanos.

Palabras clave

Comunicación, Diseño centrado en el usuario, Visualización de la información, Narración, Democratización de los datos, Visualización para las masas, Infovis.

Title: Information visualization in data democratization: proposals from journalism and storytelling

Abstract

Information visualization is reviewed from the communication perspective, which is considered a new formula for disseminating information to a heterogeneous audience of people not specialized in the analysis of quantitative data. The principle of storytelling is incorporated as a tool to hold the attention and direct the reading of the visualizations, and to complement the two traditional bases of the discipline: perception and interaction. The reflection is accompanied by visualizations published in the framework of mass communication, ranging from examples in the English-speaking mainstream press to online applications aimed at making information accessible to the general public.

Keywords

Communication, User-centred design, Information visualization, Storytelling, Data democratization, Visualization for the masses, Infovis.

Sánchez-Bonvehí, Clàudia; Ribera, Mireia (2014). "Visualización de la información en la democratización de los datos: propuestas desde el periodismo y la narratividad". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 311-318.

<http://dx.doi.org/10.3145/epi.2014.may.11>

Artículo recibido el 27-01-2014

Aceptación definitiva: 12-05-2014

1. Nuevos retos en el acto comunicativo

La sociedad de la información se caracteriza por la saturación de información y el cambio de roles en el acto comunicativo, que ha dejado hace años de ser unidireccional y estático para estar regido por una diversidad de direcciones y múltiples dinámicas. Asimismo, como consecuencia de los avances y soluciones tecnológicas, la velocidad, el volumen e incluso la diversidad de datos que se generan cada día, crece de forma exponencial y surgen nuevos retos para gestionarlos y dotarlos de sentido, es decir, para convertirlos en información.

Mientras que el ser humano ha desarrollado nuevas habilidades para adaptarse a este nuevo entorno, su atención sigue siendo una capacidad limitada. La nueva situación se caracteriza por una inundación de datos, donde nuestra habilidad cognitiva de dotarlos de sentido y conducir decisiones fiables e informadas, se mantiene relativamente constante (Heer; Kandel, 2012). Como consecuencia de este cuello de botella, el acto comunicativo puede no ocurrir, a pesar de que en él se fundamenta la transmisión de conocimiento.

Para asegurar el equilibrio de lo que metafóricamente Hal Varian (Finn, 2011) llama la “obesidad informativa”, surge el reto de promover una “dieta informativa equilibrada”. Esto exige la calidad de los datos y su explotación de forma entendible y digerible; es decir, que los datos puedan hablar de aquello que representan sin perderse por el camino de la representación. Hablamos de generar significado.

El color, la textura, el movimiento son propiedades preconscientes que llaman la atención de nuestra percepción incluso antes de ser procesadas por nuestro cerebro

Como respuesta a esta necesidad, la visualización de la información emerge para cambiar el proceso comunicativo y devolver estratégicamente el protagonismo ancestral a lo visual. No se trata de vulgarizar los datos, sino de profundizar y aprovechar las ventajas que ofrece la percepción humana y la interacción, con el objetivo de transformar realidades complejas y abstractas en realidades simples.

En nuestra propuesta vamos más allá de los modelos de visualización de información habitualmente establecidos, incorporando los beneficios del *storytelling* (narratividad) y considerando como público objetivo de las visualizaciones al ciudadano no experto en el análisis de datos ni versado en conceptos estadísticos. La narratividad se plantea como una solución para la creación de visualizaciones que no se dirigen a la comunidad de expertos —en el sentido explicado anteriormente, expertos en análisis de datos y en el manejo de conceptos estadísticos—, y que por el contrario persiguen la divulgación de información, de las historias que cuentan los datos a un público masivo y diverso con motivaciones, intereses y necesidades acentuadamente dispares.

Aproximar los datos a los ciudadanos

En los últimos años las agencias gubernamentales han intentado presentar los datos con la intención de conseguir el máximo impacto social, pero han fracasado en el intento de aproximar indicadores abstractos (Cukier, 2009), y la mayoría de innovaciones sobre visualización han quedado limitadas al ámbito de la investigación. Como consecuencia, diseñador y usuario final han coincidido en la misma figura: el innovador (Fuhrmann *et al.*, 2005). Se hace evidente pues la necesidad de adaptarse a las características del usuario que no es experto¹, y por tanto, no presuponer unos conocimientos del fenómeno representado.

La presencia de la visualización de la información (en adelante, infovis) fuera del ámbito académico es cada vez más acentuada. El responsable del departamento de infovis de la OCDE justifica la importancia de la misma como la forma más obvia para transmitir el conocimiento de los expertos a la gente (Cukier, 2009). Diferentes universidades españolas incorporan asignaturas específicas de visualización y han creado nuevos programas de formación en torno a ella². El campo del periodismo ha sido el primero en utilizar técnicas de visualización para narrar las noticias, como una de las técnicas de lo que se ha dado en denominar “periodismo de datos”. Con ello se busca llegar a la audiencia más allá del impacto, basándose en la narrativa y el intento de hacer que aquello que era complejo a priori, sea comprensible. Con ello el periodismo ha conseguido convertir la información —previamente basada en datos— en historias.

En este artículo se muestra la importancia de esta aportación, que según las autoras debería convertirse en el tercer pilar básico de la infovis, y se buscan los referentes teóricos que permitan llevarla a su máxima expresión. Finalmente se muestra la necesidad de futuras investigaciones en el área con las técnicas de diseño centrado en el usuario (DCU) para consolidar las propuestas teóricas y los primeros intentos de fusión entre infovis y narratividad.

2. Dimensiones de la visualización de la información

La bibliografía tradicional sobre infovis desde sus inicios coincide en considerar dos elementos principales: la teoría de la percepción del ser humano y la explotación de la interacción (Shneiderman, 1996; Card; Mackinlay; Shneiderman, 1999; Spence, 2001; 2007; Segel; Heer, 2010; Ware, 2012).

En este trabajo se incluye una nueva dimensión: la narratividad de la visualización o el *storytelling* tematizado por Segel y Heer (2010) y presentado como potencial ámbito de desarrollo de las visualización contemporánea por Kosara y Mackinlay (2013). En este apartado procedemos a desglosar estas tres dimensiones que consideramos fundamentales para la concepción, creación y diseño de visualizaciones que se dirigen a usuarios no expertos. Las explicaciones se acompañan de ejemplos en medios de comunicación con el objeto de ilustrar el potencial de las técnicas de visualización para la divulgación de la información.

2.1. Percepción: capacidades intrínsecas y compartidas

El aprovechamiento de los recursos visuales para representar los fenómenos, ya sean datos, relaciones o procesos, es la base que fundamenta la infovis. Como ventajas destacan:

- la amplificación de la cognición humana, puesto que centra la carga del trabajo en el sistema perceptivo en lugar del cognitivo, así como la agrupación de la información y la facilidad de representar una alta densidad de datos en poco espacio;
- reconocimiento de patrones;
- reconocimiento frente a la memorización (Molich; Nielsen, 1999);
- organización de los datos en relaciones estructurales (Card; MacKinlay; Shneiderman, 1999).

Los datos se codifican de tal forma que los ojos pueden discernir y el cerebro entender (Few, 2013). Para ello los teóricos identifican las propiedades “preatentivas” o “preconscientes”, que son aquellas que llaman la atención de nuestra percepción incluso antes de ser procesadas por nuestro cerebro, como color, forma, movimientos, texturas, posición espacial...

«Vamos más allá de los modelos de visualización de información habitualmente establecidos, incorporando el *storytelling* (narratividad)»

Una propiedad es preatentiva cuando se percibe de forma rápida y sin esfuerzos. Ware (2012) habla del fenómeno *pop-up*, para referirse a la cualidad de emergencia de estas propiedades. Escoger estas características y decidir cuál servirá para codificar cada atributo o dimensión de los datos es clave para la legibilidad de la visualización. Sumándolas se puede favorecer una lectura integrada o una lectura que codifica de forma separada estas dimensiones. El uso combinado del color y la forma para codificar atributos permite una lectura más integrada que el color y el movimiento, por ejemplo.

Según la aportación de Ware, se distinguen tres niveles en el proceso de percepción:

- emergencia de las cualidades preatentivas –forma, color, movimiento, posición espacial-,
- el segundo nivel procesa estas características para favorecer la búsqueda de patrones;
- percepción de objetos.

Un ejemplo del uso de estas propiedades lo encontramos

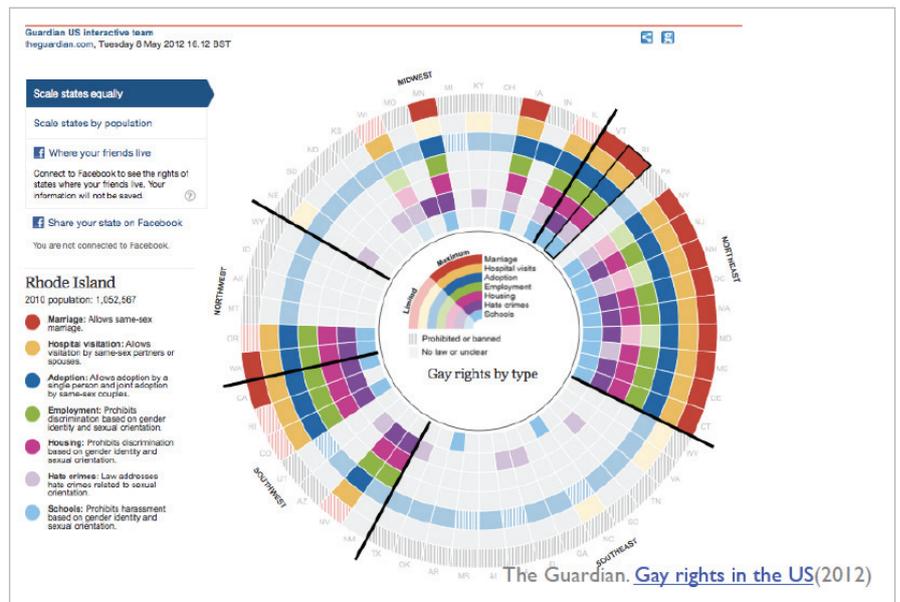


Figura 1. Derechos de los homosexuales en EUA (Guardian News & Media Ltd 2012). <http://www.theguardian.com/commentisfree/ifamerica/2012/may/08/gay-rights-activists-history-movement>

en la figura 1, que representa el grado de respeto a los derechos de los homosexuales en los diferentes estados de EUA. En ella destacamos la utilización de las propiedades preatentivas como el color para identificar los tipos de derecho y la luminancia para mostrar el grado de respeto de estos derechos (menos luminancia-mayor reconocimiento). Por ejemplo en Dakota del Norte se indica con azul claro (más luminoso) el reconocimiento a la adopción para padres solteros; en cambio en Oregon, en azul oscuro, se representa el derecho de adopción a padres solteros y parejas del mismo sexo. Además, a través de la distribución espacial se agrupan los estados por zonas. De este modo en una sola ojeada podemos afirmar que en el noreste del país, casi todos los estados reconocen el matrimonio homosexual (rojo oscuro).

2.2. Interacción: apropiación e interiorización del tema

Las nuevas tecnologías permiten dinamizar gráficos hasta ahora estáticos y así posibilitar la exploración de los datos en función de los diferentes criterios del usuario, dejando en segundo plano las visiones fruto de la voluntad del autor. A partir de la puesta en contexto se dota al usuario de un rol *interlocutivo*.

Las herramientas interactivas de análisis permiten reducir los efectos del *big data* (Heer; Kandel, 2012), porque permiten al usuario apropiarse de la visualización. En términos de percepción, un diseño eficaz posibilita y vehicula la formulación de hipótesis. La interacción dispone las condiciones para que emerja un proceso iterativo de pregunta-respuesta, paralelo al reconocimiento y la comprensión. Es decir, en contraste con los diseños estáticos, la dimensión de la interacción incorpora el análisis visual, caracterizado por un proceso de creación de visiones, exploración y refinamiento (Heer; Shneiderman, 2012). De esta forma se activa una dinámica de construcción y verificación de hipótesis, circuito necesario para desencadenar que el usuario se apropie del tema representado y entre en diálogo con el diseño visual.



Figura 2. Índice para una vida mejor. (OCDE, 2013). <http://www.oecdbetterlifeindex.org/es>



Figura 3. Índice para una vida mejor (OCDE, 2013). <http://www.oecdbetterlifeindex.org/es>

La interacción es clave para adaptarse a cada contexto de búsqueda de información donde el entorno se distingue por la cantidad de datos a codificar y la diversidad de objetivos y motivaciones de los usuarios. Por ello no es extraño que Ware (2012) considere que los elementos informativos tienen que ser activos: representar información cuando son necesarios, desaparecer cuando son innecesarios y aceptar el control del usuario para apoyar los procesos de pensamiento.

“La interacción permite dotar al usuario de un rol interlocutivo frente a la visualización”

Para conseguir esta actividad, Shneiderman (1996) conceptualiza el paradigma de la exploración visual, conocido también como el “mantra” de la búsqueda de información: “primero visión general, zoom y filtro, y luego detalles bajo demanda”.

La interfaz tiene que ser lo más transparente posible para facilitar la exploración. La virtud de transparencia dependerá de los conocimientos previos de los usuarios, pero se puede optimizar la visibilidad y aquello que se esconde, como también se puede facilitar la comprensión de las posibilidades

de la visualización y de la misma interfaz. Existen recursos que incrementan la potencialidad³ de los elementos que integran una visualización, muchos de ellos relacionados con las propiedades preatentivas, que permiten e invitan al usuario a que los objetos sean utilizados para la finalidad para la que han sido concebidos. Además existen otros recursos como la metáfora, que encontramos constantemente cuando navegamos (por ejemplo, cuando con el icono de la mano desplazamos el mapa de la pantalla, o con la lupa nos acercamos a él).

En la visualización creada por la OCDE sobre el Índice para una vida mejor (figura 2), los países se representan por una flor cuyos pétalos corresponden a temas o categorías –vivienda, ingresos, empleo...-. Ya no sólo por el texto que invita al usuario a interactuar, sino también por el panel de la derecha que permite controlar las visiones, el usuario reconoce esta capacidad y la ejecuta. Como resultado de las acciones realizadas por el usuario, la visión presentada cambia.

Pongamos por ejemplo que consideramos la “vivienda” como el tema de mayor importancia para nuestro Índice para una vida mejor (figura 3). La visión resultante muestra la organización de los datos y su representación según los criterios del usuario: se resalta el pétalo verde de cada país, que indica la vivienda, y los países cambian de posición. Esta iteración pregunta-respuesta se puede ir repitiendo ya que cada visión resultante puede plantear una duda o una hipótesis que la persona que interactúa va a querer responder, dentro y fuera de la visualización.

2.3. Storytelling: un hilo conductor para mantener el espectador

La teoría de la percepción y las posibilidades de interacción han sido estudiadas de forma amplia para desarrollar visualizaciones y han generado conocimiento en torno a la noción de espacio de diseño (Kosara; Mackinlay, 2013). A pesar de estos avances, cuando los datos tienen que ser explicados fuera de la comunidad de expertos es necesario poner en marcha un mecanismo que apele al lector y mantenga su atención. Con la finalidad de gestionar la información proporcionada para vehicular la toma de conocimiento, el hilo conductor emerge como un elemento a tener en cuenta. De forma semejante al espectador que mira una película, al usuario de una visualización la narrativa le revela la información de forma efectiva e intuitiva (Gershon; Page, 2001).

Esta nueva dimensión que empieza a considerarse en la visualización de la información emerge como un elemento clave para la comunicación de datos. Entender las estructuras narrativas tomadas de otras disciplinas como la literatura o el periodismo, y especialmente en el ámbito de la comunicación audiovisual, puede proporcionar una solución a los nuevos retos comunicativos (Quesenberg; Brooks, 2010). Esta perspectiva incluye la estructura interna del mensaje, una planificación que tiene un guión y múltiples historias que permiten introducir al usuario en la visualización. Incluso se puede considerar inherente a la visualización, pero no siempre ha sido bien gestionada ni tenida en consideración. El mantra de la búsqueda de información de Shneiderman (1996) define, más allá de una guía para facilitar la explotación, un mecanismo que se corresponde con la introducción, el nudo y el desenlace, que caracteriza las fábulas o cuentos propios de gran parte de los géneros populares.

La gran diferencia entre las películas y las visualizaciones – por ahora- desde el punto de vista de la narración radica en el control de la progresión, la secuencialidad de los eventos (Segel; Heer, 2010). En función del grado de intervención del autor, del abanico de opciones y control que tiene el usuario, podemos establecer diferentes tipos de narrativa en las visualizaciones. De la excelente categorización de Segel y Heer (2010) destaca la estructura del “vaso de Martini”, una combinación de dos esquemas: se inicia con una aproximación guiada por el autor para, más avanzada la exploración, traspasar las riendas al receptor. Como consecuencia del primer contacto guiado por el autor se consigue contextualizar al usuario, mostrar el alcance de la visualización y qué tareas puede llevar a cabo. En la segunda parte, el usuario controla y explora para verificar hipótesis y reformularlas, es decir, poner en marcha la búsqueda interactiva que comentábamos anteriormente. La gestión del tiempo tiene un papel esencial en la representación de datos. Un claro ejemplo sobre el aprovechamiento de la animación para revelar datos lo encontramos en la herramienta *Trendalyzer* de *Gapminder*. Se trata de un software de visualización de información para animar estadísticas, desarrollado por Hans Rosling.

<http://www.gapminder.org/tag/trendalyzer>

El *treemap* es la representación de un árbol de datos, en el que cada nodo puede tener 0 o más hijos, y se representa por un rectángulo cuyas medidas y color reflejan los valores de los datos. *The New York times* (figura 4) escoge un *treemap* para representar la propuesta de Obama sobre la

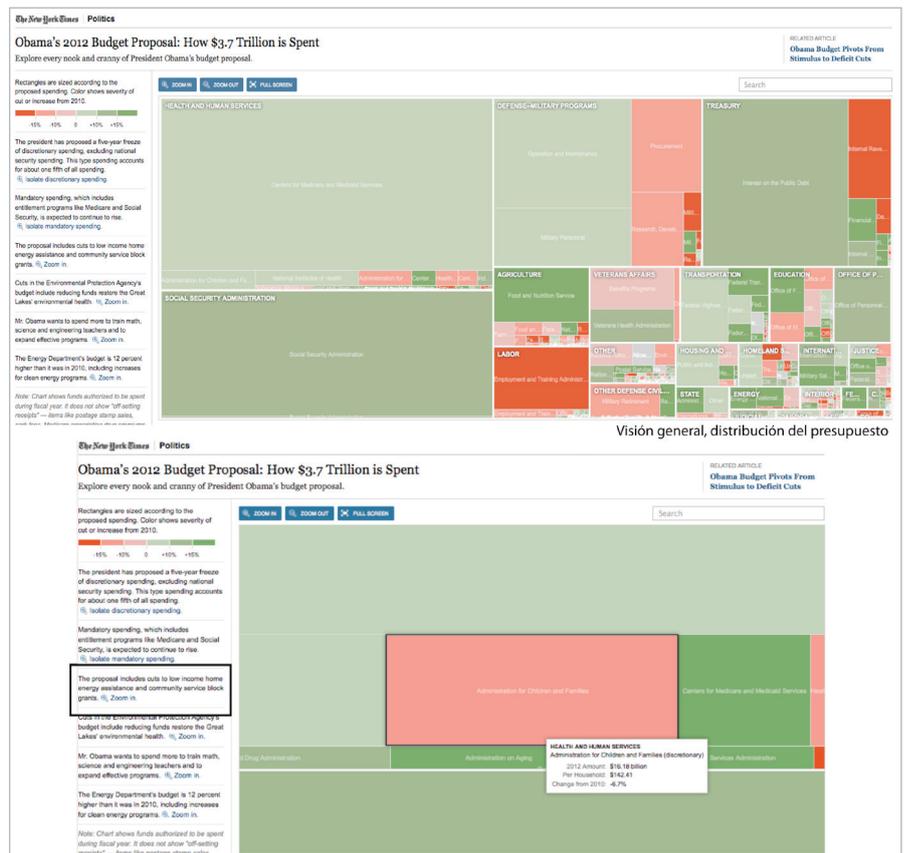


Figura 4. Propuesta de Obama para el presupuesto 2012 (*The New York times*, 2011). <http://www.nytimes.com/packages/html/newsgraphics/2011/0119-budget>

distribución del presupuesto para el 2012. La coordinación entre las historietas y las partes que corresponden a cada una en el *treemap* se activan al clicar cada *lead* para hacer un zoom en la parte representada. De esta forma el autor guía o propone diferentes lecturas, que a su vez sirven para mostrar el funcionamiento de la visualización y así actuar también como instrucciones de uso.

