

OCTUBRE/DICIEMBRE,
2021, VOL. 35, NUM. 89,
MÉXICO, ISSN 2448-8321

Investigación Bibliotecológica

ARCHIVONOMIA, BIBLIOTECOLOGIA E INFORMACIÓN



Investigación Bibliotecológica

ARCHIVONOMÍA, BIBLIOTECOLOGÍA E INFORMACIÓN

Vol. 35, núm. 89, octubre/diciembre, 2021, México, ISSN: 2448-8321

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89>



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Instituto de Investigaciones Bibliotecológicas y de la Información

Contenido

INVESTIGACIÓN BIBLIOTECOLÓGICA, VOL. 35, NÚM. 89, OCTUBRE/DICIEMBRE, 2021, MÉXICO, ISSN: 2448-8321
DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89>

ARTÍCULOS

- **Algoritmo para el análisis temático de documentos digitales** 13-31
[Algorithm for thematic analysis of digital documents]
Luis Roberto Polo Bautista y Karen Vanessa Martínez Acevedo
- **Servicios accesibles a todos los usuarios en las bibliotecas universitarias españolas: estado de la cuestión** 33-50
[Accessible services to all users in Spanish University Libraries: state of the art]
David Domingo Pàmies, Carina Rey Martín y Concepción Rodríguez Parada
- **Evolution of the scholar community in the area of informetrics in Mexico: 1971-2018** 51-78
[Evolución de la comunidad académica en el área de informetría en México: 1971-2018]
María Elena Luna-Morales, Miguel Ángel Pérez-Angón and Evelia Luna-Morales
- **Lectura y conocimiento en Bibliotecología** 79-90
[Reading and knowledge in Library Science]
Héctor Guillermo Alfaro López
- **Repercusión altométrica y tipo de acceso en artículos peruanos de Ciencias Sociales** 91-110
[Altmetric attention and access type in Peruvian articles of Social Sciences]
Joel Alhuay-Quispe y Lourdes Bautista-Ynofuente
- **SIGI, an integrated scientometric and curricular information system for teaching-research institutions** 111-131
[SIGI, un sistema integral de información cienciométrica y curricular para instituciones de investigación-enseñanza]
María Magdalena Sierra Flores, Romel Calero Ramos, María Victoria Guzmán Sánchez, Juan Claudio Toledo-Roy and Yolsy Gabriela Gamboa Calderón
- **¿Son las investigadoras de Documentación más interdisciplinarias frente a sus colegas hombres? Un análisis aplicado al profesorado español** 133-149
[Are Information Science female researchers more interdisciplinary than their male colleagues? An analysis of Spanish faculty]
Manuel Blázquez Ochando, Michela Montesi e Isabel Villaseñor Rodríguez

- **Interconexões temáticas em Ciência da Informação no ResearchGate: o caso de professores da Pós-graduação no Brasil** [Thematic interconnections in Information Science at ResearchGate: the case of postgraduate professors in Brazil]
Bruna Lessa e Hildenise Ferreira Novo

151-167

- **Evaluación UX para sitios web orientados a bibliotecas digitales con usuarios ciegos** [UX Evaluation on digital libraries-oriented websites with blind people]
Teresita Álvarez-Robles, Yadira Alejandra Orozco Osuna y Francisco Javier Álvarez Rodríguez

169-194

- **Olhando a dimensão epistemológica da organização do conhecimento brasileira com três lentes da análise de domínio: epistemológica, estruturação científica e terminológica** [Looking at the epistemological dimension of Brazilian knowledge organization with three lenses of domain analysis: epistemological, scientific structuring and terminology]
Maira Fernandes Alencar e Brígida Maria Nogueira Cervantes