“ las estructuras narrativas tomadas de otras disciplinas como la literatura o el periodismo pueden proporcionar una solución a los nuevos retos comunicativos ”

3. Asegurar la correspondencia entre emisor y receptor: el diseño centrado en el usuario (DCU)

Aunque se ha teorizado alrededor de las teorías de la percepción y la interacción, pocas veces se ha verificado el impacto real en los usuarios. Además, cuando se ha escogido un enfoque que incorpora la persona en el proceso se observa una tendencia a favor de la perspectiva de la interacción persona-ordenador (Lloyd, 2010). Wassink et al. (2009) valoran los pocos estudios que han conseguido traspasar sus resultados a guías de diseño o heurísticos en su revisión sobre las diferentes técnicas de aproximación a los usuarios en el diseño de visualizaciones interactivas destina-

das a usuarios expertos. La necesidad de disponer de guías para el diseño de visualizaciones con carácter comunicativo y destinadas a una audiencia más heterogénea es aún más patente.

La perspectiva del diseño centrado en el usuario tuvo su origen en el campo militar en los años 50, se introdujo posteriormente como aproximación para diseñar productos cotidianos, y a mediados de los años 80 tuvo su auge como perspectiva adoptada para el desarrollo de software en el entorno digital. **Norman** (1988) apunta que la finalidad del diseño centrado en el usuario (DCU en adelante) es conseguir que los productos sean utilizados y comprensibles a partir de conocer los intereses y necesidades del usuario. Para ello el DCU adopta diferentes métodos y técnicas, tanto cualitativas como cuantitativas para conocer todas las variables que rodean al usuario final –entorno, motivaciones, limitaciones.

“ El *storytelling* ayuda a gestionar la información proporcionada con el fin de vehicular la toma de conocimiento ”

Como su nombre indica, en el DCU el usuario es el centro que rige las decisiones de diseño y es tangible en todo el proceso puesto que se introduce a lo largo del ciclo de diseño. Destacan técnicas como el prototipado, la técnica personas, escenarios, casos de uso, diseños de interacción, etc.

En 2010 se publica la *ISO 9241-2014*, para sustituir la *ISO 13407* con cierto grado de abstracción (**Jokela et al.**, 2003), y dar respuesta a la cantidad de manuales publicados relativos a este enfoque. La nueva *ISO* plantea requisitos en lugar de recomendaciones:

- Entender y especificar el contexto de uso (incluyendo usuarios, tareas y entorno) (análisis).
- Especificar los requerimientos del usuario con suficiente detalle para conducir el diseño e introducir soluciones de diseño (análisis).
- Producir soluciones de diseño que encajen con los requerimientos (diseño).
- Conducir evaluaciones centradas en el usuario de las soluciones de diseño y modificar éste teniendo en cuenta los resultados (evaluación).

Nielsen (1993) además, define el concepto de usabilidad, inherente al DCU, como la suma de cinco atributos, todos ellos cuantificables: facilidad de aprendizaje, eficiencia, facilidad de memorización, tasa de errores y satisfacción.

En el caso de las visualizaciones tratadas en este trabajo, con finalidad divulgativa, será esencial definir la audiencia, por plural que sea, pues para determinar la usabilidad deben definirse los usuarios, sus objetivos concretos y el entorno de uso.

Introducir esta perspectiva para tener presente los diferentes perfiles de usuarios y poder responder a sus requisitos, puede apoyar el diseño de visualizaciones que capten el interés, que respondan a las necesidades y encajen en consultas de tipo informativo o de ocio para la multitud. Se propone a nivel práctico la incorporación de técnicas y metodologías que estudian el usuario, lo definen en diferentes grados de concreción a la vez que lo abstraen para gestionar mejor el proceso de diseño y validación.

A modo de ejemplo del uso de la aplicación de la metodología DCU en el campo de la infovis, la figura 5 presenta un prototipo elaborado para el diseño de una visualización que representa los movimientos migratorios internos en Catalunya. A partir de él se validaron los elementos utilizados para codificar los datos y se verificó si comprendía los recursos adecuados (destacados en color) para llevar a cabo las tareas.

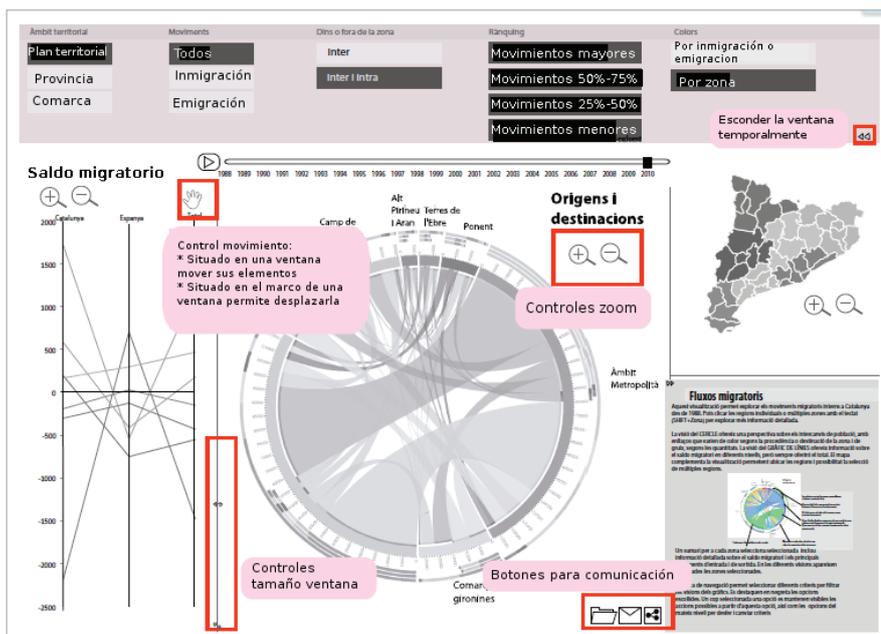


Figura 5. Prototipo de visualización sobre los movimientos migratorios en Catalunya. Controles de zoom, tamaño de las ventanas, opción para esconder las ventanas, control para mover la imagen en el espacio y botones para desarrollar las tareas de comunicación

4. Conclusiones

La visualización de la información cumple con los requisitos necesarios para situarse como herramienta privilegiada para la transmisión de conocimiento. Su capacidad representativa, donde prima lo visual por encima de lo textual, permite trabajar de forma rápida y con menor necesidad de procesar en el cerebro. Según los modelos de usabilidad, la percepción directa mejora la eficiencia pues reduce el tiempo de procesamiento, únicamente perceptivo (**Card; Mackinlay; Shneiderman**, 1999). También veremos que mostrar los datos en visualizaciones cumple con las heurísticas clásicas de usabilidad (**Molich; Nielsen**, 1990). La visualización se configura a partir de elementos que minimizan el esfuerzo del usua-

rio, pues es una representación estética y minimalista de los datos, para garantizar que su atención altamente limitada en una marea de estímulos, se centre en los aspectos importantes, la información esencial, la que se quiere comunicar. Ello cumple también con la heurística de prevención de errores al maximizar la captación de atención. Al mismo tiempo, el equilibrio entre el titular y la información en detalle puede verse satisfecho gracias a la interacción, dando al usuario un mayor control y libertad sobre los datos mostrados. Los múltiples caminos que éste puede explorar según sus criterios, sus vicisitudes, sus necesidades y sus motivaciones permiten resituarlo en un nuevo espacio privilegiado donde hay lugar para la identificación, la interiorización y, por ende, para la apropiación de la información, ya sea para acumular conocimiento como para ir más lejos aún y desencadenar decisiones.

Es necesario disponer de guías para el diseño de visualizaciones con carácter comunicativo y destinadas a una audiencia heterogénea y no experta

Una visualización interactiva ofrece flexibilidad y eficiencia de uso. Aprovechar estas propiedades para representar la complejidad de los fenómenos puede ampliar la audiencia de contenidos que antes quedaba reducida a una minoría de expertos y a la vez reducir el ruido en la comunicación del mensaje.

Para incrementar las posibilidades de éxito del acto comunicativo, la incorporación de mecanismos narrativos en las visualizaciones se presenta como una nueva dimensión complementaria a la percepción y la interacción, y vendría a completar los anteriores beneficios con ayudas y documentación para guiar al usuario. Aunque se trate de un principio relativamente reciente y poco estudiado en el marco de la infovis, la presencia de visualizaciones que acompañan artículos de divulgación en los medios de comunicación avala su eficacia. Con la tendencia del *open data* es una dimensión que cobra notoriedad para la reformulación del periodismo.

La propuesta de incluir la narratividad en la infovis debe ser validada pues hay un gran vacío de trabajos que comprueben la eficacia de las visualizaciones, y una necesidad de casos de estudio con resultados extrapolables que permitan la elaboración de guías de diseño. Para articular las dimensiones tratadas en este estudio (percepción, interacción y narrativa), el diseño centrado en el usuario puede proveer un marco de trabajo que garantice el éxito de la comunicación. El DCU permitirá evitar prejuicios sobre los modelos mentales del usuario, definir mejor la audiencia y establecer aquellos elementos que se adecúan mejor a su perfil.

5. Notas

1. Nuestro estudio parte de la definición de usuario no experto de Heer *et al.* (2008) el cual se caracteriza por tener cierto nivel en el manejo de herramientas informáticas o como usuario de internet convencional, que lo utiliza para mantenerse informado pero no es experto en la visualiza-

ción de la información. Para reforzar la finalidad comunicativa del estudio, las visualizaciones consultadas por no expertos tendrán lugar en el contexto de utilización proporcionado por Sprague y Tory (2012): fuera del trabajo, con motivaciones casuales, informativas y de ocio.

2. En España en los últimos tres años han nacido programas de *Máster en Periodismo de Datos (URJC)* y posgrados específicos en *Visualización de la información (IDEC, UPF)*, al mismo tiempo que se han introducido asignaturas de Visualización de datos como en el programa de *Máster en Gestión de Contenidos Digitales (UB)*.

3. Del inglés *affordance* introducido inicialmente por Gibson y adaptado por Norman (1988).

4. ISO 9241-210:2010. *Ergonomics of human-system interaction. Part 210: Human-centered design for interactive systems*. TC/SC: ISO/TC 159/SC 4

6. Bibliografía

Card, Suart; Mackinlay, Jock; Shneiderman, Ben (1999). *Readings in information visualization: using vision to think*. Academic Press. Morgan Kaufman Publishers. San Francisco. ISBN: 1 55860 533 9

Cukier, Jerome (2009). "Can data visualization improve knowledge and decision-making?" En: Kosara, Robert; Cohe, Sarah; Cukier, Martin. *Panel: Changing the world with visualization. IEEE Visualization Conference Compendium*. http://kosara.net/papers/2009/Kosara_InfoVisPanel_2009.pdf

Few, Stephen (2013) "A more thoughtful but no more convincing view of big data". *Perceptual edge. Visual bussines intelligence. A blog by Stephen Few*. <http://www.perceptualedge.com/blog/?p=1671>

Finn, Holly (2011). "Lunch with Hal Varian". *Think quarterly. The data issue 2011*. <http://www.google.co.uk/think/articles/lunch-with-hal.html>

Fuhrmann, Sven; Ahonen-Rainio, Paula; Edsall, Robert; Fabrikant, Sara; Koua, Etien; Tobón, Carolina; Ware, Colin; Wilson, Stephanie (2005) "Making useful and usable geo-visualization". En: Dykes, Jason; Maceachren, Alan; Kraak, Menno-Jan *Design and Evalutaion Issues. Exploring Geovisualization*, pp. 553-566. ISBN: 978 0 08 044531 1 <http://dx.doi.org/10.1016/B978-008044531-1/50446-2>

Gershon, Nahum; Page, Ward (2001). "What storytelling can do for information visualization". *Communications of the ACM*, v. 44. n.8, pp. 31-37 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.7625&rep=rep1&type=pdf> <http://dx.doi.org/10.1145/381641.381653>

Heer, Jeffrey; Kandel, Sean (2012). "Interactive analysis of big data". En: *XRDS: Crossroads. The ACM magazine for students*, v. 19, n. 1, pp 50-54. <http://dx.doi.org/10.1145/2331042.2331058>

Heer, Jeffrey; Shneiderman, Ben (2012). "Interactive dynamics for visual analysis". *Communications of the ACM*, v. 55, n. 4, pp. 45-54.

<http://idl.cs.washington.edu/files/2012-InteractiveDynamics-CACM.pdf>

<http://dx.doi.org/10.1145/2133806.2133821>

Heer, Jeffrey; Van-Ham, Frank; Caprendale, Sheelagh; Weaver, Chris; Isenberg, Petra (2008). "Creation and collaborations: engaging new audiences for information visualizations" *Information visualization LNCS 4950*. Springer-Verlag, pp. 92-133.

http://dx.doi.org/10.1007/978-3-540-70956-5_5

Jokela, Timo; Iivari, Netta; Matero, Juha; Karukka, Minna (2003). "The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11" En: *Proceedings of the Latin American conference on Human-computer interaction (CLIH'03)*. ACM, New York, pp. 53-60.

<http://mcom.cit.ie/staff/Computing/prothwell/HCI/papers/ISO%20Standards%20Analysis%202003%20Jokela.pdf>

<http://dx.doi.org/10.1145/944519.944525>

Kosara, Robert; Mackinlay, Jock (2013). "Storytelling: next step for visualization" En: *IEEE Computer Society*, v. 46, n. 5, pp. 44-50.

http://kosara.net/papers/2013/Kosara_Computer_2013.pdf

<http://dx.doi.org/10.1109/MC.2013.36>

Lloyd, David (2010). *Evaluating human-centered approaches to geovisualization*. City University London.

Molich, Rolf; Nielsen, Jakob (1990). "Improving a human-computer dialogue". *Communications of the ACM*, v. 33, n. 3, pp. 338-348.

http://cpe.njit.edu/dlnotes/cis/cis732_447/cis732_2r.pdf

<http://dx.doi.org/10.1145/77481.77486>

Nielsen, Jakob (1993) *Usability engineering*. Boston, Massachusetts: Morgan Kaufmann Publishers, San Francisco. ISBN: 0125184050

Norman, Donald (1988). *The psychology of everyday things*.

Basic Books. ISBN: 0465067093

Quesenbery, Whitney; Brooks, Kevin (2010). *Storytelling for user experience*. Ronsenfeld Media. ISBN: 1 933820 03 9

Segel, Edward; Heer, Jeffrey (2010) "Narrative visualization: telling stories with data". En: *IEEE Transactions on Visualization and Computer Graphics*, pp. 1139-1148.

<http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf>

Shneiderman, Ben (1996) "The eyes have it: a data type taxonomy for information visualizations" En: *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336-343.

<http://www.cs.umd.edu/~ben/papers/Shneiderman1996eyes.pdf>

<http://dx.doi.org/10.1109/VL.1996.545307>

Spence, Robert (2001). *Information visualization*. ACM, New York. ISBN: 978 0201596267

Spence, Robert (2007). *Information visualization: design for interaction*. Pearson/Prentice Hall, Second edition. ISBN: 978 0132065504

Sprague, David; Tory, Melanie (2012). "Exploring how and why people use visualization in casual contexts: modeling user goals and regulated motivations". *Information visualization*, v. 11, n. 2, pp. 106-203.

<http://dx.doi.org/10.1177/1473871611433710>

Ware, Colin (2012). *Information visualization: perception for design*. Morgan Kaufmann Publishers. Third edition, San Francisco. ISBN: 0123814642

Wassink, Ingo; Kulyk, Olga; Van-Dijk, Betsy; Van-der-Veer, Gerrit; Van-der-Vet, Paul (2009). "Applying user-centered approach to interactive visualisation design". En: *Trends in interactive visualisation, advanced information and knowledge processing*, pp. 175-199.

<http://doc.utwente.nl/64840/1/WassinkKulyk.pdf>

http://dx.doi.org/10.1007/978-1-84800-269-2_8

Suscripción EPI sólo online

Pensando sobre todo en los posibles suscriptores latinoamericanos, ya no es obligatorio pagar la suscripción impresa de EPI para acceder a la online.

EPI se ofrece a instituciones en suscripción "sólo online" a un precio considerablemente más reducido (101,65 +21% IVA euros/año = **123 euros/año**), puesto que en esta modalidad no hay que cubrir los gastos de imprenta ni de correo postal.

Para los suscriptores particulares, el precio de la suscripción "sólo online" es de **62 + 21% IVA euros/año = 75 euros/año**.



VISUALIZACIÓN Y RECUPERACIÓN DE INFORMACIÓN EN LA WEB SEMÁNTICA



Jorge Morato, Sonia Sánchez-Cuadrado, Alejandro Ruiz-Robles y José-Antonio Moreiro-González



Jorge Morato es doctor en documentación y trabaja en el *Departamento de Informática* de la *Universidad Carlos III de Madrid*. Sus principales publicaciones están relacionadas con la utilización y construcción automática de tesauros y ontologías en la mejora de la recuperación de información y gestión del conocimiento.

<http://orcid.org/0000-0002-7530-9753>

Universidad Carlos III de Madrid, Departamento de Informática
Avda. Universidad, 30. 28911 Leganés (Madrid), España
jmorato@inf.uc3m.es



Sonia Sánchez-Cuadrado es doctora en documentación y ha sido profesora del *Departamento de Informática* de la *Universidad Carlos III* durante más de una década. Actualmente trabaja en sistemas de procesamiento del lenguaje natural en publicidad SEO. Su actividad investigadora está enfocada a la extracción de información, reconocimiento de patrones, procesamiento del lenguaje natural y sistemas de organización del conocimiento.

<http://orcid.org/0000-0002-7722-1982>

JOT Internet Media
Gral. Ramírez de Madrid, 6-8. 28020 Madrid, España
ssanhec@gmail.com



Alejandro Ruiz-Robles es máster en ciencia y tecnología informática por la *Universidad Carlos III de Madrid (UC3M)* y profesor a tiempo completo por la *Universidad de Piura (Perú)*. Actualmente cursa el doctorado en la *UC3M* y su principal área de investigación es la gestión del conocimiento.