195-211

Investigación Bibliotecológica: archivonomía, bibliotecología e información [en línea] / ed. por el Centro Universitario de Investigaciones Bibliotecológicas. -Vol. 1, No. 1 (ago. 1986) - . Ciudad de México: Universidad Nacional Autónoma de México, Centro Universitario de Investigaciones Bibliotecológicas, 1986 - V. Semestral, 1986 – 2007; a partir de vol. 22, no. 44 (ene. / abr. 2008) - , la periodicidad cambió a cuatrimestral. A partir de enero de 2018, la periodicidad cambió a trimestral. Resúmenes en español e inglés, a partir del vol. 3, no. 1 (jul / dic. 1987) Disponible también en idioma inglés a partir del vol. 28, no. 62 (ene. / abr. 2014) Publicado por la misma dependencia bajo su nombre actual: Universidad Nacional Autónoma de México, Instituto de Investigaciones Bibliotecológicas y de la Información vol. 26, no. 56 (ene. / abr. 2012) - . Disponible para su consulta en línea a partir del vol. 1, no. 1 (ago. 1986) - . Publicado en formato electrónico a partir del vol. 30, no. 70 (sep. / dic. 2016) - . Todos los artículos cuentan con DOI en forma individual. Disponible en: <http://rev-ib.unam.mx/ib> ISSN 0187-358X (impreso) ISSN 2448-8321 (en línea)



Investigación Bibliotecológica: archivonomía, bibliotecología e información, Vol. 35, No. 89, octubre-diciembre 2021, es una publicación trimestral editada por la Universidad Nacional Autónoma de México, Ciudad Universitaria, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, a través del Instituto de Investigaciones Bibliotecológicas y de la Información, Circuito Escolar s/n, Torre II de Humanidades, Piso 12, Ciudad Universitaria, Col. Copilco, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, Tel. 56230349 y 56230337, <http://rev-ib.unam.mx/ib>, correos electrónicos: revista@iibi.unam.mx, drevista@iibi.unam.mx. Editor responsable: Dr. Egbert John Sánchez Vanderkast, Reserva de Derechos al uso Exclusivo No. 04-2016-041813344600-203, ISSN: 2448-8321, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número, Dra. María de Jesús Madera Jaramillo, Instituto de Investigaciones Bibliotecológicas y de la Información, Circuito Escolar s/n, Torre II de Humanidades, Piso 12, Ciudad Universitaria, Col. Copilco, Alcaldía Coyoacán, Ciudad de México, C.P. 04510, Tel. 56230337, fecha de la última modificación, octubre 2021.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor de la publicación. Se autoriza la reproducción total o parcial de los textos aquí publicados siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación.

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89>

REVISTA INDIZADA EN:

- | | |
|---|-----------------------------|
| •Clarivate Analytics
Web of Science | •Latindex |
| •Clarivate Analytics JCR | •DOAJ |
| •Scopus | •Elsevier
Science Direct |
| •SCImago Journal
& Rank | •LISA |
| •Índice de Revistas Mexicanas de Investigación Científica y Tecnológica del CONACyT | •LISTA Full Text |
| •SciELO Citation Index | •LISS |
| •SciELO | •INFOBILA |
| •Google Academics | •CSIC e-revist@s |
| | •Dialnet |
| | •CLASE |
| | •HAPI |

Esta revista está disponible en texto completo y en acceso abierto en:

- Revista IIBI: <http://rev-ib.unam.mx/ib>
- SciELO: http://www.scielo.org.mx/scielo.php?script=sci_serial&pid=0187-358X&lng=es&nrm=iso
- Science Direct: <https://www.sciencedirect.com/journal/investigacion-bibliotecologica-archivonomia-bibliotecologia-e-informacion>

DIRECTOR DE LA REVISTA

DR. EGBERT JOHN SÁNCHEZ VANDERKAST

CONSEJO EDITORIAL

DRA. ESTELA MORALES CAMPOS
Universidad Nacional Autónoma de México
Ciudad de México, México

DR. CARLOS ADOLFO PIÑA GARCÍA
Universidad Veracruzana
México

PhD BIRGER HJØRLAND
University of Copenhagen
Copenhagen, Dinamarca

DR. JOSÉ LÓPEZ YEPES
Universidad Complutense de Madrid
Madrid, España

DRA. GLORIA PÉREZ SALMERÓN
Stichting IFLA Foundation
La Haya, Holanda

DRA. FERNANDA RIBEIRO
University of Porto
Porto, Portugal

DR. ELÍAS SANZ CASADO
Universidad Carlos III
Madrid, España

COLABORADORES EN ESTE NÚMERO

INVESTIGACIÓN BIBLIOTECOLÓGICA, vol. 35, núm. 89, octubre/diciembre, 2021, México,
ISSN: 2448-8321