<http://orcid.org/0000-0002-5280-679X>

Universidad de Piura, Dep. Ingeniería Industrial y de Sistemas
Av. Ramón Múgica, 131. Apdo. postal 352, Piura, Perú
alejandro.ruiz@udep.pe



José-Antonio Moreiro-González es catedrático de biblioteconomía y Documentación de la *Universidad Carlos III de Madrid*. Es autor de trabajos sobre técnicas de análisis de contenido documental y cuestiones conceptuales de la documentación. Ha sido director del *Departamento de Biblioteconomía y Documentación* y decano de la *Facultad de Humanidades, Comunicación y Documentación*.

<http://orcid.org/0000-0002-8827-158X>

Univ. Carlos III de Madrid, Depto. de Biblioteconomía y Documentación
C/ Madrid, 126. 28903 Getafe (Madrid), España
jamore@bib.uc3m.es

Resumen

Se cuestiona el funcionamiento de los motores de búsqueda para la web semántica y si la visualización gráfica de los resultados mejora la recuperación. Se han analizado los tipos de consultas que se pueden realizar y las principales características de un buscador semántico. Se ha seleccionado una consulta y se ha ejecutado en buscadores semánticos de acceso online con la *DBpedia*. Además, se han estudiado las funciones básicas y deseables en un buscador de la web semántica y se han revisado algunos estudios sobre la evaluación de buscadores semánticos. Por último se analizan las soluciones gráficas para mostrar los resultados de algunos buscadores. Las principales conclusiones inciden en la idea de que las soluciones de visualización de los resultados de las consultas no cumplen las expectativas propuestas en la bibliografía. .

Palabras clave

Datos abiertos, Datos enlazados, Web semántica, Buscadores, Motores de búsqueda, Consultas, Visualización, Representaciones gráficas, Evaluación, Navegación conceptual, *DBpedia*.

Artículo recibido el 31-01-2014

Aceptación definitiva: 06-06-2014

Title: Information visualization and retrieval in the semantic web

Abstract

The authors have questioned how the semantic web search engines work and if the graphical display of the search results improves the retrieval. The types of queries that can be performed, as well as the main features of a semantic search engine, have been analyzed. A query was selected and executed in semantic search engines with online access to *DBpedia*. The basic and desirable functionality in a semantic web search engine, and reviewed studies about the evaluation of semantic search engines have been studied. Finally, there have been analyzed the graphic solutions for displaying the results of some search engines. The main findings underlines the idea that graphical solutions to display query results do not meet the expectations proposed in the bibliography.

Keywords

Linked open data, LOD, Semantic web, Queries, Search engines, Visualization, Graphical display, Evaluation, Conceptual browsing, *DBpedia*.

Morato, Jorge; Sánchez-Cuadrado, Sonia; Ruiz-Robles, Alejandro; Moreiro-González, José-Antonio (2014). "Visualización y recuperación de información en la web semántica". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 319-329.

<http://dx.doi.org/10.3145/epi.2014.may.12>

1. Introducción

El proyecto *linked open data* (LOD) se basa en la propuesta de **Berners-Lee** de publicar datos enlazados en la web mediante el uso de URIs, que permiten nombrar las entidades y establecer enlaces que relacionen recursos (**Bizer; Heath; Berners-Lee**, 2009). Esta información se formaliza con estándares como RDF, con sentencias en forma de tripletas con la secuencia recurso-atributo-valor. Posteriormente estas tripletas pueden ser interrogadas con lenguajes como *Sparql*, donde la potencia en la recuperación se produce por la combinación de tripletas de diferentes recursos.

En 2011 el proyecto *LOD* ya contaba con 295 *datasets*, con más de 31 millones de tripletas (**Bauer; Kaltenböck**, 2011) y diferentes recursos interrelacionados, caracterizándose por las enormes dimensiones de sus componentes. Por ejemplo *FreeBase* está formado por 43 millones de tópicos y 2,3 millones de hechos. *DBpedia* contiene 3,7 millones de tópicos, incluyendo 832.000 personas o 639.000 lugares.

<http://dbpedia.org/About>

<http://www.freebase.com/browse>

El elevado número de conceptos interrelacionados hace patente la necesidad e importancia de un método intuitivo para organizar, recuperar, filtrar, visualizar y navegar entre conceptos y acceder a sus recursos vinculados. En el estudio de **Méndez y Greenberg** (2012) se recogen y sintetizan diversas propuestas desde la perspectiva de *LOD*.

2. Consulta y navegación en la web semántica

Tipos de consultas y respuestas

Las consultas que caracterizan la web semántica consisten en la formulación de una única respuesta a una pregunta dada o a la obtención de un listado relacionado con una entidad, más que a una localización de palabras claves contenidas en documentos en lenguaje natural. Se interroga sobre un dato simple como una fecha, un nombre, un lugar (ele-

mentos también conocidos como entidades del nombre), o bien sobre listados de entidades con atributos o facetas comunes que permitan filtrar la información. Según la propuesta clásica de **Broader** (2002), los tipos de consultas son:

- Informacionales: pueden ser tanto de instancias de tipos de datos (listado de músicos alemanes, fecha de las olimpiadas de Londres, etc.) como de refinamiento o filtrado de resultados por características, facetas, etc. (músicos alemanes del siglo XX con más de cinco sinfonías).
- Navegacionales: para, dado un recurso encontrar otros similares (webs sobre tecnologías para almacenar energía), o bien, dado un concepto encontrar documentos relacionados (sitios con fotografías de monumentos en París).
- Transaccionales: aquellas que permiten interactuar con el sitio una vez localizado.

“ La interrogación semántica implica conocer a priori dónde preguntar y cómo se representaron los conceptos y sus relaciones ”

Obviamente esta clasificación no está orientada, por ejemplo, a la localización de recursos de conocimiento formalizados (como ontologías, vocabularios de metadatos, etc.). Un ejemplo podría ser localizar un vocabulario de metadatos que incluya conceptos relacionados con un corpus de documentos. Ante estas carencias, investigadores como **Uren et al.** (2007) sugieren otros tipos de consultas como: consulta de entidad, relacional (que busca relaciones entre entidades) y parametrizada, que sería similar a las conocidas en documentación como facetadas. De forma análoga, **Morato et al.** (2013) identifican consultas conceptuales, contextuales y las que permiten el acceso a un recurso. También se han categorizado según las necesidades de información

del usuario (Wei; Barnaghi; Bargiela, 2008; Stranskas; Tomassen, 2010). Otras tipologías hacen hincapié en la formulación de la consulta: Fazzinga y Lukasiewicz (2010) clasificaron las consultas semánticas en estructuradas, con palabras clave y en lenguaje natural.

Los distintos tipos de consultas pueden devolver tipos de resultados diferentes: una consulta sobre un dato normalmente puede requerir una única tripleta; un listado puede precisar un conjunto de tripletas; y la petición de una o más páginas asociadas a un concepto, puede devolver una página enlazada mediante una o más tripletas. Por ejemplo, para obtener “un listado de municipios españoles y la población correspondiente a cada uno de ellos”, se crea una consulta y los datos resultantes pueden proceder de varias combinaciones de tripletas.

Tipos de buscadores semánticos

Presentan una gran diversidad según sus características. Su interfaz de búsqueda puede presentar formularios, grafos o permitir el uso de un lenguaje de recuperación. Lopez et al. (2011) centraron la clasificación en aspectos como: el ámbito, el grado de estructuración de las fuentes, los tipos de datos para expresar la consulta y el contexto de búsqueda (desambiguación, multilingüismo, ranking o fiabilidad). Además, el buscador puede tener o no una ontología subyacente para navegar semánticamente entre los conceptos.

La satisfacción del usuario ante la complejidad de la web semántica no puede medirse con métricas típicas de recuperación de información

Algunos permiten búsquedas sólo contra un *dataset* o un vocabulario, mientras que otros permiten consultas federadas contra más de un conjunto de datos. También varían en cuanto a la usabilidad y conocimiento técnico requerido por el usuario. Por último cada buscador utiliza un lenguaje de interrogación, y una ontología que puede determinar la potencia y complejidad para realizar la consulta.

Algunas propuestas plantean buscadores similares a Google en cuanto a un formulario sencillo como interfaz de búsqueda. Es el caso de Swoogle, Sindice, Falcons o Dr.Watson, que

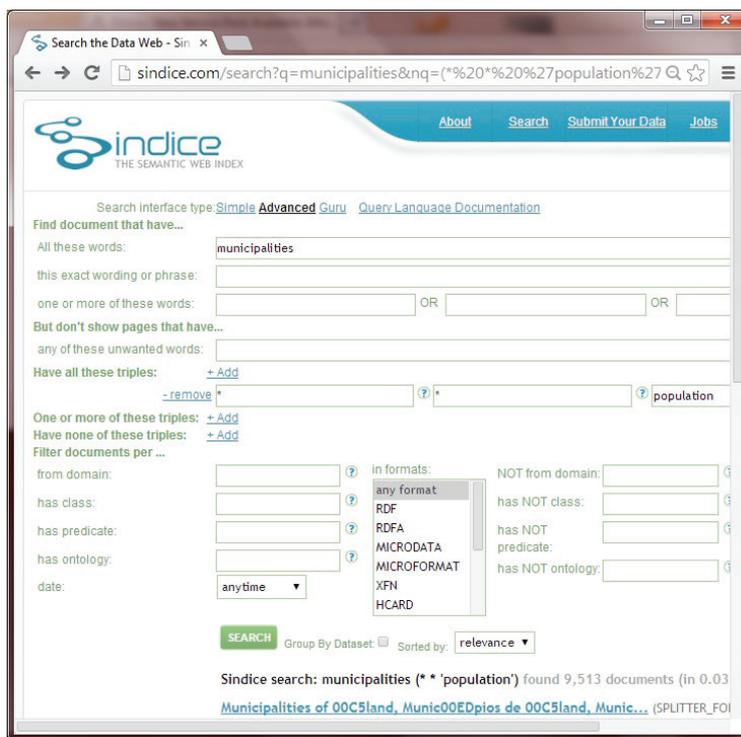


Figura 1. Búsqueda avanzada en Sindice para buscar la población de municipios. <http://sindice.com>

presentan una interfaz despejada que permite una consulta sencilla basada en una palabra clave. Sin embargo, esta aparente usabilidad contrasta con los interrogantes planteados al usuario sobre cómo se ha representado cada concepto, como se observa en la figura 1, Lo que supone un conocimiento especializado previo sobre las consultas y las respuestas que se pueden obtener.

Estos buscadores proporcionan una solución parcial para resolver los tipos de consultas de la web semántica. Por una parte por la dificultad de interrogar al buscador, cuando se desconoce el recurso o si no se tiene un conocimiento de los lenguajes de la web semántica (Miličić, 2011; Morato et al., 2013), y por otra, por la dificultad añadida que supone la interpretación de los resultados, bien por el modo en el que se ordenan, o bien por cómo se muestra el resultado y su contexto.

En cuanto a las opciones de búsqueda y el lenguaje de interrogación para los buscadores semánticos, dependiendo del tipo de buscador y de consulta se exigen algunos conocimientos a los usuarios como:

- Conocer un recurso que contenga esta información.
- Conocer la estructura de conocimiento que defina el con-

Tabla 1. Comparación de las definiciones para el elemento “agent”

Vocabulario	Definición elemento “agent”
NSDL Registry http://metadataregistry.org	The person or organization having a role with the vocabulary
MuDiCat http://www.ifla.org	A person (author, publisher, sculptor, editor, director, composer, etc.) or a group (family, organization, corporation, library, orchestra, country, federation, etc.) or an automaton that has a role in the lifecycle of a resource
FOAF http://www.foaf-project.org	An agent (eg. person, group, software or physical artifact)
VCard http://tools.ietf.org/html/rfc6350	An entity who may sometimes act on behalf of the entity associated with the vCard

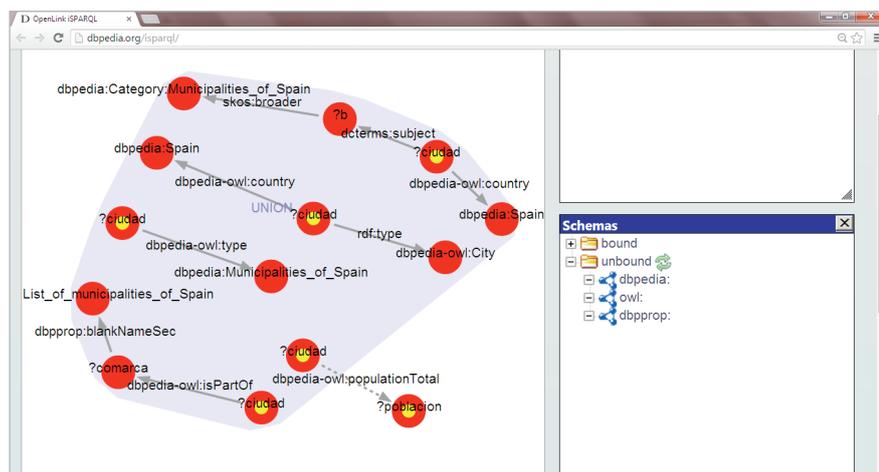


Figura 2. Grafos alternativos sobre población de municipios en España con iSPARQL <http://dbpedia.org/isparql>

cepto general en ese recurso y cómo está representado el dato. Un mismo concepto puede estar representado por criterios diferentes.

- Ser consciente de las diferencias de matices entre las definiciones alternativas referidas a la representación de un concepto. Esto es, una misma codificación puede representar conceptos más genéricos, más amplios o más específicos dependiendo del vocabulario.
- Para consultas complejas, conocer el lenguaje de interrogación del buscador (como Sparql).
- Valorar la consulta más eficaz en el caso de que existan alternativas.

En el siguiente punto se detallan ejemplos de los elementos discutidos en esta enumeración.

Representación y codificación de los conceptos

Un mismo concepto puede ser representado en distintos vocabularios o esquemas de metadatos con diferente terminología y codificado en diferentes lenguajes de representación, por eso se habla de representación y codificación de conceptos. Si se desconoce en qué vocabulario se encuentra el concepto, una aproximación razonable para realizar la consulta sería utilizar recursos que abarquen múltiples vocabularios como *Open metadata registry (OMR)* o *Linked open vocabularies (LOV)*. No obstante, nos encontraremos con los problemas asociados a la polisemia y falta de contexto. Por ejemplo en la búsqueda para el concepto “agent”, el OMR lo encuentra en 4 vocabularios y LOV lo localiza en 88. Si extrae-

Tabla 2. Consultas Sparql para recuperar la población de municipios: prefijos dbpedia: <http://dbpedia.org/resource>; dbpedia-owl: <http://dbpedia.org/ontology>; category: <http://dbpedia.org/resource/Category>; dbpprop: <http://dbpedia.org/property>

1	SELECT DISTINCT ?ciudad ?poblacion WHERE {{?ciudad dbpedia-owl:type dbpedia:Municipalities_of_Spain} OPTIONAL {?ciudad dbpedia-owl:populationTotal ?poblacion.}}
2	SELECT DISTINCT ?ciudad ?poblacion WHERE {{?ciudad rdf:type dbpedia-owl:City. ?ciudad dbpedia-owl:country dbpedia:Spain} OPTIONAL {?ciudad dbpedia-owl:populationTotal ?poblacion.}}
3	SELECT DISTINCT ?ciudad ?poblacion WHERE {{?ciudad dbpedia-owl:country dbpedia:Spain. ?ciudad dcterms:subject ?b. ?b skos:broader category:Municipalities_of_Spain} OPTIONAL {?ciudad dbpedia-owl:populationTotal ?poblacion.}}
4	SELECT DISTINCT ?ciudad ?poblacion WHERE {{?ciudad dbpedia-owl:isPartOf ?comarca. ?comarca dbpprop:blankNameSec dbpedia:List_of_municipalities_of_Spain} OPTIONAL {?ciudad dbpedia-owl:populationTotal ?poblacion.}}

mos y comparamos los resultados se aprecia que el solapamiento de las definiciones es parcial (tabla 1), por lo que el concepto que representan podría no ser exactamente el mismo. <http://lov.okfn.org/dataset/lov/index.html>

<http://metadataregistry.org>

Algunos trabajos, como el de **Fuentes-Lorenzo, Morato y Gómez (2009)**, proponen mostrar el contexto en lenguaje natural y anotadores que conectados a vocabularios simplifiquen su codificación. No obstante, sin visualizar las relaciones del término, el problema persiste.

Formulación de la consulta

Hemos desarrollado el ejemplo propuesto previamente sobre “un listado de municipios españoles y la población correspondiente a cada uno de ellos” en la *DBpedia*. En la tabla 2 observamos la formulación de consultas alternativas en Sparql, pero no equivalentes en su resultado. En el ejemplo las entidades buscadas pueden estar categorizadas como: *Municipalities_of_Spain* (consulta 1), una consulta más limitada con *city* con un valor de filtrado igual a *Spain* (consulta 2), ciudades que tengan el término genérico *Municipalities_of_Spain* (consulta 3), o una consulta equivalente a la tres para el caso de que haya una subdivisión administrativa como provincia o similar (consulta 4).

Se confirma que la formulación de esta consulta implica al menos tres dificultades:

- conocimiento de la sintaxis del lenguaje de búsqueda y la posibilidad de combinar tripletas y filtros;
- identificación de los conceptos sobre los que se han representado los datos;
- en caso de consultas alternativas, selección de la formulación con datos más confiables.

3. Resultados de las consultas

La figura 2 muestra los grafos resultado de las cuatro consultas alternativas sobre población de municipios en España del apartado anterior en *OpenLink iSPARQL*. Observamos como cada secuencia de nodos muestra la relación que los

une, y cada una de estas secuencias corresponde a una de las consultas realizadas en el apartado anterior.

Aunque todas las búsquedas devuelven resultados válidos, al compararlas se observa que no son equivalentes: no se obtienen los mismos datos en cuanto a nombres, número de municipios y población.