Brígida Maria Nogueira Cervantes

Centro de Educação, Comunicação e Artes
(CECA), Universidade Estadual de Londrina
Campus Universitário Rodovia Celso Garcia
Cid (PR 445), KM380 Caixa Postal: 10011
Londrina – Paraná CEP 86057-970, Brasil
Tel. +55 4399945 2910
brigidacervantes@gmail.com

Bruna Lessa

Departamento de Documentação e Informação,
Instituto de Ciência da Informação, Universida-
de Federal da Bahia
Avenida Reitor Miguel Calmon s/n parte supe-
rior Campus do Canela 40.110- 100, Brasil
Tel. +55 713283 7759
brunalesa@ufba.br

Carina Rey Martín

Departament de Biblioteconomia, Documen-
tació i Comunicació Audiovisual, Facultat de
Información y Medios Audiovisuales, Universitat
de Barcelona
C/ Melcior de Palau, 140, 08032 Barcelona
España
Tel. +34 60614 4742
carina.rey@ub.edu

Concepción Rodríguez Parada

Departament de Biblioteconomia, Documen-
tació i Comunicació Audiovisual, Facultat de
Información y Medios Audiovisuales, Universitat
de Barcelona
C/ Melcior de Palau, 140, 08032 Barcelona
España
Tel. +34 63657 9246
crodriguezp@ub.edu

David Domingo Pàmies

Centre de Recursos per a l'Aprentatge i la
Investigació, Campus Catalunya, Universitat
Rovira i Virgili
Av. Catalunya, 35 edif. C. 43002 Tarragona
España
Tels. +34 97755 9524, +34 65499 7794
david.domingop@urv.cat

Evelia Luna-Morales

Coordinación General de Servicios
Bibliográficos, Centro de Investigación y de
Estudios Avanzados del IPN (Cinvestav-IPN)
Av. Instituto Politécnico Nacional #2508,
Colonia San Pedro Zacatenco, CDMX
México
Tel. +52 555303 8010
eveorama@gmail.com

Francisco Javier Álvarez Rodríguez

Departamento de Ciencias Básicas,
Universidad Autónoma de Aguascalientes
Paseo de los Álamos #42, Fraccionamiento
Pulgas Pandas, Aguascalientes, Aguascalientes
México
Tel. +52 1449105 0925
fjalvar@correo.uaa.mx

Héctor Guillermo Alfaro López

Instituto de Investigaciones Bibliotecológicas
y de la Información, Universidad Nacional
Autónoma de México
Circuito Escolar s/n, Torre II de Humanidades,
piso 11, Ciudad Universitaria, Col. Copilco,
Alcaldía Coyoacán, CDMX, C.P. 04510
México
Tel. +52 555623 0358
galvaro@unam.mx

Hildenise Ferreira Novo

Departamento de Documentação e
Informação, Instituto de Ciência da Informação,
Universidade Federal da Bahia
Avenida Reitor Miguel Calmon s/n parte
superior Campus do Canela 40.110- 100
Brasil
Tel. +55 713283 7759
hnovo@ufba.br

Isabel Villaseñor Rodríguez

Facultad de Ciencias de la Documentación,
Universidad Complutense de Madrid
Calle de Santísima Trinidad 37, 28010 Madrid
España
Tel. +34 91394 6675
isavilla@ucm.es

Joel Alhuay-Quispe

Facultad de Ingeniería, Universidad Privada San
Juan Bautista
Calle Andrómeda Mz. Q Lote 7B, APV. Alejan-
dro Álvarez – Lima 03. Código Postal: Lima
15498, Perú
Tel. +51 94712 3856
joel.alhuay@upsjb.edu.pe

Juan Claudio Toledo-Roy

Instituto de Ciencias Nucleares, Universidad
Nacional Autónoma de México
Circuito Exterior s/n, Ciudad Universitaria, Col.
Universidad Nacional Autónoma de México,
Deleg. Coyoacán. C.P. 04510, CDMX
México
Tel. +52 555622 4660 Ext. 3312
juan.toledo@nucleares.unam.mx

Karen Vanessa Martínez Acevedo

Escuela Nacional de Biblioteconomía y
Archivonomía
Calz. Ticomán 645, La Laguna Ticoman,
Gustavo A. Madero, 07340, CDMX, México
Tel. +52 558451 4370
miwalakia83@gmail.com

Lourdes Bautista-Ynofuente

Universidad Nacional Mayor de San Marcos
Calle 4, Mz. A Lote 2 Asoc. Tambo Real – Lima
33. Código postal: Lima 15054
Perú
Tel. +51 97753 3424
lourdes.bautista1@unmsm.edu.pe

Luis Roberto Polo Bautista

Escuela Nacional de Biblioteconomía y
Archivonomía
Calz. Ticomán 645, La Laguna Ticoman,
Gustavo A. Madero, 07340 CDMX
México
Tel. +52 557851 0088
luispolo221@yahoo.com.mx

María Elena Luna-Morales

Coordinación General de Servicios
Bibliográficos, Centro de Investigación y de
Estudios Avanzados del IPN (Cinvestav-IPN)
Av. Instituto Politécnico Nacional #2508,
Colonia San Pedro Zacatenco, CDMX
México
Tel. +52 555303 8010
elena.5280@gmail.com

Maíra Fernandes Alencar

Centro de Educação, Comunicação e Artes
(CECA), Universidade Estadual de Londrina
Campus Universitário Rodovia Celso Garcia
Cid (PR 445), KM380 Caixa Postal: 10011
Londrina – Paraná CEP 86057-970
Brasil
Tel. +55 4399160 9753
alencarmfa@gmail.com

Manuel Blázquez Ochando

Facultad de Ciencias de la Documentación,
Universidad Complutense de Madrid
Calle de Santísima Trinidad 37, 28010 Madrid
España
Tel. +34 91394 6719
manublaz@ucm.es

María Magdalena Sierra Flores

Instituto de Ciencias Nucleares, Universidad
Nacional Autónoma de México
Circuito Exterior s/n, Ciudad Universitaria, Col.
Universidad Nacional Autónoma de México,
Deleg. Coyoacán. C.P. 04510, CDMX
México
Tel. +52 555622 4689 Ext. 3322
sierra@nucleares.unam.mx

María Victoria Guzmán Sánchez

Instituto Finlay de Vacunas
Avenida No. 21, No. 19810, entre 198 y 200,
Reperto Atabey, Playa, La Habana
Cuba
Tels. +53 7208 6086, +53 7208 0976
mvguzman@finlay.edu.cu

Michela Montesi

Facultad de Ciencias de la Documentación,
Universidad Complutense de Madrid
Calle de Santísima Trinidad 37, 28010 Madrid
España
Tel. +34 91394 6693
mmontesi@ucm.es

Miguel Ángel Pérez-Angón

Departamento de Física, Centro de
Investigación y de Estudios Avanzados del IPN,
(Cinvestav-IPN)
Av. Instituto Politécnico Nacional #2508,
Colonia San Pedro Zacatenco, CDMX
México
Tel.: +52 555303 8010
mperez@fis.cinvestav.mx

Romel Calero Ramos

Centro de Ciencias de la Complejidad,
Universidad Nacional Autónoma de México
Circuito Centro Cultural s/n (frente a
Universum), Cd. Universitaria, Delegación
Coyoacán, 04510, CDMX, México
Tel. +52 555622 6730 Ext. 0002
romel.calero@c3.unam.mx

Teresita Álvarez-Robles

Departamento Académico de Sistemas
Computacionales, Universidad Autónoma de
Baja California Sur
Norte #10-C Col. Los Laureles, C.P.91037,
Xalapa, Veracruz, México
Tel. +52 1228148 5095
alvarez88.t@gmail.com

Yadira Alejandra Orozco Osuna

Centro Universitario de Ciencias Exactas e
Ingenierías, Universidad de Guadalajara
Pedro M. Vélez #785 Col. Lomas del paradero,
Guadalajara, Jalisco, México
Tel. +52 1331551 7862
yadira.orozco4156@alumnos.udg.mx

Yolsy Gabriela Gamboa Calderón

Instituto de Ciencias Nucleares, Universidad
Nacional Autónoma de México
Circuito Exterior s/n, Ciudad Universitaria, Col.
Universidad Nacional Autónoma de México,
Deleg. Coyoacán. C.P. 04510, CDMX
México
Tel. +52 555622 4689 Ext. 3305
gamboa.yolsy@nucleares.unam.mx