Los datos obtenidos se han comparado con los publicados por el *Instituto Nacional de Estadística (INE)* para 2013. El *INE* recoge 8.249 municipios en España. Se han considerado las variaciones ortográfica o idiomática. En la tabla 3 se observa el número de lugares recuperados y coincidentes con los recogidos en el *INE*. Se han calculado también las tasas de precisión y exhaustividad sin tener en cuenta el dato de población ni el año al que corresponde ese valor

Una observación más pormenorizada de los datos muestra que entre los resultados de las consultas 1 y 2 el porcentaje de ciudades coincidentes es sólo de un 3% y aproximada-

Tabla 3. Exhaustividad y precisión respecto a los datos del *INE* <http://www.ine.es>

Pregunta	Coincidentes <i>INE</i> (T test 95%)	Exhaustividad	Precisión	Ciudades encontradas sin dato población (%)
1	1.442 (p=0,022)	0,17	0,88	7
2	192 (p=0,084)	0,02	0,85	1
3	4.911 (p=0,003)	0,59	0,92	2
4	1.325 (p=0,043)	0,16	0,86	3

Tabla 4. Propuestas orientadas a mostrar de un modo gráfico las relaciones

Propuestas	Principal función	Problema que afronta
<i>VisualSparql</i> y <i>Welkin</i> http://graves.cl/visualSparql http://simile.mit.edu/welkin	Visualizar gráficamente consultas <i>Sparql</i>	Dificultad de uso del lenguaje de recuperación
Di-Martino (2010)	Transformar consulta <i>PLN</i> en <i>Sparql</i> y representarla gráficamente	Dificultad de traducir el lenguaje natural al de consulta
<i>Cypher</i> http://docs.neo4j.org/chunked/milestone	Lenguaje de consulta gráfico con bases de datos gráficas <i>Neo4j</i>	Dificultad de uso del lenguaje de recuperación
<i>Iconvis</i> y <i>Wikimindmap</i> http://iconvis.polito.it/iconvis http://www.wikimindmap.org	Búsqueda gráfica secuencial y facetada de entidades del nombre	Problemas para restringir el ámbito de búsqueda
<i>Sgvizler</i> http://dev.data2000.no/sgvizler	Resultados con distintos tipos de gráficos (histogramas, dispersión, geolocalización, etc.)	Escasa adaptación de los buscadores a distintos tipos de consultas
Mejía-Sánchez-Bermejo (2013)	Desambiguación mediante el cálculo de la distancia semántica en el grafo	Problemas asociados a la denominación de un concepto en un vocabulario debido a sinonimia y polisemia
<i>Thinkpedia</i> y <i>Thinkbase</i> (Hirsch; Hosking; Grundy , 2009)	Visualización de vocabularios	Dificultad de navegar debido a la abundancia de conceptos

ciudad	poblacion
http://dbpedia.org/resource/Barcelona	3218071
http://dbpedia.org/resource/Barcelona	1620943
http://dbpedia.org/resource/Murcia	442573
http://dbpedia.org/resource/Palma_Majorca	401270
http://dbpedia.org/resource/Las_Palmas	383308
http://dbpedia.org/resource/Bilbao	351629
http://dbpedia.org/resource/C%C3%B3rdoba,_Andalusia	325453
http://dbpedia.org/resource/Valladolid	311501
http://dbpedia.org/resource/Gij%C3%B3n	281649
http://dbpedia.org/resource/L'Hospitalet_de_Llobregat	258642
http://dbpedia.org/resource/A_Coru%C3%B1a	246056
http://dbpedia.org/resource/Vitoria-Gasteiz	238247
http://dbpedia.org/resource/Jundiz_recycling_plant	235661
http://dbpedia.org/resource/Elche	230112
http://dbpedia.org/resource/Oviedo	224005
http://dbpedia.org/resource/Santa_Cruz_de_Tenerife	222417
http://dbpedia.org/resource/Badalona	220977
http://dbpedia.org/resource/Cartagena_Spain	218210

Figura 3. Resultado de la consulta 1 sobre la población de municipios

mente del 84% para las parejas 1 y 3, y 3 y 4. En el caso de ciudades coincidentes entre las consultas 2 y 3 es un 96%. Por ejemplo Madrid sólo se obtiene en la consulta 3 y Barcelona se obtiene en todas salvo en la consulta 2. Los datos de las poblaciones coincidentes de la *DBpedia* y el *INE* son significativamente diferentes. Con estos datos se muestra que no todas las formulaciones tienen los mismos niveles de ruido y silencio en la recuperación.

A modo de ejemplo, en la figura 3 se muestran los resultados de la consulta 1. En el listado completo se detectan ausencias significativas, incoherencias, e incluso datos erróneos. Otro ejemplo, que dificultaría las consultas relacionadas (*Uren et al., 2007*) sería la consulta “*Select ?a ?b Where {?a skos:broader ?b. ?b skos:broader ?a.}*” que proporciona como resultado 1.566 conceptos cuyo genérico y específico es el mismo concepto.

Los principales problemas encontrados en la recuperación son:

- dificultad para formular consultas en el lenguaje de recuperación debido a su complejidad;

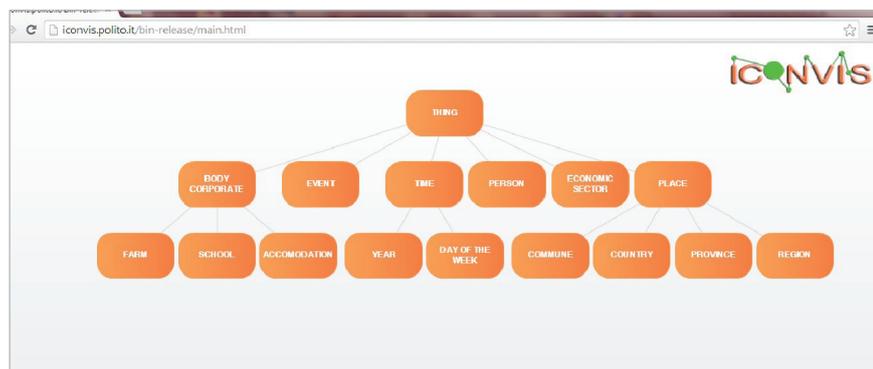


Figura 4. Consulta en Iconvis <http://iconvis.polito.it/iconvis>

- posibles inconsistencias en las definiciones entre vocabularios al definir el mismo concepto;
- problemas asociados a la denominación de un concepto en un vocabulario debido a sinonimia y polisemia;
- dificultad para navegar debido al volumen de conceptos;
- dificultad para saber a qué vocabulario/s debemos interrogar;
- escasa usabilidad y funcionalidad de los buscadores de la web semántica a los distintos tipos de consultas y usuarios no expertos;
- problemática con moléculas RDF, es decir dificultad de expresar determinadas ideas en una única tripleta.

Visualización y navegación web

El volumen de datos que puede producir una consulta y la dificultad para interpretar los resultados ha dado lugar a iniciativas encaminadas a visualizar gráficamente los vocabularios, tratando de simplificar la consulta y contextualizando los conceptos. **Hirsch, Hosking y Grundy (2009)** señalan que la web social, la web semántica y las distintas técnicas de

visualización de información pueden interactuar entre sí para simplificar y optimizar el proceso de consulta. Es decir, se trata de mejorar la visualización de los recursos de la web semántica, dado que esta visualización tiene como objetivo representar objetos de información de forma que el acceso a la información y al contenido subyacente sea más sencillo y comprensible (**Baeza-Yates; Rivera-Loaiza; Velasco, 2004**). Dependiendo de la información, ésta podrá ser visualizada como un dato, una imagen, una tabla o un grafo.

La complejidad del lenguaje *Sparql* y las deficiencias en la visualización limitan la popularidad de los recursos semánticos

Algunos ejemplos de herramientas orientadas a la visualización y la navegación web son *VisualSparql*, *Welkin*, *Iconvis*, *Cypher*, entre otras muchas. Por ejemplo, *Iconvis* (figura 4) permite navegar entre nodos asociados de forma gráfica. Cada nodo se expande y muestra las clases y propiedades asociadas, convirtiendo el evento en un conjunto de consultas *Sparql*. Tras varios nodos se muestra el resultado en la *Wikipedia*.

La tabla 4 recopila propuestas que muestran de modo gráfico las relaciones entre conceptos enlazados semánticamente, y expone la función más relevante y las principales dificultades que conlleva su uso.

Tabla 5. Elementos para el análisis de sistemas de recuperación semántica

Uren et al. (2007)	Mangold (2007)	Strasunskas y Tomassen (2010)	Morato et al. (2013)
Entorno de búsqueda: escala, heterogeneidad y portabilidad	Arquitectura	Arquitectura	
Tipos de consulta	Necesidades de información del usuario	Objetivo de la búsqueda	Tipos de consulta
Posibilidad de realizar consultas de forma iterativa y exploratoria: refinamiento, recomendación y reutilización	Posibilidad de modificación de la consulta. Transparencia e interactividad	Expresión de búsqueda Input del usuario (palabras clave, lenguaje natural, gráficos, consulta formal e interactiva)	
	Estructura de la ontología	Sistemas de organización del conocimiento (KOS)	Gestión semántica: desambiguación, multilingüismo, sinonimia, reuso
	Formalización de la ontología	Formalización de la ontología (RDFS, OWL, ...)	Lenguaje en que se expresa Posibilidad de modificar el esquema para mejorar la interoperabilidad, formalización, interactividad
	Acoplamiento entre ontología y documentos		Posibilidad de gestionar el marco semántico y capacidad de modificar ámbito, extensibilidad, modificabilidad
		Alcance (según sea aplicación de escritorio o web)	
Problemas intrínsecos: Comprensión, ranking de resultados y alineamiento			

4. Criterios para la evaluación de buscadores semánticos

Algunos autores sostienen que las medidas clásicas de precisión y exhaustividad son aplicables a la evaluación de motores semánticos de recuperación (Tümer; Shah; Bitirim, 2009; Andago; Phoebe; Thanoun, 2010). En general estos estudios no evalúan recursos de la web semántica, sino motores de documentos en lenguaje natural que aplican técnicas de extracción semántica y sistemas de organización del conocimiento.

Otros autores como Strasunskas y Tomassen (2010) señalan que la satisfacción del usuario al enfrentarse a la complejidad de una aplicación de un recuperador semántico no puede ser medida con estas métricas de recuperación de la información. La solución según estos autores es una evaluación que incluya aspectos cualitativos, como la calidad de los documentos formalizados, la calidad de la ontología subyacente, la arquitectura, usabilidad o la dificultad de la consulta. Diversos autores (Esmaili; Abolhassani, 2006; Uren et al., 2007; Mangold, 2007; Morato et al., 2013; Strasunskas; Tomassen, 2010) citan los siguientes elementos a evaluar:

- análisis de los buscadores según el tipo de consulta;
- calidad de los sistemas de organización de conocimiento (relaciones de dominio, meronimia, sinonimia, jerarquía, instancias y negación);
- similitud entre documentos y la ontología;
- criterios de indización de los documentos con la ontología;
- capacidad para representar, modificar o expandir la consulta o interactuar con el usuario;
- presentación o visualización de los resultados.

Muchos de estos factores son interdependientes, por ejemplo, los criterios de indización o la capacidad de expandir la consulta dependen de la riqueza semántica representada en el sistema de organización de conocimiento.

Como se puede observar en la tabla 5, todos los autores consideran relevante los tipos de consulta en la evaluación. Por otro lado, algunos señalan la necesidad de incluir una evaluación de la interfaz de búsqueda (Strasunskas; Tomassen, 2010), mientras otros se centran en las características de la ontología subyacente (Morato et al., 2013). Hirsch, Hosking y Grundy (2009) mencionan que dada la diferente calidad de los recursos subyacentes, se dificultan las pruebas de usabilidad

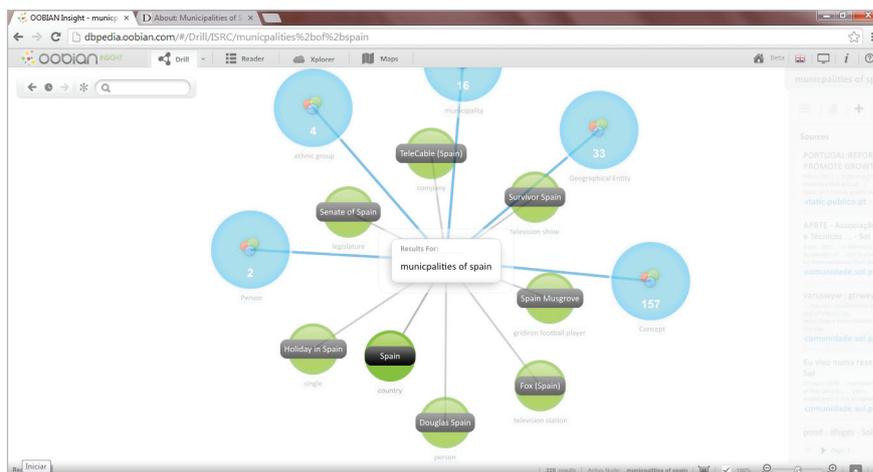


Figura 5. Oograph

heurística en los usuarios. Como alternativa, estos investigadores proponen recopilar los comentarios de los usuarios al navegar por el recurso.

Existen variantes para representar el mismo concepto y formular la misma consulta

5. Análisis de las interfaces gráficas de consulta semántica

Vistos los problemas en la recuperación, se han analizado interfaces gráficas orientadas a facilitar la formulación de consultas semánticas. El objetivo es determinar en qué medida son una solución a algunos de los problemas discutidos. Se ha restringido el estudio a aquellas interfaces que pueden ser consultadas online sin precisar una instalación local (como es el caso de Allegrograph). También para evitar problemas derivados de la existencia de varias ontologías se ha limitado el estudio a aquellos recursos que operan con DBpedia.

Es necesario analizar los buscadores semánticos en escena-

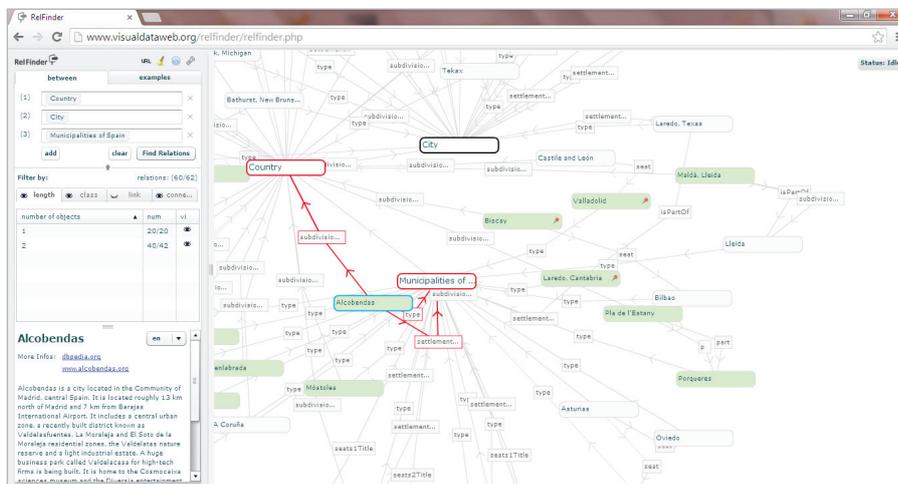


Figura 6. Visual dataweb

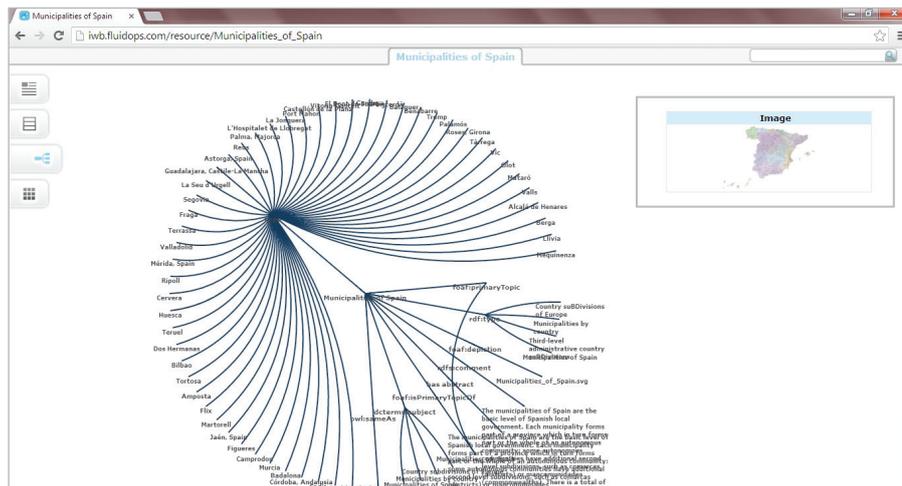


Figura 7. *Linked open data demonstrator*

rios que incorporen elementos como interactividad, filtrado y diversos tipos de consulta. Se ha propuesto una tarea consistente en una ampliación de la consulta de ejemplo sobre municipios españoles, posteriormente refinando a los diez más poblados y accediendo a páginas web relacionadas. Señalando las principales ventajas y desventajas a partir de nuestra experiencia con el buscador. Finalmente se ha realizado una comparativa de algunos aspectos indicados en la tabla 5.

A continuación se analizan algunas de las características de los buscadores:

Oobian

<http://dbpedia.oobian.com>

Aplicación desarrollada por *Maisis*, compañía especializada en gestión del conocimiento y de contenidos. Es una de las que mejores resultados ha proporcionado para probar el escenario propuesto.

- Ventajas: posee una interfaz intuitiva. Permite refinar la consulta y la visualización es rápida, accediendo a recursos web relacionados.
- Desventajas: escasa capacidad de combinar entidades. La consulta *Sparql* generada no es visible.
- Escenario: presentó problemas para combinar entidades, descargar los listados, ordenar o filtrar resultados.

Visual dataweb relfinder

<http://www.visualdataweb.org/relfinder/relfinder.php>

La aplicación se subdivide en: *Relfinder*, especializado en buscar relaciones entre datos *RDF*, permitiendo acceder a nuevas tripletas. Utiliza *Adobe Flex*, lo cual le dota de mayor portabilidad. Por otro lado, *GFacet* permite una exploración a partir de cualquier elemento de la tripleta, con capacidad de filtros y de unir nuevos nodos.

- Ventajas: posibilidad de interrogar por varias tripletas. Muestra mucha información sobre las entidades interrogadas.

-Desventajas: al tener sus funciones divididas entre varias herramientas su uso es más difícil. Además de la ambigüedad de los términos sugeridos, el abarrotamiento de términos, la dificultad de filtrado y tiempos excesivos complican su uso.

-Escenario: aunque ha sido una de las más completas y permite interrogar por varias tripletas y elegir caminos alternativos, presenta problemas para filtrar o establecer un ranking sobre cuál es el camino más acertado.

Linked open data demonstrator
<http://iwb.fluidops.com>

Es una plataforma para *linked data*, que integra varias fuentes (*DBpedia*, *Twitter* o *Facebook*). Permite integrarlo con otros recursos como *big data*.

- Ventajas: fácil de utilizar, y gráficos novedosos. Contiene recursos asociados a los términos. Fácil de consultar en lenguaje natural, pero no para todos los tipos de consultas. Es sencillo obtener un listado con las distintas tripletas a partir de un recurso.
- Desventajas: la visualización no es intuitiva, incluso la fuente llega a ser ilegible. Las consultas relacionales son complicadas de realizar.
- Escenario: las dificultades en interrogación simultánea por varios conceptos, obtención de listados y la gran densidad de conceptos dificultan su uso.

Wikistalker

<http://sepans.com/wikistalker/#>

Muestra los enlaces de cada artículo de la *Wikipedia*.

- Ventajas: visualización en un solo vistazo de todos los enlaces del artículo, es de las pocas que posee un ranking visual de resultados.
- Desventajas: la visualización se resiente del elevado número de enlaces, no permite búsquedas más allá de las de un artículo.