Algoritmo para el análisis temático de documentos digitales

Luis Roberto Polo Bautista*
Karen Vanessa Martínez Acevedo*

Artículo recibido:
21 de febrero de 2021

Artículo aceptado:
14 de junio de 2021

Artículo de investigación

RESUMEN

El objetivo del artículo es presentar un algoritmo para asignar áreas temáticas a documentos digitales que sirva como herramienta de apoyo al análisis temático dentro de la organización de la información, con el fin de ser implementado en el desarrollo de vocabularios controlados. La metodología utilizada consistió en aplicar el Reconocimiento Óptico de Caracteres (ROC) y la Asignación Latente de Dirichlet (ALD) como las principales herramientas para el desarrollo de un algoritmo basado en el lenguaje de programación Python, que permite la lectura de archivos con extensión PDF para la obtención de los principales temas del *corpus* textual. Los resultados de la aplicación del

* Escuela Nacional de Biblioteconomía y Archivonomía, México
luispolo221@yahoo.com.mx miwalakia83@gmail.com

algoritmo demuestran su utilidad en el área de la indexación como un sistema para identificar y extraer temas relevantes de un documento específico en formato electrónico, permitiendo la automatización de procesos por parte del profesional de la información. De esta forma, se concluye su uso como desarrollo de puntos de acceso alternativos en función del contenido de los textos.

Palabras clave: Asignación Latente de Dirichlet; Algoritmos; Análisis Temático; Documentos Digitales

Algorithm for thematic analysis of digital documents

Luis Roberto Polo Bautista and Karen Vanessa Martínez Acevedo

ABSTRACT

The objective of the article is to present an algorithm for assigning subject areas to digital documents which serve as a support tool for thematic analysis within the organization of information, in order to be implemented in development of controlled vocabularies. The methodology used consisted in applying Optical Character Recognition (OCR) and Latent Dirichlet Allocation (LDA) as main tools for developing an algorithm based on Python programming language, which allows reading of files with a PDF extension in order to obtain the main themes of textual corpus. Results of the algorithm's application demonstrate its usefulness in the area of indexing as a system for identifying and extracting relevant topics from a specific document in electronic format, and allow automation of processes by the information professional. This way, its use as a development of alternative points of access based on the content of texts is concluded.

Keywords: Latent Dirichlet Allocation; Algorithms; Thematic Analysis; Digital Documents

INTRODUCCIÓN

En la actualidad, gran parte del conocimiento colectivo se encuentra digitalizado y almacenado en bases de datos en forma de noticias, páginas de internet, literatura científica, entre otras formas, lo que dificulta el proceso de búsqueda de información por parte de un usuario específico (Blei, 2012: 77). Por lo tanto, una tendencia dentro del ámbito bibliotecológico referente a solucionar el problema anterior es la indización y recuperación temática tomando como base el estudio de la organización de contenidos en internet (Naumis, 2015: 223).

Para ello, “necesitamos nuevas herramientas computacionales para ayudar a organizar, buscar y comprender esta gran cantidad de información” (Blei, 2012: 77). Piepenbrink y Gaur (2017: 11335) mencionan que, a “pesar de muchos avances en el dominio del análisis de contenido, todavía carecemos de herramientas sofisticadas que se puedan utilizar para analizar grandes datos textuales [...] como un enfoque del big data para identificar temas en el análisis de contenido de documentos digitales”.

“Desde el punto de vista de los desarrolladores de tecnología, la solución está en disponer de herramientas potentes y sumamente rápidas, que nos ayuden a tomar decisiones proactivas, y conducidas por un conocimiento acabado de la información” (Naumis, 2015: 224). Los métodos que pueden utilizar estas herramientas pueden tener distintas finalidades, como la extracción de términos, la clasificación, la agrupación de documentos y la identificación de relaciones (Contreras, 2018: 111).

En las bibliotecas tradicionales, el proceso de indización es efectuado de forma manual e intelectual por personal especializado, dando como resultado palabras clave en forma de descriptores o encabezamientos de materia que son utilizados como puntos de acceso en la búsqueda de la información (Contreras, 2018: 112). De acuerdo con la norma UNE 50-121-91 emitida por la Asociación Española de Normalización y Certificación (1991: 3), el proceso de indización consiste en las siguientes etapas:

- a) Examen del documento y determinación de su contenido: las partes más importantes del texto deben examinarse cuidadosamente y se debe prestar atención al título, resumen (si lo tiene), sumario o tabla de contenido, introducción, párrafos iniciales de capítulos y conclusiones.
- b) Identificación y selección de los conceptos principales del contenido: el indizador debe identificar los términos más apropiados que reflejen las nociones esenciales de la descripción del contenido.