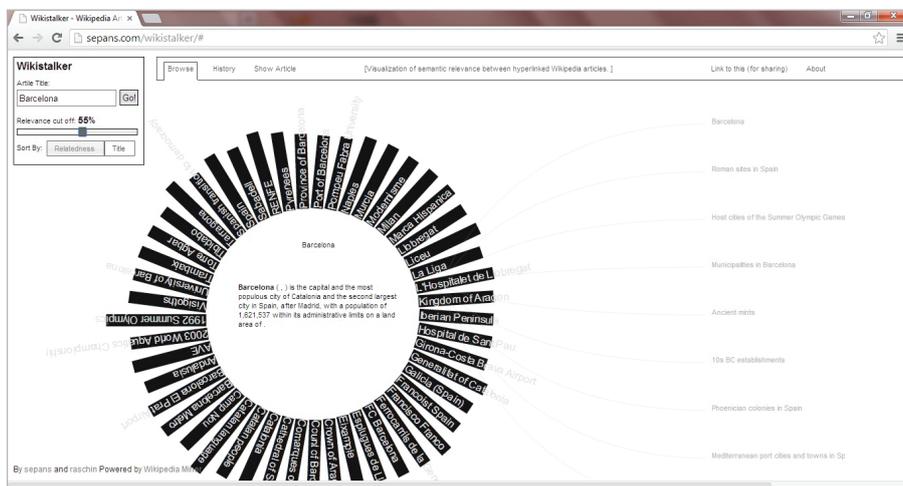


Figura 8. *Wikistalker*

- Escenario: aunque indica enlaces relacionados, el acceso al documento o la descarga de los resultados de la búsqueda no es posible. Como en otros casos, el acceso a varias tripletas de forma simultánea no forma parte de las funciones del sistema.

Lodlive
<http://en.lodlive.it>

Se trata de una interfaz usable y funcional. Se basa en tecnologías de gran difusión y libres como *JavaScript*, *jQuery*, *html* y *JSON*.

- Ventajas: facilidad para navegar entre resultados relacionados semánticamente.
- Desventajas: dificultad para consultar varias tripletas al mismo tiempo y acceder a los documentos de resultados.
- Escenario: la imposibilidad de combinar varias tripletas, descargar listados y documentos asociados y volumen de datos mostrados sin limitar el espacio de búsqueda hacen imposible la ejecución de las tareas.



Figura 9. Lodlive

Las interfaces gráficas muestran grandes volúmenes de datos pero escasas posibilidades de filtrar y combinar consultas

unitérminos, mostrando listas de conceptos relacionados. Por último, se vio la posibilidad de realizar un ranking bien entre consultas alternativas o con los resultados, como propone **Uren et al.** (2007).

En algunos datos en la comparativa de la tabla 6, la implementación es parcial, así en *Lodlive* la consulta *Sparql* se puede introducir, pero no se puede obtener la conseguida con la navegación. Como conclusión se puede observar que la ausencia de ranking, de descarga de tripletas, de búsqueda de elementos concurrentes e integración con *Sparql* limitan estas herramientas. Por lo tanto el escenario propuesto no ha podido completarse dado el escaso soporte para filtrados y consultas a múltiples conceptos (sólo *Visual dataweb* lo soporta).

Conclusiones

El uso de los recursos semánticos depende de la disponibilidad de buscadores orientados a los usuarios. Para ello, los motores de la web semántica deben optar por simplificar la interfaz de recuperación y representar los recursos y/o los resultados de forma gráfica, con la presunción de que una visualización de los recursos interrelacionados mejora la interpretación de los datos.

Tras probar la ejecución en diferentes buscadores semánticos se ha comprobado que no es factible a día de hoy pres-

En la tabla 6, basada en la tabla 5, se analizan aspectos relacionados con los tipos de consulta, viendo la posibilidad gráfica de expansión de la consulta a elementos relacionados semánticamente, acceso documental a páginas web relacionadas y posibilidad de descargar elementos de las tripletas resultantes (como listas de entidades). El refinamiento (**Uren et al.**, 2007; **Mangold**, 2007) se ha analizado observando la capacidad de buscar instancias en las que se solapan varios conceptos (ejemplo: solapamiento de conceptos población y municipio) y la posibilidad de filtrado (ejemplo: municipios con más población). Elementos como portabilidad (**Uren et al.**, 2007), multilingüismo y desambiguación (**Morato et al.**, 2013) y expresión de búsqueda (**Strasunskas**; **Tomassen**, 2010) se estudiaron mediante la posibilidad de desambiguación de homógrafos y visualización de la consulta generada en lenguaje *Sparql* o similar. Casi todas las herramientas permiten comenzar la consulta en lenguaje libre, pero con

Tabla 6. Comparación de herramientas para algunas funciones

	Oograph	Visual dataweb	LOD demonstrator	Wikistalker	Lodlive
Desambiguación homógrafos en la consulta	Sí	Parcial	Sí	Sí	Sí
Multilingüismo	Sí	Sí	Sí	Sí	Parcial
Búsqueda con expansión a relacionados	Sí	Sí	Sí	Sí	Sí
Búsqueda por varios conceptos concurrentes	No	Sí	No	No	No
Filtrado de resultados	Parcial	No	No	No	No
Ranking consultas o resultados	No	No	No	Sí	No
Acceso documentos resultantes	Sí	Sí	No	No	Sí
Descarga tripletas resultantes (ejemplo: listas)	No	No	No	No	No
Visualización de la consulta <i>Sparql</i>	No	No	No	No	Parcial

cindir del lenguaje de interrogación tipo *Sparql*. En cuanto a cómo se representan los datos, se obtienen resultados diferentes condicionados a la formulación de la consulta, que requiere conocer además los vocabularios de metadatos por los que está representado el recurso que se desea interrogar. Conocer la representación de los datos en los recursos también es una tarea ardua, porque aunque existe control de vocabulario en cada uno de los recursos mediante los vocabularios de metadatos, son muchos los conceptos similares pero representados de forma diferente. Existe por tanto control de vocabulario a nivel metadato, pero no al fusionarse con otros vocabularios, por lo que es necesario conocer previamente cómo están representados los datos antes de poder interrogarlos.

« Aun se requieren elementos que faciliten la recuperación y comprensión de los datos y recursos recuperados »

La consecuencia más inmediata es la complicación para usar el motor de búsqueda. En comparación con los buscadores de la web tradicional, quedan patentes las diferencias en cuanto al tipo de consulta (que está lejos de parecerse a una consulta por palabras clave), a los requisitos necesarios para formular la interrogación, y al tipo de recurso que se va a recuperar.

Estos buscadores difícilmente pueden ser evaluados sólo mediante métricas de precisión y exhaustividad. Creemos que deben añadirse características cualitativas asociadas a la usabilidad y funcionalidad. La capacidad de los buscadores para localizar determinada información dentro del gran volumen de documentos en la web semántica podría ser mejorada con algoritmos de posicionamiento para determinar los *datasets* o vocabularios más fiables.

La función de los buscadores visuales actuales se limita a mostrar los nodos directamente relacionados, y con la ejecución de las consultas complejas no se obtienen los resultados esperados. La visualización de los datos no es clara en aquellas consultas cuyos resultados son muy abundantes y con muchas relaciones. La navegación es compleja puesto que no existen directrices que orienten al usuario. Consideramos que aún se requieren elementos que faciliten la recuperación y comprensión de los datos y recursos recuperados.

Agradecimientos

Agradecemos la financiación recibida por el proyecto *Alma-histo* (HAR2011-27540) del *Mincin*.

Bibliografía

Andago, Martin O.; Phoebe, The P. L.; Thanoun, Bassam A. M. (2010). "Evaluation of a semantic search engine against a keyword search engine using first 20 precision". *International journal for the advancement of science & arts*, v. 1, n. 2, pp. 55-63.
<http://www.ucsiuniversity.edu.my/cervie/pdf/ijasa/paperV1N2IT3.pdf>

Baeza-Yates, Ricardo; Rivera-Loaiza, Cuauhtémoc; Velasco, Javier (2004). "Arquitectura de la información y usabilidad en la web". *El profesional de la información*, mayo-junio, v. 13, n. 3, pp. 168-178.
<http://www.elprofesionaldelainformacion.com/contenidos/2004/mayo/1.pdf>

Bauer, Florian; Kaltenböck, Martin (2011). *Linked open data: the essentials*. Viena: Semantic Web Company. ISBN: 978 3 902796 05 9
<http://www.semantic-web.at/LOD-TheEssentials.pdf>

Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2009). "Linked data—the story so far". *International journal on semantic web and information systems*, v. 5, n.3, pp. 1-22.
<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
<http://dx.doi.org/10.4018/jswis.2009081901>

Broader, Andrei (2002). "A taxonomy of web search". En: *Sigir Forum*, v. 36, n. 2, pp. 3-10.
<http://www.cis.upenn.edu/~nenkova/Courses/cis430/p3-broder.pdf>
<http://dx.doi.org/10.1145/792550.792552>

Di-Martino, Beniamino (2010). "An approach to semantic information retrieval based on natural language query understanding". En: *ICWE 2010 Workshop*, pp. 211-222.
http://dx.doi.org/10.1007/978-3-642-16985-4_19

Esmaili, Kyumars-Sheykh; Abolhassani, Hassan (2006). "A categorization scheme for semantic web search engines". En: *Aiccsa*, pp. 171-178.
<http://dx.doi.org/10.1109/AICCSA.2006.205086>

Fazzinga, Bettina; Lukasiewicz, Thomas (2010). "Semantic search on the web". *Semantic web: interoperability, usability, aplicability*, v. 1, n. 1-2, pp. 89-96.
<http://www.semantic-web-journal.net/sites/default/files/swj51.pdf>
<http://dx.doi.org/10.3233/SW-2010-0023>

Fuentes-Lorenzo, Damaris; Morato, Jorge; Gómez, Juan-Miguel (2009). "Knowledge management in biomedical libraries: a semantic web approach". *Information systems frontiers*, v. 11, n. 4, pp. 471-480.
http://eprints.networks.imdea.org/111/1/Knowledge_management_in_biomedical_libraries_a_semantic_web_approach.pdf
<http://dx.doi.org/10.1007/s10796-009-9159-y>

Hirsch, Christian; Hosking, John; Grundy, John (2009). "Interactive visualization tools for exploring the semantic graph of large knowledge spaces". En: *Workshop on visual interfaces to the social and the semantic web* (Vissw2009).
<http://ceur-ws.org/Vol-443/paper6.pdf>

Lopez, Vanessa; Uren, Victoria; Sabou, Marta; Motta, Enrico (2011). "Is question answering fit for the semantic web? A survey". *Semantic web: interoperability, usability, aplicability*, n. 2, pp. 125-155.
<http://www.semantic-web-journal.net/content/question-answering-fit-semantic-web-survey>
<http://dx.doi.org/10.3233/SW-2011-0041>

Mangold, Christoph (2007). "A survey and classification of

semantic search approaches". *International journal of meta-data, semantics and ontologies*, v. 2, n. 1, pp. 23-34.

<http://dee.srv1.eu/refs/files/1/65.pdf>

<http://dx.doi.org/10.1504/IJMSO.2007.015073>

Mejía-Sánchez-Bermejo, Antonio (2013). *Similitud semántica entre conceptos de Wikipedia*. Trabajo fin de grado. Universidad Carlos III de Madrid.

<http://hdl.handle.net/10016/17170>

Méndez, Eva; Greenberg, Jane (2012). "Linked data for open vocabularies and HIVE's global framework". *El profesional de la información*, mayo-junio, v. 21, n. 3, pp. 236-244.

http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_eng.pdf

<http://dx.doi.org/10.3145/epi.2012.may.03>

Miličić, Vuk (2011). "Introducing hypernotation an alternative to linked data". *Bew citnames*.

<http://milicicvuk.com/blog/2011/11/17/introducing-hypernotation-an-alternative-to-linked-data>

Morato, Jorge; Sánchez-Cuadrado, Sonia; Dimou, Christos; Yadav, Divakar; Palacios, Vicente (2013). "Evaluation of semantic retrieval systems on the semantic web". *Library hi tech*, v. 31, n. 4, pp. 638-656.

<http://dx.doi.org/10.1108/LHT-03-2013-0026>

Strasunskas, Darijus; Tomassen, Stein (2010). "On variety of semantic search systems and their evaluation methods".

En: *Procs of the Intl conf on information management and evaluation*, pp. 380-387.

Tümer, Duygu; Shah, Mohammad-Ahmed; Bitirim, Yiltan (2009). "An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, MSN and Hakia".

En: *4th Intl conf on internet monitoring and protection (Icimp'09)*, pp. 51-55.

<http://dx.doi.org/10.1109/ICIMP.2009.16>

Uren, Victoria; Lei, Yuangui; Lopez, Vanessa; Liu, Haiming; Motta, Enrico; Giordanino, Marina (2007). "The usability of semantic search tools: a review".

The knowledge engineering review, v. 22, n. 4, pp. 361-377.

<http://oro.open.ac.uk/23501/1/download.pdf>

<http://dx.doi.org/10.1017/S0269888907001233>

Wei, Wang; Barnaghi, Payam; Bargiela, Andrzej (2008). "Search with meanings: an overview of semantic search systems".

International journal communications of SIWN, n. 3, pp. 76-82.

<http://personal.ee.surrey.ac.uk/Personal/P.Barnaghi/doc/siwn2008Survey.pdf>

<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.324.2897>

iralis®

International Registry for Authors:
Links to Identify Scientists

es:

- una **guía** para los autores hispanos para que firmen sus trabajos en el formato internacional usual
- una **base de datos** que registra las variantes de firma usadas por cada autor en diferentes épocas
- un **buscador** que usa automáticamente todas las variantes registradas

¡Ni un día más sin decidir tu firma!

Inscríbete en <http://iralis.org>

**El formato de
firma internacional**

Con el apoyo de:

El profesional de la
información





BÚSQUEDA DE EXPEDIENTES DE FORMA VISUAL



Isabel Medrano-Corrales, Diego Ruiz-Macías y María-José Escalona-Cuaresma



Isabel Medrano-Corrales es licenciada en geografía e historia por la *Universidad de Sevilla*. Desde 1998 es jefa de archivo y documentación de la *Agencia de Obra Pública* de la *Junta de Andalucía*, donde ha participado en el proceso de diseño e implantación del módulo de gestión documental integrado en un sistema de gestión empresarial. Continúa trabajando y estudiando la optimización de un *enterprise resource planning* (ERP) en un entorno de documento electrónico e interoperabilidad. Es profesora colaboradora del máster de documentación digital de la *Universidad Pompeu Fabra*.
<http://orcid.org/0000-0002-3419-9180>

Agencia de Obra Pública de la Junta de Andalucía
Av. Diego Martínez Barrio, 10. 41071 Sevilla, España
isabel.medrano@aopandalucia.es



Diego Ruiz-Macías es ingeniero de telecomunicación, con intensificación de telemática por la *Escuela Superior de Ingenieros de Sevilla*. Es consultor senior en la empresa *Tecnocom España Solutions* desde 2007, especializado en proyectos de consultoría e implantación de sistemas de gestión documental y en el desarrollo de aplicaciones de gestión a media en tecnología *Java*. Anteriormente fue jefe de proyectos en la empresa *Caymasa El Sendero* especializado en proyectos basados en software libre con tecnología *Java*. Tiene especial interés en el uso de metodologías ágiles para el desarrollo de los proyectos.
<http://orcid.org/0000-0002-2819-3970>

Tecnocom España Solutions
Av. de la Innovación, s/n. Edif. Rentasevilla pl. 5. 41020 Sevilla, España
diego.ruiz@tecnocom.es



María-José Escalona-Cuaresma es ingeniera superior en informática por la *Universidad de Sevilla* donde consiguió su título de doctora europea en 2004. Es directora del *Grupo de Investigación de Ingeniería Web y Testing Temprano* y está adscrita al *Departamento de Lenguajes y Sistemas Informáticos* de la misma universidad donde trabaja como profesora titular. Como investigadora, sus líneas de trabajo están orientadas al campo de la ingeniería de requisitos y del aseguramiento de la calidad del software. Tiene una dilatada trayectoria en dirección de proyectos de investigación e investigación aplicada.
<http://orcid.org/0000-0002-6435-1497>

Universidad de Sevilla. ETS Ingeniería Informática
Av. Reina Mercedes, s/n. 41012 Sevilla, España
mjescalona@us.es

Resumen

Se muestran los resultados de la aplicación de técnicas de visualización de grandes conjuntos de documentos en el marco del *Proyecto de innovación de la gestión documental aplicada a expedientes de contratación de servicios y obras de infraestructuras de transporte*. En un entorno de documento electrónico, se parte de un esquema de metadatos como origen de la información que servirá para construir las visualizaciones. De los cuatro modos de visualización utilizados -*timeline*, *cluster*, nube de puntos y mapa- se describen con exhaustividad los metadatos empleados para cada uno de ellos y las representaciones visuales resultantes.

Palabras clave

Documentos de archivo, Expedientes de contratación, Archivos, Visualización, Búsqueda y recuperación de documentos, Líneas de tiempo, Nubes de puntos, Clusters, Mapas.

Title: Searching records visually

Abstract

We present the results of the application of visualization techniques to search large sets of records, under the framework of the "Project of innovation in record management applied to transport infrastructure services and works procurement dos-

Artículo recibido el 05-02-2014
Aceptación definitiva: 26-05-2014

siers". We started from a metadata schema as the origin of the information used to create visualizations. Four visualization techniques –timeline, cluster, cloud and map- and their metadata are exhaustively described.

Keywords

Records, Procurement dossiers, Archives, Visualization, Search and records retrieval, Timelines, Network diagrams, Clusters, Maps.

Medrano-Corrales, Isabel; Ruiz-Macías, Diego; Escalona-Cuaresma, María-José (2014). "Búsqueda de expedientes de forma visual". *El profesional de la información*, mayo-junio, v. 23, n. 3, pp. 330-338.

<http://dx.doi.org/10.3145/epi.2014.may.13>

1. Introducción

La búsqueda de documentos de archivo en un entorno electrónico conlleva el diseño de estrategias para traducir nuestra necesidad de información a las posibilidades de una interfaz de búsqueda más o menos avanzada (Marcos, 2004; Red.es, 2013). Cuando se trata de realizar búsquedas difusas en un buscador simple, nuestras expectativas de éxito rápido dependerán de las posibilidades de filtrado posterior que ofrezca el sistema. En los casos de búsquedas de fondos, divisiones de fondos y series documentales, las herramientas de búsqueda aún se vuelven más estrictas que si se trata de documentos.

Los referentes visuales que siempre hemos tenido en un entorno papel y nos han servido para orientarnos, se han desvanecido

Los referentes visuales que siempre hemos tenido en un entorno papel y nos han servido para orientarnos, se han desvanecido. Por eso consideramos interesante explorar nuevas formas de visualización. Reconocemos como pionera en España la iniciativa del *Archivo de la Ciudad de Arganda del Rey*, que ofrece las opciones de visualizar los documentos que resultan de una búsqueda en una línea del tiempo y un mapa.

<http://archivo.ayto-arganda.es/Búsquedas.aspx>

En el *Proyecto de innovación de la gestión documental aplicada a expedientes de contratación de servicios y obras de infraestructuras de transporte*¹, se incluye el empleo de técnicas de visualización de grandes conjuntos de datos a la localización de documentos. Los resultados de este proyecto –aún en fase de ejecución- se han diseñado para poder reutilizarse por cualquier órgano contratante de la *Junta de Andalucía* y, por extensión, por cualquier administración con competencias en materia de contratación de servicios y obras (Dietrich; Myrntinen, 2012). El proceso de adopción de los modelos y herramientas de software por parte de cualquier organismo interesado requerirá previa o simultáneamente dotarse de un nivel de identificación, valoración y descripción documental cualificada. Es decir, deberá contarse con los estudios de identificación y valoración de las series documentales de expedientes de contratación de servicios y obras, así como la descripción de dichos expedientes y los documentos que contienen. El

tratamiento archivístico es, si cabe, aún más imprescindible en un entorno electrónico.