- c) Selección de los términos de indización: el indizador debe verificar si los términos identificados en el paso anterior se encuentran en diccionarios y enciclopedias de autoridad, tesauros u clasificaciones temáticas.

Con base en estas etapas y al control de calidad respecto a la asignación temática mencionada en la misma norma, se podría mencionar que la indización tradicional es un proceso que conlleva tiempo y esfuerzo adicional por parte del bibliotecario profesional, considerando que en algunos casos particulares debe conocer o estar familiarizado con el área de conocimiento de un documento, con el objetivo de identificar y comprender con mayor precisión el vocabulario utilizado.

Asimismo, el proceso de indización puede conllevar a la asignación de descriptores o encabezamientos erróneos que no describan la totalidad o una parte esencial del contenido de un documento, dificultando el proceso de recuperación de información por parte de un usuario específico.

El algoritmo presentado en este artículo pretende optimizar de forma eficiente el tiempo de procesamiento de información, así como mejorar la identificación de temas de un documento, empleando el método de extracción de términos o palabras clave mediante el modelo de Asignación Latente de Dirichlet (ALD, en inglés Latent Dirichlet Allocation) por medio del aprendizaje no supervisado, así como también el Reconocimiento Óptico de Caracteres (ROC, en inglés Optical Character Recognition) para la conversión del texto completo de documentos digitales en archivos editables.

La ALD es un modelo probabilístico generativo de un *corpus*, cuya idea básica es que los documentos se representan como mezclas aleatorias sobre temas latentes, donde cada tema se caracteriza por una distribución sobre palabras. Es importante mencionar que el modelo no está necesariamente ligado al texto y tiene aplicaciones que involucran otros tipos de datos (Blei, Ng y Jordan, 2003: 996).

La notación y la terminología que utiliza el modelo ALD es la siguiente: una palabra es la unidad básica de datos discretos, definida como un elemento de un vocabulario indexado por $\{1, \dots, V\}$. Representamos palabras usando vectores de base unitaria que tienen un solo componente igual a uno y todos los demás componentes iguales a cero. Así, utilizando superíndices para denotar componentes, una palabra en una posición determinada en el vocabulario está representada por un V -vector w , de tal manera que $w^v = 1$ y $w^u = 0$ para $u \neq v$ (Blei, Ng y Jordan, 2003: 995).

“Un documento es una secuencia de N palabras indicadas por $w = (w_1, w_2, \dots, w_N)$ donde w_n es la posición de una palabra en la secuencia” (Blei, Ng y Jordan, 2003: 995).

“Un corpus es una colección de M documentos indicados por $D = \{w_1, w_2, \dots, w_M\}$ ” (Blei, Ng y Jordan, 2003: 995).

Por otro lado, el ROC es definido como “un software que convierte el texto impreso y las imágenes en forma digitalizada para que pueda ser manipulado por una máquina” (Islam, Islam y Noor, 2017: 1). A diferencia del cerebro humano, las computadoras no son lo suficientemente inteligentes para percibir la información disponible en una imagen o un documento impreso. Por lo tanto, se han presentado una gran cantidad de esfuerzos de investigación que intentan transformar este tipo de documentos a un archivo en un formato comprensible para una computadora (Islam, Islam y Noor, 2017).

El ROC asegura que la información de dichos documentos se digitalice a través de sistemas de tecnología de la información (TI) para su análisis. El procesamiento de lenguaje natural enriquece este proceso al permitir que esos sistemas reconozcan conceptos relevantes en el texto resultante, lo que es beneficioso para los análisis de aprendizaje automático (Smolaks, 2019). De esta manera, la tecnología ROC permitió convertir el texto completo de diversos documentos digitales en archivos editables con la finalidad de ser procesado por el modelo ALD. La combinación del ROC y el modelo ALD permitió la elaboración de un algoritmo basado en aprendizaje no supervisado que tome un documento en formato PDF como dato de entrada, convierta el texto completo a un archivo editable y automáticamente lo analice con el objetivo de mostrar los temas relevantes de ese documento de una forma dinámica.

El desarrollo de esta herramienta tiene como premisa principal facilitar la organización de la información por medio del análisis temático de documentos digitales, con la finalidad de ser implementado en la construcción de vocabularios controlados, como tesauros, taxonomías u ontologías que representen puntos de acceso alternativos en función al contenido de un texto.

METODOLOGÍA

Como se mencionó anteriormente, el algoritmo se elaboró tomando como base el ROC y la ALD a través del lenguaje de programación Python por medio del entorno Jupyter Notebook, el cual es “una aplicación web de código abierto que permite crear y compartir documentos que contienen código, ecuaciones, visualizaciones y texto narrativo” (Project Jupyter, 2021).

El método empleado para la aplicación del algoritmo se divide en dos partes fundamentales (*Figura 1*). La primera se refiere a la conversión del texto completo del documento y la segunda corresponde a la identificación de temas y su modelación visual por medio de HTML. Cabe señalar que el proceso

para comprobar la similitud de los temas generados por el algoritmo y temas propios de un documento (palabras clave) se realizó a partir del índice de Jaccard, el cual se mostrará posteriormente.

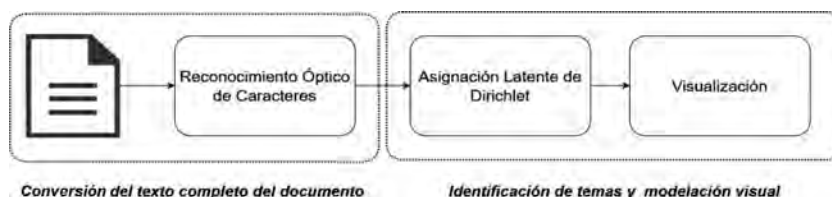


Figura 1. Proceso empleado para la aplicación del algoritmo

Conversión del texto completo del documento

En esta parte se utilizó el módulo PyMuPDF de Python, que entre otras funciones permite la identificación y conversión del texto de diferentes tipos de documentos a archivos editables. Es importante señalar que el funcionamiento de este algoritmo toma como base documentos en formato PDF, con el módulo PyMuPDF se hace uso del ROC, que permita generar archivos de texto simple (archivos con extensión txt) que contengan el texto completo de los documentos. A continuación, se muestra el pseudocódigo que se utilizó en este proceso:

Algoritmo: Conversión del texto completo de documentos

INICIO

- 1 Importar los módulos necesarios
- 2 Importar el documento en formato PDF
- 3 Leer el documento mediante el módulo PyMuPDF
- 4 Mostrar el número de páginas y los datos bibliográficos
- 5 Declarar una variable que contenga el nombre del nuevo archivo en formato txt
- 6 Para todas las páginas del documento hacer
- 7 Conversión del texto en formato UTF-8
- 8 Escribir el texto en un archivo con el formato definido en la línea 5
- 9 Definir dentro del archivo las divisiones de cada página
- 10 Cerrar el documento

FIN

Identificación de temas y modelación visual

En la sección de identificación de temas se utilizaron los módulos Numpy y Scikit-Learn para el cálculo de frecuencia y vectores de palabras y el modelo de Asignación Latente de Dirichlet (ALD) para la identificación de los temas relevantes. Para la visualización dinámica de los temas por medio de HTML se utilizó el módulo pyLDAvis. A continuación, se muestra el pseudocódigo que se utilizó en este proceso:

Algoritmo: Identificación de temas relevantes y modelación visual

INICIO

- 1 Importar los módulos necesarios
- 2 Importar el archivo de texto simple (archivo en formato txt)
- 3 Eliminar los artículos, pronombres, preposiciones, etc., del texto y normalizar el texto en letras minúsculas
- 4 Normalizar todas las formas de una misma palabra
- 5 Calcular los vectores de las palabras individuales que conforman cada frase
- 6 Crear una función que extraiga los vectores de las palabras y los guarde en una variable
- 7 Declarar en una variable el número de temas a procesar
- 8 Declarar en una variable el número de palabras relacionadas a los temas
- 9 Procesar los datos ingresados en la línea 6 con base al módulo ALD
- 10 Mostrar los datos procesados referentes a la línea 8 con base al vector asignado a cada palabra, así como el número de palabras relacionados a los temas
- 11 Declarar en una variable el nombre del archivo en formato HTML
- 12 Procesar los datos referentes a la línea 9 para que se puedan mostrar de forma dinámica
- 13 Preparar el archivo con los datos a mostrar
- 14 Guardar el archivo en una ruta determinada con extensión HTML

FIN

PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

Para analizar los temas obtenidos automáticamente mediante el algoritmo, existen diversos métodos que calculan la consistencia de la indización entre dos sistemas (manual y automatizado); en este trabajo se utilizó el índice de Jaccard para identificar el coeficiente de similitud entre los conjuntos de términos generados automáticamente y los conjuntos de términos generados manualmente de un documento específico.