Las visualizaciones se diseñaron en un principio para representar documentos –simples o expedientes-, pero durante el desarrollo de los trabajos se vio la oportunidad de aplicarlas también para visualizar relaciones entre otros niveles de agrupación: grupos de fondos, fondos, divisiones de fondos y series documentales.

2. ¿Qué entendemos por "forma visual"?

Entendemos por búsqueda y recuperación de documentos de "forma visual" la asistencia que se puede recibir de representaciones visuales que ayudan a comprender el volumen de información que se muestra y, además, sirven de orientación para afinar la búsqueda mediante la interacción con los elementos mostrados en la visualización.

Los objetivos generales que se pretenden alcanzar son:

- Facilitar la comprensión por los usuarios de grandes conjuntos de documentos de archivo; una práctica habitual en el caso de los documentos bibliográficos, para los que existe una tradición de estudios bibliométricos y de visualización. Ligado a ello se pretende intensificar la orientación al usuario, también adoptando maneras de hacer ya consolidadas en el ámbito bibliotecario (Bengochea; Patricia, 2005).
- Reducir el tiempo de búsqueda, ofreciendo al usuario opciones intuitivas con referentes visuales que le asistan.
- Visualizar la información contextual de los documentos contenida en metadatos; por ejemplo, visualizar los documentos en el tiempo, en el espacio, en los hitos de la tramitación administrativa y en sus relaciones con otros documentos.
- Evidenciar visualmente relaciones que habitualmente permanecen ocultas o escondidas entre las categorías de la entidad documento y que son muy útiles en la búsqueda de información; por ejemplo, visualizar las relaciones de un expediente con otros o de una serie documental con otras es algo que en un entorno papel es información registrada en un formulario o informe de uso casi exclusivamente interno.

Adicionalmente identificamos otros objetivos específicos en relación a la contratación de obras y servicios de transporte:

- Facilitar las búsquedas de documentos que componen los expedientes de contratación de obras y servicios de infraestructuras del transporte, que se caracterizan por ser voluminosos -por su gran número de documentos-, exten-

sos -sobre todo las obras de infraestructuras se prolongan en el tiempo hasta su recepción definitiva- y complejos – al contener documentos que en sí son objetos complejos, como los proyectos de construcción-.

- Presentar visualizaciones de los documentos orientadas al usuario, para lo cual se emplean los referentes que se conocen como más intuitivos para esta tipología documental: espacio (situación geográfica del objeto de cada expediente) (**Bosque-Sendra; Zamora-Ludovic**, 2002), tiempo (localización temporal de cada documento) y génesis (relación con la tarea que lo genera).
- Ofrecer posibilidades de interacción en las visualizaciones de documentos para que los usuarios puedan adaptar determinados parámetros de la representación de manera personalizada. Para ello previamente es precisa una labor de identificación de los metadatos que intervienen en estos procesos de interacción.

El principal obstáculo para la consecución de los objetivos anteriores es el riesgo de deficiencias en la calidad de los metadatos, que provoquen una extracción incierta y una visualización incorrecta.

Debemos explorar nuevas formas de visualización para hacer más comprensible la realidad de los documentos en el entorno electrónico

3. El origen de los datos para la visualización

La naturaleza y estado de los datos sobre los documentos objeto de visualización son una cuestión sustancial en el proceso. En un entorno electrónico los datos sobre los documentos son los metadatos y requieren estar definidos en un perfil de aplicación de un modelo de metadatos (**Chen**, 1999). Esta circunstancia propicia que la información estructurada pueda ser seleccionada, tratada y explotada mediante representaciones visuales de datos.

En el caso que nos ocupa, el perfil de metadatos sigue el *Esquema de metadatos para la gestión del documento electrónico (e-Emgde²)* propuesto por el *Ministerio de Hacienda y Administración Pública* en el marco de aplicación de la *Norma técnica de interoperabilidad de política de gestión de documentos electrónicos*.

Se ha optado por un modelo multientidad con cinco entidades:

- documento;
- agente;
- actividad;
- regulación;
- relación.

De todas ellas, será la entidad documento la que suministre la mayoría de la información que servirá para elaborar las visualizaciones. Dentro ella, se distinguen varias categorías con el objetivo de crear una estructura jerárquica documental: grupo de fondos, fondo, división de fondo, serie documental, expediente, documento.

El lugar natural donde se implementa el perfil de metadatos es el gestor documental. En él se debe situar toda la lógica de la información sobre los documentos electrónicos. De forma complementaria, se podrán desarrollar servicios que optimicen la extracción de datos para la visualización de los documentos. Esta ha sido la solución empleada en el proyecto que exponemos.

En lo que respecta a las visualizaciones, se han considerado cuatro tipos: timeline, clúster, nube de puntos y mapa.

A modo de esquema, en la figura 1 se muestran los metadatos ofrecidos por la entidad documento, en cada una de sus categorías, y que son utilizados por las visualizaciones. A continuación se analiza de dónde procede la información empleada en los distintos tipos de visualización propuestos:

3.1. Vista *timeline*

Muestra los documentos de un expediente situados a lo largo de una línea de tiempo, y las fases y tareas realizadas en su tramitación. Es por ello que en esta visión están involucradas la entidad “documento” -y en concreto, sus categorías “expediente” y “documento simple”- y la entidad “actividad”. La información utilizada para construir esta vista se obtiene de:

Entidad “documento”

a) Categoría “expediente”:

- El nombre del expediente <Nombre.NombreNatural> figura en la pantalla inicial de la visualización.
- La descripción del expediente <Descripción>.

b) Categoría “documento simple”:

- El nombre de cada documento limitado a 40 caracteres <Nombre.NombreNatural> se presenta en una viñeta en la barra temporal. Si seleccionamos esta viñeta, visualizamos en la parte superior el nombre completo del documento.
- La descripción <Descripción>.
- La fecha de captura <Fecha.FechaInicio>, que refleja la fecha de creación del documento en el sistema.
- Metadatos específicos que el usuario puede elegir para modificar la visualización, por ejemplo, otras fechas del documento relativas a su tramitación y que hacen que cambie la ubicación de cada uno en el timeline <Fecha.FechaDocumento>; o la valoración o no de un documento como esencial, criterio que reduce la visualización del expediente a sus documentos esenciales.

Entidad “actividad”

a) Las categorías “fase” y “tarea” están relacionadas con los tipos de documentos identificados en el procedimiento de contratación:

- De esta forma, el nombre de la fase o de la tarea <Nombre.NombreNatural> nos ayuda –junto con la fecha del documento- a contextualizar el documento.

3.2. Vista clúster

Modo de visualización que el usuario puede elegir para tener una visión de las relaciones de asociación que existen entre los nodos del cuadro de clasificación de documentos

(Fernández-Molina; Peis, 1999), hasta el nivel de serie documental. Para construir esta representación se utiliza información procedente de:

Entidad “relación”, categoría “relación de procedencia”

a) El metadato <TipoRelación>, que expresa como uno de los valores el tipo de relación de procedencia asociativo, que es el que nos interesa.

b) El número de relaciones de asociación, cuyo cómputo se emplea para aproximar los nodos con más relaciones y viceversa.

c) El número de relaciones de tipo “contiene” entre un elemento y sus hijos, para ajustar el tamaño de cada nodo al volumen de objetos existentes. Por ejemplo, los nodos que representan cada serie documental tienen el tamaño proporcional al número de expedientes de esa serie.

Entidad “documento”

a) Categorías “fondo”, “división de fondo” y “serie documental”:

- Nombre <Nombre.NombreNatural> del fondo, división de fondo o serie documental, que se presenta como etiqueta sobre su punto correspondiente en la nube.
- Descripción <Descripción> del fondo, división de fondo o serie documental que hayamos seleccionado y cuyo texto visualizamos en la parte inferior de la pantalla.
- Otros metadatos específicos que se hayan configurado en la pantalla de preferencias para que aparezcan como información de detalle del objeto seleccionado en la parte inferior de la pantalla.

3.3. Vista nube de puntos

Modo de visualizar un expediente mostrando su diagrama de relaciones –de subordinación o asociación– con otros expedientes. Obtiene la información de la entidad “documento” y, en concreto, de las categorías “expediente” y “documento simple”, ya que la nube de puntos mostrará solamente las relaciones entre los elementos de estas categorías:

- Nombre del expediente o documento <Nombre.NombreNatural>.
- Descripción del expediente o documento <Descripción>.
- Otros metadatos específicos que hayamos configurado en nuestra pantalla de preferencias para que nos aparezcan como información de detalle del objeto seleccionado, en la parte inferior de la pantalla.

3.4. Vista mapa

Se posicionan los expedientes en un sistema de cartografía. La información se extrae de la entidad “documento” y, en concreto, de la categoría “expediente”:

- Las coordenadas del objeto del expediente –es decir, las del proyecto u obra de infraestructura– se obtienen de los metadatos <PuntoAcceso.Lugar.Coordenadas.Latitud> y <PuntoAcceso.Lugar.Coordenadas.Longitud>.
- El nombre se obtiene del metadato <Nombre.NombreNatural> y se visualiza en un pop-up sobre el punto correspondiente en el mapa

- La descripción se obtiene del metadato <Descripción>.
- La información del pop-up puede completarse con otros metadatos específicos que se hayan configurado en la pantalla de preferencias para que aparezcan como información de detalle del objeto seleccionado.

Hay que ofrecer al usuario opciones más intuitivas para buscar documentos, en las que encuentre referentes visuales que le asistan

4. Tipos de visualizaciones de conjuntos de documentos de archivo según el modelo de grafo

A continuación se describen los cuatro tipos de visualizaciones utilizadas y se exponen los requisitos de uso:

4.1. Timeline

- Propósito

El objetivo es ofrecer una panorámica visual de los documentos que contiene un expediente, con dos referencias en la banda inferior:

- la línea de tiempo que marca la fecha del documento
- el nombre de la tarea que da origen al documento.

Es una alternativa a las habituales interfaces tabulares que presentan una lista de documentos, distribuida en pantallas consecutivas por las que hay que navegar. Para el caso de expedientes voluminosos y largos como los de contratación de obras y servicios de infraestructuras, es un intento de hacer más accesible la complejidad de este tipo documental.

El usuario puede moverse con rapidez en el expediente cronológicamente, ampliar períodos concretos para ver con más detalle las viñetas de documentos y visualizarlos en la parte superior cuando marque la viñeta que le interese.

Este modo de visualización puede resultar especialmente útil a usuarios que no vayan buscando una información concreta, sino que quieran ver el panorama global de un contrato para luego navegar de forma más específica según las opciones que se vayan presentando. No obstante, si el usuario personaliza la visualización mediante el menú de sus preferencias, puede lograr un resultado mucho más afinado, por ejemplo, puede ver solamente los documentos seleccionados como “esenciales” para auditoría.

Los destinatarios de este tipo de visualización pueden ser gestores, auditores e incluso el público en general si se produce la difusión web en abierto. Según el perfil del usuario, tendrá permiso para ver todos o determinados documentos.

- Aspectos visuales

En el diseño de pantalla predomina el blanco y otros colores suaves que aportan un aspecto limpio y resaltan la información que se quiere mostrar. No se usan motivos de decoración para no restar protagonismo a los elementos informativos.

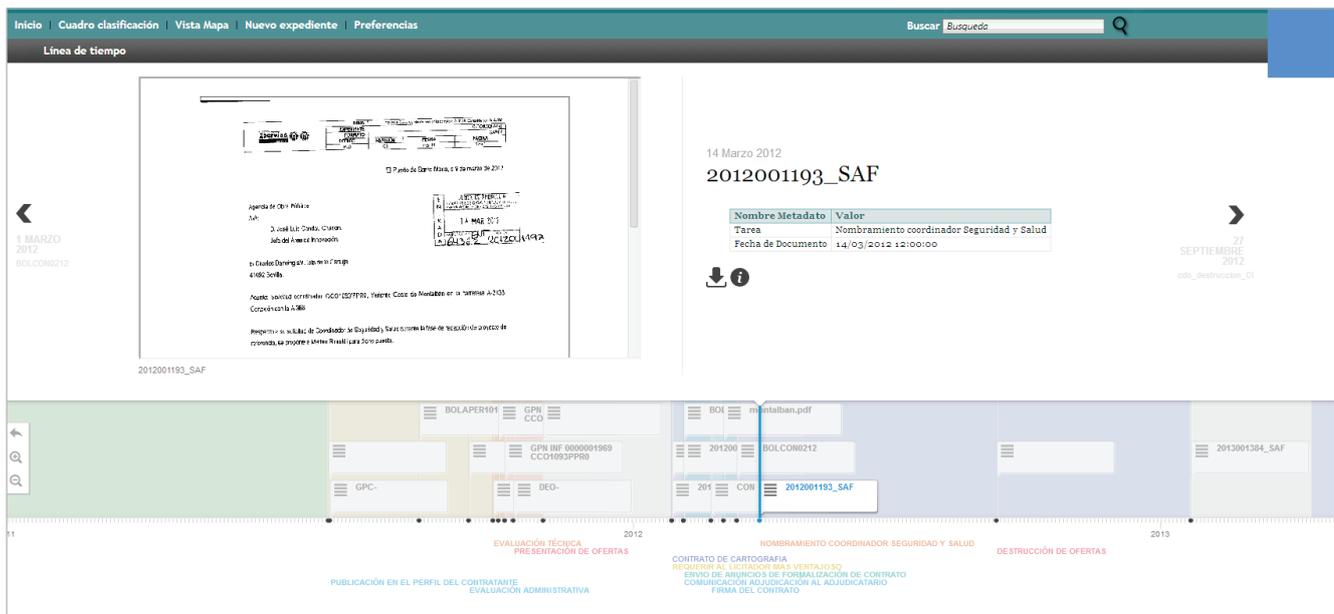


Figura 1. Visualización de un expediente en *timeline*

La interfaz está dividida en tres bandas horizontales: en la inferior se sitúa la línea de tiempo y el detalle de las tareas asociadas al cronograma, en la siguiente están las viñetas de los documentos sobre una franja de colores que representan la tarea que origina dichos documentos, y en la superior se visualiza la imagen del documento seleccionado en el *timeline* y un detalle del mismo.

La escala de tiempo se ajusta automáticamente en base a la información contenida en el expediente, o sea, al número de documentos, de forma que la visualización inicial del mismo resulte lo más compacta posible. Mediante una etiqueta flotante se accede al menú contextual, que permite cambiar a otro tipo de vistas del expediente: tabular, árbol, sólo documentos esenciales o todos, y establecer el tipo de fecha por la que se quieren visualizar los documentos en el *timeline*.

Cada icono sin leyenda tiene su descripción emergente; es el caso de los controles de la línea de tiempo –aumentar, disminuir zoom y volver al inicio de la línea del tiempo– y los iconos de “detalle del documento” y “descarga” referidos al documento que se ve en pantalla. Se han elegido letras claras y fuentes de una única familia; según la configuración del equipo será *trebuchet ms*, *arial*, *verdana* o *helvética*.

- Funciones

La función principal consiste en visualizar de forma gráfica y asociada a una línea temporal y a tareas de tramitación, los documentos de un expediente en una sola búsqueda, pudiendo el usuario desplazarse de dos formas: eligiendo pasar a ver el próximo/anterior documento mediante las flechas situadas a mitad de pantalla, o desplazando el cursor o rueda del ratón sobre la línea de tiempo. Las funciones del *timeline* son:

- aumentar/disminuir la escala;
- volver al inicio de la línea de tiempo;
- seleccionar un documento en el *timeline*.

Al seleccionar un documento, en el marco superior se muestra la imagen de la página primera y una ficha con metadatos mínimos descriptivos, pudiendo descargar el documento o acceder al registro descriptivo completo del documento.

“ La vista de *timeline* es útil a usuarios que no buscan una información concreta, sino el panorama global de un contrato para luego navegar de forma más específica ”

- Interacción

Para modificar la visualización, el usuario puede configurar sus preferencias a través del menú contextual que se le ofrece en una ventana flotante. Por ejemplo los metadatos que quiere ver en la ficha descriptiva del documento, el tipo de fecha de los documentos o si los quiere ver todos o una selección de los “esenciales”.

- Aplicaciones, componentes y servicios utilizados

Esta vista se ha creado con *TimelineJS*, un componente *JavaScript* para construir líneas de tiempo basadas en eventos asociándoles elementos multimedia. Permite integración directa con los servicios de *Twitter*, *Flickr*, *Google Maps*, *YouTube*, *Vimeo*, *Vine*, *Dailymotion*, *Wikipedia*, *SoundCloud*. Este componente ha sido modificado para facilitar funciones adicionales así como para realizar su integración en el sistema final.
<http://timeline.knightlab.com>

4.2. Clúster

- Propósito

Se pretende aproximar al usuario a las relaciones entre fondos, entre divisiones de fondos y entre series documenta-

les, una información que tradicionalmente ha estado restringida a la consulta de los estudios de identificación de series documentales, que por otra parte no suelen tener una difusión muy abierta. En cambio sí que contienen información de utilidad tanto para los profesionales como para los usuarios.

- Aspectos visuales

En el diagrama (figura 2) se visualiza la entidad seleccionada -en distinto color- y sus relaciones de asociación con otros nodos. El tamaño de los nodos es proporcional al número de expedientes que contiene cada serie documental y la distancia entre nodos revela la intensidad de la relación, o sea, el número de relaciones entre los nodos. En este caso las series documentales de expedientes de contratación de obras y de servicios son las más voluminosas y las que más estrecha relación tienen. Es totalmente lógico, ya que la redacción de un proyecto (contrato de servicio) está directamente vinculada a la realización de la obra (contrato de obra) y a la dirección de obra (contrato de servicio).

Los nodos presentan el rótulo con su nombre y están unidos por líneas de igual grosor. El marco de la parte inferior muestra los metadatos básicos del objeto seleccionado, que puede configurar el usuario en su pantalla de preferencias.

Podemos visualizar los documentos en el tiempo, en el espacio, en los hitos de la tramitación administrativa y en sus relaciones con otros documentos

- Funciones

Esta vista propicia una visión contextualizada de las agrupaciones de documentos superiores al nivel de expediente. Elijiendo una de estas agrupaciones -fondo, división de fondo o serie documental- se representa el diagrama de relaciones de asociación con otras entidades.

- Interacción

El usuario puede navegar a través de los nodos y la visualización va cambiando, cargándose las relaciones de asociación del nuevo nodo con otros.

- Aplicaciones, componentes y servicios utilizados

Las vistas de nube de puntos y clúster basan su funcionamiento en el componente *JSF mindmap* de *PrimeFace*, al que se le ha ampliado su función básica con el desarrollo de clúster.

4.3. Nube de puntos

- Propósito

El objetivo de esta vista es que sean evidentes para el usua-

rio las relaciones de "contiene" y de "asociación" entre expedientes, y las relaciones de "asociación" entre documentos. Un expediente según esta vista aparecerá enlazado a otros expedientes subordinados (relación de "contiene") y a otros expedientes relacionados (relación de "asociación"). Es una visualización sencilla y útil para los gestores, profesionales de la información y usuarios.

La vista de clúster muestra una visión contextualizada de las agrupaciones de documentos superiores al nivel de expediente

Se diferencia del clúster en que aquí la distancia entre nodos no tiene significado, mientras que el clúster aproxima los elementos según la intensidad de su relación.

- Aspectos visuales

El resultado es un diagrama de nodos unidos por líneas. En el centro está el objeto seleccionado y a su alrededor los demás objetos con los que se relaciona. El nodo principal figura en el centro en color destacado y los nodos relacionados en dos colores distintos, según estén unidos por relación de "asociación" o "contiene".

Los nodos tienen etiquetas con su nombre limitado a un número de caracteres, ya que los nombres de los expedientes o documentos pueden ser bastante largos. Al situar el cursor sobre un nodo, se visualiza una etiqueta emergente con el nombre completo. Las líneas de relación son iguales en todos los casos, tanto en grosor como en distancia. El marco de la parte inferior muestra los metadatos básicos del objeto seleccionado, que puede personalizar el usuario.

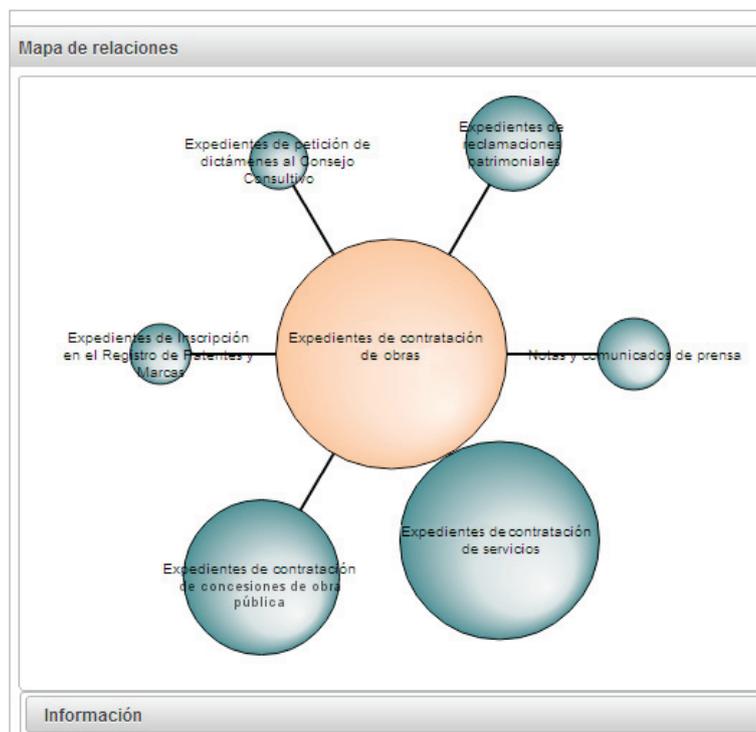


Figura 2. Visualización de series documentales relacionadas

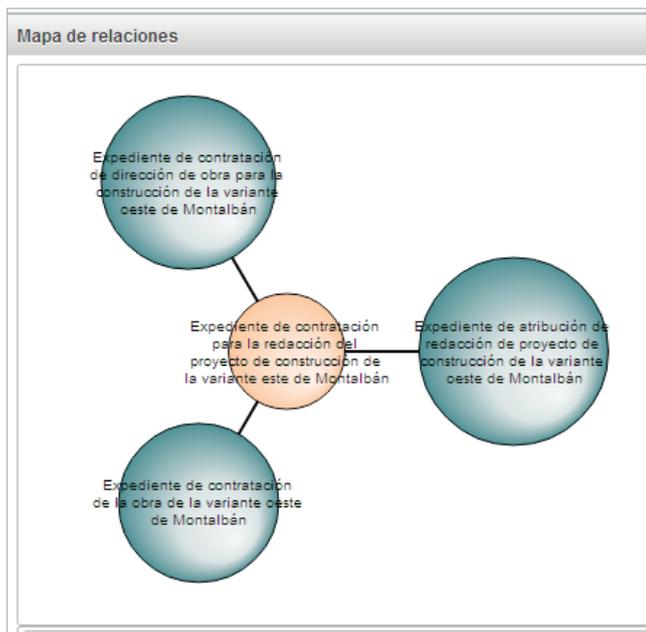


Figura 3. Visualización de la relación de un expediente con otros

- Funciones principales

Proporciona una visión contextualizada de documentos y expedientes, al mostrar sus relaciones de “asociación” o “contiene”.

- Interacción

El usuario puede modificar la visualización organizando los nodos para mejorar la composición del diagrama. En el caso del clúster, los nodos no se podían mover porque la distancia estaba fijada en función de la intensidad de la relación entre nodos. También se pueden seleccionar otros nodos y se cargan las relaciones de ese objeto con otros.

La vista de nube de puntos proporciona una visión contextualizada de documentos y expedientes

4.4. Vista mapa

- Propósito

Cuando el objeto de los contratos tiene un alto significado geográfico –como es el caso de los contratos de obras y servicios de infraestructuras de transporte- la visualización de los expedientes de contratación posicionados en un mapa resulta una forma muy intuitiva de ofrecer la información al usuario.

Esta vista es una opción que tiene el usuario en la pantalla de resultados, tras realizar una búsqueda. El propio mapa puede convertirse en una prolongación de la búsqueda si el usuario interacciona y llega a los documentos a través de los iconos que los señalan.

- Aspectos visuales

En la pantalla inicial se muestra la cartografía centrada en Andalucía porque es el territorio donde se sitúan los docu-

mentos tratados. Los expedientes se visualizan mediante puntos que, si están muy cercanos y la escala es grande, se agrupan en puntos mayores y proporcionales al número de documentos. A medida que se aumenta el zoom se van individualizando los puntos.

- Funciones

Esta vista proporciona de forma gráfica la ubicación de los expedientes en una posición geográfica y facilita al usuario el análisis de la información que visualiza, a la vez que proporciona asistencia si quiere seguir la búsqueda y su referente es el espacio.

- Interacción

El mapa permite realizar zoom más/menos. Al colocar el cursor sobre un punto emerge el nombre del expediente y al hacer clic se verán sus metadatos principales y se podrá acceder directamente al mismo. Estos metadatos son configurables por el usuario en su perfil de preferencias.

- Aplicaciones, componentes y servicios utilizados

La vista mapa basa su funcionamiento en los siguientes componentes:

- *OpenStreetMap*: servicio de cartografía bajo licencia abierta (con licencia *Creative Commons Reconocimiento-CompartirIgual 2.0 (CC BY-SA)*) y creado por las aportaciones de su comunidad de usuarios.
<http://www.openstreetmap.org>
- *Leaflet*: librería *JavaScript open source* que permite añadir capas de forma sencilla a una cartografía.
<http://leafletjs.com>
- *MarkerClusterer* plugin para *Leaflet*: librería *open source JavaScript* que permite la creación de clústeres de puntos de interés en una cartografía, de forma que los puntos de interés se agrupan o se separan en función del zoom aplicado
<https://github.com/Leaflet/Leaflet.markercluster>

La vista de mapa muestra la ubicación de los expedientes en una posición geográfica y facilita al usuario el análisis de la información que visualiza

4.5. Requisitos de uso

Los mecanismos de visualización descritos se encuadran en una aplicación web desarrollada con tecnología *Java* y basada en los siguientes *frameworks*:

- *JSF 2.0* con implementación *Mojarra 2.1.2* y librerías de componentes *Jboss RichFaces* y *PrimeFaces*.
<http://www.jboss.org/richfaces>
<http://www.primefaces.org>
- *Spring Framework 3.1.2* con *Spring Security* para la gestión de la seguridad.
- Estándar de servicios web *JAX-WS* con la implementación de *Apache CXF 2.7.7*.

Los requisitos mínimos que debe cumplir el servidor en el que se instale la aplicación son:

- Máquina virtual *Java 1.6* o superior.
- Servidor de aplicaciones *Jboss 7.1* con implementación de servicios *JBossWS-CXF-4.1.1.Final*.

A nivel de usuario de la aplicación, los requisitos que debe disponer para usar los mecanismos de visualización son:

- Navegador web *Internet Explorer 9* o superior, *Google Chrome* en su última versión o *Mozilla Firefox* en su última versión.
- Resolución de pantalla mínima de 1.024 x 768 y resolución recomendada de 1.920 x 1.080.

5. Conclusiones

Es evidente que los documentos de archivo son una fuente de información única. Pero ganarían mucho los usuarios si utilizamos todos los medios a nuestro alcance para facilitarles el contacto con los documentos. La orientación al usuario es la máxima que debe inspirar la labor de difusión de los documentos de archivo.

Este es el motivo que nos ha llevado a emplear técnicas de visualización para hacer más comprensivo y directo el conocimiento (Dürsteler, 2002; Costa, 1998) de un tipo de expedientes complejo por la cantidad de documentos que contienen y por la extensa tramitación en el tiempo: los expedientes de contratación de servicios y obras de infraestructuras de transporte. Estos expedientes se podrán consultar asociados a su localización geográfica o a través de las relaciones –ya sea de pertenencia o de asociación- con otros expedientes. Una vez dentro del expediente, se podrán visualizar los documentos que lo componen, asociados a una línea de tiempo o de forma tabular inscritos en la fase y tarea del procedimiento que le corresponda a cada uno.

Con el objetivo de optimizar estas visualizaciones y extender su aplicación a más conjuntos de documentos, las herramientas han sido parametrizadas para que puedan representar también otro tipo de expedientes, así como series documentales, divisiones de fondos y fondos. Todos estos elementos se pretenden incluir en un intento de ofrecer visualmente al usuario un volumen de información estructurada pero de difícil comprensión.

Las visualizaciones son consecuencia de la gestión de los metadatos de los documentos, cuyos valores contienen los datos que precisamos para producirlas. De la calidad de estos datos estructurados dependerá el éxito y la certeza de las visualizaciones que obtengamos.

La experiencia que se expone en este artículo no es original en su desarrollo, pero sí en su aplicación en un ámbito falto de estímulos para buscar documentos de forma visual. Se-

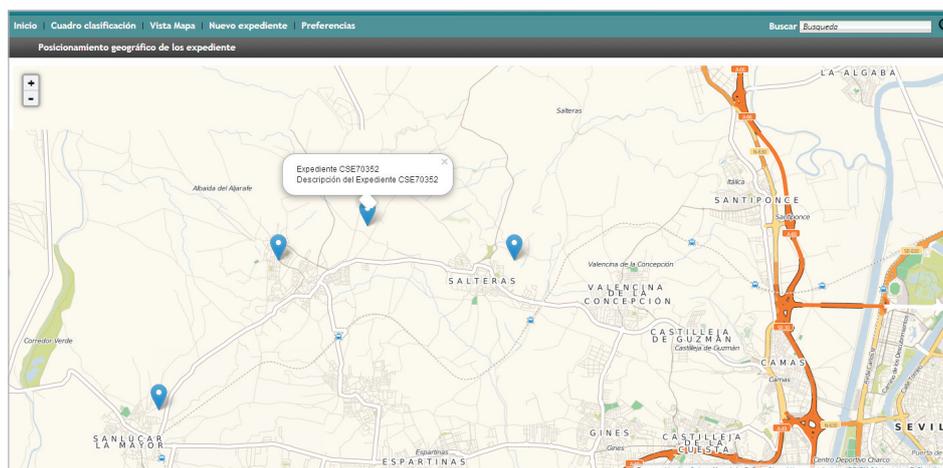


Figura 4. Visualización en mapa de expedientes de contratación

ría deseable que surgiesen más iniciativas en este sentido, aprovechando la oportunidad de tratamiento que nos ofrece la información altamente estructurada propia del entorno del documento electrónico.

6. Reconocimientos

Los autores agradecen la financiación de la Unión Europea al Proyecto de innovación de la gestión documental aplicada a expedientes de contratación de servicios y obras de infraestructuras de transporte a través del Programa Operativo Feder de Andalucía 2007-2013. También agradecen su labor a los miembros del equipo de investigación de Ingeniería Web y Testing Temprano de la Universidad de Sevilla, a la empresa *Tecnocom España Solutions*, así como a los directivos e investigadores de la Agencia de Obra Pública de la Junta de Andalucía, por su dedicación y profesionalidad.

Las técnicas de visualización hacen más comprensivo y directo el conocimiento de un expediente complejo

7. Notas

1. Uno de los 28 proyectos de I+D+i relacionados con los ámbitos y competencias de la *Consejería de Fomento y Vivienda*, que fueron licitados y contratados en 2012 por la *Agencia de Obra Pública* de la *Junta de Andalucía*, todos cofinanciados por la Unión Europea (*Programa Operativo Feder* de Andalucía 2007-2013). Este en concreto fue adjudicado al *Grupo de investigación de Ingeniería Web y Testing Temprano* de la *Universidad de Sevilla*, continúa en ejecución y está prevista su finalización para junio de 2014.

2. *Esquema de metadatos para la gestión del documento electrónico (e-Emgde)*. Documentación complementaria a la *Norma técnica de interoperabilidad de política de gestión de documentos electrónicos*. Ministerio de Hacienda y Administraciones Públicas.

8. Bibliografía

Bengochea, Luis; Patricio, Miguel-Ángel (2005). “Sistemas de visualización para bibliotecas digitales”. *Revista española*

de documentación científica, v. 28, n. 3, pp. 273-292.
<http://redc.revistas.csic.es/index.php/redc/article/view/170/224>
<http://dx.doi.org/10.3989/redc.2005.v28.i3.170>

Bosque-Sendra, Joaquín; Zamora-Ludovic, Hernán-Eduardo (2002). "Visualización geográfica y nuevas cartografías". *Geofocus: revista internacional de ciencia y tecnología de la información geográfica*, n. 2, pp. 61-77.
http://geofocus.rediris.es/docPDF/Articulo4_2002.pdf

Costa, Joan (1998). *Información visual y conocimiento. La esquemática. Visualizar la información*. Barcelona: Paidós. ISBN: 9788449306112

Chen, Chaomei (1999). *Information visualization and virtual environments*. London: Springer. ISBN: 1852331364

Dietrich, Daniel; Myrntinen, Henri (2012). *Data visualization*. European Public Sector Information Platform.
<http://www.epsiplatform.eu/sites/default/files/Final%20TR%20Data%20Visualisation.pdf>

Dürsteler, Juan-Carlos (2002). *Visualización de la*

información: proceso de interiorización del conocimiento mediante la percepción de información.
<http://www.infovis.net/printMag.php?num=100&lang=1>

Fernández-Molina, Juan-Carlos; Peis, Eduardo (1999). "Uso de esquemas de clasificación para mejorar las prestaciones de visualización (*browsing*) de los catálogos en línea". En: *Organización del conocimiento en sistemas de información y documentación: III Encuentro de ISKO-España*, Getafe, 19-21 nov. 1997, pp. 295-306.
<http://dialnet.unirioja.es/descarga/articulo/2036344.pdf>

Inf@Vis! La revista digital de InfoVis.net.
<http://www.infovis.net>

Marcos, Mari-Carmen (2004). *Interacción en interfaces de recuperación de información: conceptos, metáforas y visualización*. Gijón (Asturias): Trea. ISBN: 8497041186

Red.es (2013). *Recopilación de herramientas de procesamiento y visualización de datos*. Madrid
http://datos.gob.es/sites/default/files/Herramientas_de_Visualizaci%C3%B3n_vf.docx

Directorio EXIT (EXpertos en Tratamiento de la Información) <http://directorioexit.info>

El *Directorio EXIT* permite localizar colegas para participar en cursos, conferencias, proyectos o recibir ofertas de trabajo

¿Te apuntas?
Ya somos 2.900
profesionales de
la información

Para bibliotecarios, documentalistas, periodistas, comunicadores, informáticos y diseñadores que ya figuren al menos en *Google Scholar* o tengan experiencia ejerciendo en empresas o instituciones

The screenshot shows the profile of Nieves González-Fernández-Villavicencio on the Directorio EXIT website. The profile includes a photo, a curriculum vitae icon, and a list of social media links (Google Scholar, ResearchGate). The main information provided is: Exit ID: 299; Email: gonzalez-fernandez-villavicencio@unipablo.es; Institution: Universidad Pablo de Olavide; Area: Área de Bibliotecología y Documentación; Address: Ctra. de Utrera, km. 3; Postal code: 41033; City: Sevilla; Province: Sevilla; Country: España; Personal email: nievesvillavicencio@gmail.com; Institutional web: http://www.ujo.es; Personal web: http://www.nievesglez.com; Personal descriptor: Social Media Marketing; Specialties: Biblioteca universitaria, Educación, formación, affir, Marketing y promoción; Recuperación de información y búsquedas; Web 2.0 y redes sociales; Groups: Buscar en Google, iCard, Buscar en Google Académico, XOL, Buscar en Bing, ROP, and Buscar en Yahoo!.

Acabamos de estrenar nuevo
look y nuevas funciones

Una ventana a las oportunidades profesionales

junio-octubre 2014

13-15 de junio de 2014

12TH EASE CONF. The complex world of science editing
Split, Croacia

European Association of Science Editors

<http://www.ease.org.uk/ease-events/major-conferences>

16-20 de junio de 2014

LIBRARIES IN THE DIGITAL AGE (LIDA). Assessing libraries and library users and use

Zadar, Croacia

<http://ozk.unizd.hr/lida>

19-20 de junio de 2014

COMMUNICATION AND INFORMATION SCIENCES IN NETWORK SOCIETY: Experiences and insights

Vilnius

Vilnius University

http://www.conference.kf.vu.lt/?page_id=23&lang=en

19-20 de junio de 2014

18TH INTL CONF ON ELECTRONIC PUBLISHING (EIPub). Let's put data to use: digital scholarship for the next generation
Tsalónica

Alexander Technological Education Institute of Thessaloniki (ATEITH)

<http://elpub2014.teithe.gr>

19-20 de junio de 2014

3^a CONF ESPAÑOLA SOBRE RECUPERACIÓN DE INFORMACIÓN (CERI)

A Coruña

<http://ceri2014.udc.es>

23 de junio de 2014

DCMI REGIONAL MEETING. Making metadata work
Londres

http://www.iskouk.org/events/metadata_June_2014.htm

<http://dcevents.dublincore.org/BibData/mmw>

25-28 de junio de 2014

KNOWLEDGE MANAGEMENT CONF 2014

Blagoevgrad, Bulgaria

Intl Institute for Applied Knowledge Management (IIAKM)

<http://iiakm.org/conference>

1-4 de julio de 2014

FIRST EUROPEAN SOCIAL NETWORKS CONF

Barcelona

<http://www.eusn.org>

2-5 de julio de 2014

43RD LIBER. Research libraries in the 2020 information landscape
Riga, Letonia

Ligue des Bibliothèques Européennes de Recherche

<http://www.lnb.lv/en/liber2014>

10-11 de julio de 2014

EUROPEAN CONF ON SOCIAL MEDIA (ECSM 2014)

Brighton, UK

Academic Conferences & Publishing Intl

<http://academic-conferences.org/ecsm/ecsm2014/ecsm14-home.htm>

17-20 de julio de 2014

LIBEURO 2014. European conf on literature & librarianship

Brighton

Iafor (The International Academic Forum)

<http://iafor.org/iafor/conferences/the-european-conference-literature-and-librarianship-2014>

23-25 de julio de 2014

JIAI 2014. Jornadas internacionales sobre acceso a la información: Activismo ciudadano y ambientes virtuales

Quito

Centro Internacional de Estudios Superiores de Comunicación para América Latina (Ciespal)

jornadas@ciespal.net

23-25 de julio de 2014

ELECTRONIC THESES AND DISSERTATIONS 2014 (ETD2014)

Leicester, Reino Unido

University of Leicester Library

<http://www2.le.ac.uk/library/etd2014>

14-15 de agosto de 2014

WHAT FUTURE PUBLIC LIBRARIES? (IFLA satellite meeting)

Birmingham, Reino Unido

IFLA Public Libraries Section

<http://conference.ifla.org/ifla80/satellite-meetings>

12-14 de agosto de 2014

ART LIBRARIES MEET THE CHALLENGES OF E-PUBLISHING: new formats, new players, new solutions (IFLA satellite meeting)

París

IFLA Art Libraries Section

<http://conference.ifla.org/ifla80/satellite-meetings>

13 de agosto de 2014

RDA-RESOURCE DESCRIPTION AND ACCESS (IFLA satellite meeting)
Frankfurt

IFLA Cataloguing Section

<http://conference.ifla.org/ifla80/satellite-meetings>

13-14 de agosto de 2014

SPACE AND COLLECTIONS EARNING THEIR KEEP. Transformation, technologies and re-tooling (IFLA satellite meeting)

París

IFLA Library Buildings & Equipment Section and IFLA Acquisition & Collection Development Section

<http://iflasection20.org/space-and-collections-earning-their-keep-transformation-technologies-and-re-tooling>

13-14 de agosto de 2014

PARLIAMENTARY LIBRARIES: PAST AND FUTURE (IFLA satellite meeting)

París

IFLA Library & Research Services to Parliaments Section

<http://conference.ifla.org/ifla80/satellite-meetings>

13-14 de agosto de 2014

COPYRIGHT AND BEYOND. Libraries in the public sphere (IFLA satellite meeting)

Estrasburgo

IFLA Committee on Copyright and other Legal Matters (CLM) Strategic Programme with European Bureau of Library Information and Documentation Associations (Eblida)

<http://conference.ifla.org/ifla80/satellite-meetings>

13-14 de agosto de 2014

THEORY AND RESEARCH ON THE CONVERGENCE OF PROFESSIONAL IDENTITY AND TECHNOLOGY IN CULTURAL HERITAGE INSTITUTIONS (Libraries, museums, and archives) (IFLA satellite meeting)

Turín

IFLA Library Theory and Research Section
<http://conference.ifla.org/ifla80/satellite-meetings>

13-14 de agosto de 2014

CULTURAL HERITAGE IN THE DIGITAL AGE (IFLA satellite meeting)
Ginebra
IFLA Preservation and Conservation Section
<http://conference.ifla.org/ifla80/calls-for-papers/cultural-heritage-digital-age>

13-14 de agosto de 2014

RESTRUCTURING RESOURCE SHARING. New organizations, technologies, methods (IFLA satellite meeting)
Nancy, France
IFLA Document Delivery and Resource Sharing Section
<http://www.ifla.org/node/8055>

14 de agosto de 2014

LINKED DATA IN LIBRARIES. Let's make it happen! (IFLA satellite meeting)
París
IFLA Information Technology Section with the Semantic Web Special Interest Group
<http://conference.ifla.org/ifla80/satellite-meetings>

14-15 de agosto de 2014

ETHICAL DILEMMAS IN THE INFORMATION SOCIETY. How codes of ethics help to find ethically based solutions (IFLA satellite meeting)
Ginebra
IFLA (Intl Federation of Library Associations and Institutions) FAIFE (IFLA Committee on Freedom of Access to Information and Freedom of Expression)
<http://conference.ifla.org/ifla80/satellite-meetings>

14-15 de agosto de 2014

FACING THE FUTURE. Librarians and information literacy in a changing landscape (IFLA satellite meeting)
Limerick, Irlanda
IFLA Information Literacy Section
<http://www.iflasatellitelimerick.com>

15 de agosto de 2014

LIBRARIES AS MEDIA. How libraries can contribute to explain the world by producing information and organize themselves as medias? (IFLA satellite meeting)
París
IFLA Management & Marketing Section
<http://conference.ifla.org/ifla80/satellite-meetings>

15 de agosto de 2014

KNOWLEDGE MANAGEMENT AS A VITAL TOOL FOR CHANGE MANAGEMENT (IFLA satellite meeting)
Lyon
IFLA Knowledge Management Section
<http://conference.ifla.org/ifla80/satellite-meetings>

16-22 de agosto de 2014

IFLA WORLD LIBRARY AND INFORMATION CONGRESS. Libraries, citizens, societies: confluence for knowledge
Lyon
Intl Federation of Library Associations and Institutions
<http://conference.ifla.org/ifla80>

22-23 de agosto de 2014

EBOOKS FOR EVERYONE! AN OPPORTUNITY FOR MORE INCLUSIVE LIBRARIES (IFLA satellite meeting)
París
IFLA section Libraries Serving Persons with Print Disabilities (LPD)
http://inova.snv.jussieu.fr/ifla-lpd/colloques/colloques/83_index_en.html

23 de agosto de 2014

15 TO 20 YEAR OLDS READING ACROSS THE WORLD. Why? Where? How? (IFLA satellite meeting)

París

IFLA Libraries for Children & Young Adults Section
<http://conference.ifla.org/ifla80/satellite-meetings>

23-25 de agosto de 2014

FRANCOPHONIES, BIBLIOTHÈQUES ET CONFLUENCES. Managing and marketing libraries in a world where digitalisation and new technologies are changing boundaries (IFLA satellite meeting)
Limoges
IFLA Management & Marketing Section
<http://conference.ifla.org/ifla80/satellite-meetings>

23-25 de agosto de 2014

3RD INTL ASSOCIATION OF FRENCH-SPEAKING LIBRARIANS AND DOCUMENTALISTS (AIFBD) CONGRESS
Limoges
Association Internationale Francophone des Bibliothécaires et Documentalistes (AIFBD)
http://www.aifbd.org/images/Congres2014/Appel_AIFBD_2014.pdf

25-26 de agosto de 2014

HISTORY OF LIBRARIANSHIP (IFLA satellite meeting)
Lyon
IFLA Rare Books & Manuscripts Section
<http://conference.ifla.org/ifla80/satellite-meetings>

1-4 de septiembre de 2014

25TH ACM CONF ON HYPERTEXT AND SOCIAL MEDIA (ACM Hypertext 2014)
Santiago, Chile
<http://ht.acm.org/ht2014>

2-5 de septiembre de 2014

ISIC (INFORMATION SEEKING IN CONTEXT). INFORMATION BEHAVIOUR CONF
Leeds, Inglaterra
Univ. of Leeds, Univ. of Aberystwyth y Univ. of Sheffield
<http://isic2014.com>

3-5 de septiembre de 2014

10TH INTL CONF ON WEBOMETRICS, INFORMETRICS AND SCIENTOMETRICS (WIS) & 15TH COLLNET MEETING
Ilmenau, Alemania
<http://www.tu-ilmenau.de/collnet2014>

3-5 de septiembre de 2014

INTL SYMPOSIUM ON PHILOSOPHY OF LIBRARY AND INFORMATION SCIENCE. Ethics: theory and practice
Kastamonu, Turquía
Kastamonu Univ, Dept of Information and Records Management
<http://www.kbbf2014.org>

4-5 de septiembre de 2014

15TH EUROPEAN CONF ON KNOWLEDGE MANAGEMENT (ECKM 2014)
Santarém, Portugal
Academic Conferences & Publishing Intl
<http://academic-conferences.org/eckm/eckm2014/eckm14-home.htm>

5-7 de septiembre de 2014

SOCIETY OF INDEXERS ANNUAL CONF
Cirencester, Inglaterra
Society of Indexers
<http://www.indexers.org.uk/index.php?id=578>

5-8 de septiembre de 2014

2ND WORKSHOP ARCHIVES IN THE DIGITAL AGE: STANDARDS, POLICIES AND TOOLS. Part of the 4th Intl conf on integrated information (IC-Ininfo 2014)
Madrid
<http://icinfo.net>

8-10 de septiembre de 2014

STM 13TH MASTER CLASS - EUROPE. Developing leadership and innovation
Oxford

Intl Association of Scientific, Technical & Medical Publishers (STM)
<http://www.stm-assoc.org/events/13th-master-class-europe-developing-leadership-and-innovation>

8-12 de septiembre de 2014

JOINT CONF ON DIGITAL LIBRARIES (JCDL). Theory & practice of
digital libraries (TPDL) 2014

Londres
JCDL Steering Committee
<http://www.jcdl.org/index.php>

9-11 de septiembre de 2014

X COLÓQUIO HABERMAS
I COLÓQUIO DE FILOSOFIA DA INFORMAÇÃO
Rio de Janeiro

Unirio; Ppgci UFRJ; Ibcit; Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (Ancib)
<http://coloquiohabermas.wordpress.com>

12 de septiembre de 2014

2ND WORKSHOP ON LINKING AND CONTEXTUALIZING PUBLICATIONS AND DATASETS. Growing a global data publishing culture
Londres

City University of London
<http://lcpd2014.research-infrastructures.eu>

10-12 de septiembre de 2014

ALPSP INTL CONF 2014

Londres
Association of Learned and Professional Society Publishers (ALPSP)
<http://www.alpsp.org/Ebusiness/TrainingAndEvents/ALPSPInternationalConference.aspx>

10-13 de septiembre de 2014

34^º CONGRESO INTERNACIONAL DEL IBBY
México DF

The International Board on Books for Young People
<http://www.ibbycongress2014.org>

17-19 de septiembre de 2014

6TH CONF ON OPEN ACCESS SCHOLARLY PUBLISHING (OASPA)

Paris
info@oaspa.org
<http://oaspa.org>

17-21 de septiembre de 2014

SHARP'S 22ND ANNUAL CONF

Amberes
Society for the History of Authorship, Reading & Publishing (SHARP)
<http://www.sharp2014.be>

24-26 de septiembre de 2014

VI ENCUENTRO LATINOAMERICANO DE BIBLIOTECARIOS, ARCHIVISTAS Y MUSEÓLOGOS (VI EBAM)

Medellín, Colombia
<http://www.ebam.com.ar>
<https://www.facebook.com/pages/VI-EBAM/256266957856994>

24-26 de septiembre de 2014

5TH GLOBAL CONF: DIGITAL MEMORIES

Oxford
<http://www.inter-disciplinary.net/critical-issues/cyber/digital-memories/call-for-papers>

22-25 de septiembre de 2014

8TH INTL CONF ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS (FOIS 2014)

Rio de Janeiro
<http://fois2014.inf.ufes.br/p/home.html>

1-3 de octubre de 2014

XIX IBERSID. Encuentros internacionales sobre sistemas de información y documentación

Zaragoza
Francisco-Javier García-Marco, Universidad de Zaragoza
<http://www.ibernid.org>

1-3 de octubre de 2014

X ENCUENTRO DE DIRECTORES Y IX DE DOCENTES DE ESCUELAS DE BIBLIOTECOLOGÍA Y CIENCIA DE LA INFORMACIÓN DEL MERCOSUR. La cooperación y el compromiso en la formación profesional

Buenos Aires
Biblioteca Nacional Mariano Moreno
encuentrodd2014@gmail.com

2-3 de octubre de 2014

XIII WORKSHOP REBIUN SOBRE PROYECTOS DIGITALES. Libros electrónicos en las bibliotecas

Zamora
III Plan Estratégico Rebiun 2020
<http://www3.ubu.es/proyectosdigitales>

6-8 de octubre de 2014

5^ª CONFERÊNCIA LUSO-BRASILEIRA DE ACESSO ABERTO
Coimbra

Universidade de Coimbra
<http://www.acesolivre.pt/c/index.php/confoa2014/2014>

7 de octubre de 2014

STM FRANKFURT CONF 2014

Frankfurt
Intl Association of Scientific, Technical, and Medical Publishers (STM)
<http://www.stm-assoc.org/events/frankfurt-conference-2014>

7-9 de octubre de 2014

XIII CONGRESO NACIONAL DE BIBLIOTECOLOGÍA Y CIENCIA DE LA INFORMACIÓN

XX JORNADA NACIONAL y V INTERNACIONAL DE ACTUALIZACIÓN Y CAPACITACIÓN DE BIBLIOTECAS MÉDICAS. Bibliotecas: espacios de aprendizaje interdisciplinario
Bogotá
<http://www.unirecscolombia.org/jornada2014>

8-9 de octubre de 2014

4TH INTL CONF ON INFORMATION COMMUNICATION AND MANAGEMENT (ICICM 2014)

Ginebra
Intl Assoc of Computer Science and Information Technology (Iacsit)
<http://www.icicm.org>

8-12 de octubre de 2014

FRANKFURT BOOK-FAIR / BUCHMESSE

Frankfurt
Gabi Rauch-Kneer Exhibition Management
<http://www.book-fair.com/en>

12-14 de octubre de 2014

26TH INTL CONF ON TRENDS FOR SCIENTIFIC INFORMATION PROFESSIONALS (ICIC 2014)

Heidelberg
Dr. Haxel Congress and Events
<http://www.haxel.com/icic/2014/meeting>

13-15 de octubre de 2014

2^ª CONF ANUAL DEL CONSEJO INTERNACIONAL DE ARCHIVOS
9^ª CONF EUROPEA DE ARCHIVOS. Archivos e industria cultural
13^{ES} JORNADES IMATGE I RECERCA

Girona, España
Joan Boadas-Raset; Lluís-Esteve Casellas, Archivo Municipal de Girona; Eurbica; Ajuntament de Girona.
<http://www.girona.cat/sgdap>

PUBLICACIONES EL PROFESIONAL DE LA INFORMACIÓN

REVISTA EL PROFESIONAL DE LA INFORMACIÓN

<input type="checkbox"/> Suscripción anual institucional	212 €
<input type="checkbox"/> Suscripción anual individual	98 €
<input type="checkbox"/> Suscripción anual institucional sólo online.....	123 €
<input type="checkbox"/> Suscripción anual individual sólo online.....	75 €

2014 v. 23

<input type="checkbox"/> n. 1: Documentación audiovisual	37 €
<input type="checkbox"/> n. 2: Políticas de información	37 €
<input type="checkbox"/> n. 3: Visualización de información	37 €

2013 v. 22

<input type="checkbox"/> n. 6: Formación y aprendizaje	37 €
<input type="checkbox"/> n. 5: Gestión de contenidos	37 €
<input type="checkbox"/> n. 4: Economía de la información	37 €
<input type="checkbox"/> n. 3: Bibliotecas y documentación de museos	37 €
<input type="checkbox"/> n. 2: Educación y biblioteca	37 €
<input type="checkbox"/> n. 1: Soportes digitales.....	37 €

2012 v. 21

<input type="checkbox"/> n. 6: Bibliotecas universitarias	30 €
<input type="checkbox"/> n. 5: Información y derecho	30 €
<input type="checkbox"/> n. 4: Comunicación digital	30 €
<input type="checkbox"/> n. 3: Organización del conocimiento	30 €
<input type="checkbox"/> n. 2: Publicación científica y acceso abierto	30 €
<input type="checkbox"/> n. 1: El futuro de la web	30 €

2011 v. 20

<input type="checkbox"/> n. 6: Fuentes de información 2.0	25 €
<input type="checkbox"/> n. 5: Inteligencia competitiva	25 €
<input type="checkbox"/> n. 4: Fotografía y bancos de imágenes	25 €
<input type="checkbox"/> n. 3: Información de las administraciones públicas	25 €
<input type="checkbox"/> n. 2: Archivos administrativos e intranets	25 €
<input type="checkbox"/> n. 1: Psicología y sociología de la información.....	25 €

Formulario de compra EPI

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



En todos los casos, excepto en los Anuarios, hay que añadir los gastos de envío

LIBROS EL PROFESIONAL DE LA INFORMACIÓN

(Editorial UOC)

<input type="checkbox"/> 1. Tecnologías de la web semántica	12,00 €
Juan-Antonio Pastor	
<input type="checkbox"/> 2. La revolución del libro electrónico	11,50 €
José-Antonio Cordón	
<input type="checkbox"/> 3. Sistemas de información en la empresa	11,50 €
Josep Cobarsi-Morales	
<input type="checkbox"/> 4. Información en el móvil	12,00 €
Natalia Arroyo-Vázquez	
<input type="checkbox"/> 5. Acceso abierto a la ciencia	12,00 €
Ernest Abadal	
<input type="checkbox"/> 6. Fuentes de información médica	12,00 €
Pablo Medina-Aguerebere	
<input type="checkbox"/> 7. Gestión de la reputación online	11,50 €
Javier Leiva-Aguilera	
<input type="checkbox"/> 8. <i>Wikipedia</i> de la A a la W	12,50 €
Tomás Saorín	
<input type="checkbox"/> 9. Etiquetar en la web social	12,00 €
Raquel Gómez-Díaz	
<input type="checkbox"/> 10. Mejorar las búsquedas de información	12,00 €
Silvia Argudo y Amadeu Pons	
<input type="checkbox"/> 11. Clubes de lectura	11,50 €
Óscar Carreño	
<input type="checkbox"/> 12. Plan social media y community manager	13,50 €
Julián Marquina-Arenas	
<input type="checkbox"/> 13. Documentación audiovisual en televisión	12,00 €
Jorge Caldera y Pilar Arranz	
<input type="checkbox"/> 14. Gestión de documentos en la e-administración	12,00 €
Elisa García-Morales	
<input type="checkbox"/> 15. <i>El film researcher</i>	12,00 €
Iris López-de-Solis	
<input type="checkbox"/> 16. Preservación digital	12,00 €
Miquel Térmens	
<input type="checkbox"/> 17. Gestión de contenidos	13,00 €
Ricardo Eito-Brun	
<input type="checkbox"/> 18. Documentación fotográfica	14,00 €
Juan-Miguel Sánchez-Vigil y Antonia Salvador-Benítez	
<input type="checkbox"/> 19. Documentación cinematográfica	12,00 €
Elena De la Cuadra	
<input type="checkbox"/> 20. Archivos	14,00 €
Ramón Alberch-Figueras	
<input type="checkbox"/> 21. Inteligencia en redes sociales	11,00 €
Eva Moya	
<input type="checkbox"/> 22. Bibliotecas escolares	10,00 €
Concepción Mª Jiménez-Fernández y Raúl Cremades-García	
<input type="checkbox"/> 23. Marca y comunicación empresarial	11,00 €
Pablo Medina-Aguerebere	
<input type="checkbox"/> 24. El content curator	14,00 €
Javier Guallar y Javier Leiva-Aguilera	
<input type="checkbox"/> 25. Gestión de la calidad en la biblioteca	13,50 €
Nuria Balagué y Jarmo Saarti	

Próxima aparición:

<input type="checkbox"/> 26. Innovación en bibliotecas.	Isabel Riaza
<input type="checkbox"/> 27. La intranet social.	Ana Carrillo Pozas

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>



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



Deseo recibir todos los números de 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 American Express

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 relleno 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 300 €

Precios 2014

Suscripción normal: 203,94 € + IVA = 212 €

Suscripción personal: (sólo a domicilios particulares) 94,23 € + IVA = 96 €

Gastos de envío fuera de España: Europa: 40 € Américas: 60 €

Suscripción sólo online: 101,65 € + IVA = 123 €

Número suelto: 35,57 € + IVA = 37 €

Suscripción sólo online personal: 62 € + IVA = 75 €

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 (año). "Título del artículo". *Título de la publicación periódica*, mes, v., n., 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 (año). "Título de ponencia".

En: *Nombre del congreso*, 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 (año). *Título del trabajo*.

Lugar de publicación: editor. 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 (año). "Título del capítulo".

En: *Apellido, Nombre; Apellido2, Nombre2*. *Título de la monografía*. Lugar

de publicación: editor, 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*.

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>