Kosub (2019: 36) menciona que “el índice de Jaccard es una medición clásica de similitud con varias aplicaciones prácticas en la recuperación de información, extracción de datos, aprendizaje automático, entre otras”. Campos (2017) describe el índice de Jaccard como “la división entre el número de elementos en común que tienen los dos conjuntos sobre el número de elementos únicos que tiene la unión de ambos conjuntos”.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Donde:

A = Conjunto de datos 1

B = Conjunto de datos 2

$|A \cap B|$ = Número de elementos en común de ambos conjuntos

$|A \cup B|$ = Número de elementos únicos de ambos conjuntos

El algoritmo de análisis de temas descrito en la metodología fue implementado de forma individual a cada uno de los documentos de un conjunto de artículos científicos tomados de la base de datos SciELO, de tal manera que se pueda determinar el coeficiente de similitud por medio del índice de Jaccard entre los temas identificados automáticamente y los identificados de forma manual, que son principalmente palabras clave generadas por los propios autores.

Los artículos utilizados para este análisis fueron 50, divididos en cinco áreas del conocimiento (Ciencias agrarias, Ciencias biológicas, Ciencias de la salud, Ciencias exactas y de la tierra, y Ciencias sociales aplicadas) con 10 documentos de cada área; cinco en español y cinco en inglés.

A continuación, se muestra la aplicación del algoritmo en un artículo en español, donde se visualizan los resultados obtenidos. La *Figura 2* muestra un fragmento de texto del documento original y su modificación tras la eliminación de los artículos, pronombres, preposiciones, etc., así como su

normalización a letras minúsculas (este proceso se encuentra descrito en la identificación de temas y modelación visual).

Este proceso de eliminación de los artículos, pronombres, preposiciones, entre otros, facilitó el tratamiento automático del documento, permitiendo una mayor precisión en los resultados referentes a la identificación de palabras clave como temas relevantes del *corpus* textual.

De acuerdo con Gil Leiva (1997: 120), la supresión de estos términos, que comúnmente se conocen como palabras vacías o *stopwords* dentro del ámbito de la bibliotecología y ciencias de la información, queda justificado por los siguientes motivos:

- a) El descarte de las palabras vacías provoca que disminuya el número de palabras a procesar.
- b) Se reduce en un menor tiempo el análisis. Uno de los objetivos generales que se persigue al automatizar la indización es que el tiempo empleado por el programa sea similar o inferior al de un profesional.
- c) Es una ventaja no contar con este tipo de palabras en la etapa de búsqueda de términos contruidos de forma diferente respecto a los términos autorizados.
- d) “Para los textos de diferentes áreas temáticas e idiomas se ha comprobado que, aproximadamente el cincuenta por ciento de las palabras manejadas son palabras de este tipo”.

Estas palabras se eliminaron utilizando el módulo NLTK de Python por medio de las *stopwords* en idioma español, así como también de una lista donde se incorporaron algunas palabras vacías que no contempla el módulo, con el objetivo de extender aquellas palabras que no se requieran considerar para el análisis.

En este artículo se afrontan los desafíos para acceder y tratar los fondos de las hemerotecas nacionales de Colombia, Ecuador, México y Uruguay, que recogen noticias sobre eventos meteorológicos entre los siglos XIX-XX. Sobre estos periódicos se conforma un corpus de noticias que mediante lecturas técnicas y la aplicación de un proceso de bibliominería, utilizando diversas herramientas, permite iniciar la construcción de una red de ontologías.	artículo afrontan desafíos acceder tratar fondos hemerotecas nacionales colombia ecuador méxico uruguay recogen noticias eventos meteorológicos siglos xixxx periódicos conforma corpus noticias lecturas técnicas aplicación proceso bibliominería herramientas iniciar construcción red ontologías
---	--

Figura 2. Texto original y texto modificado tras la eliminación de las *stopwords*

En la *Tabla 1* se muestran los cuatro principales temas identificados por el algoritmo, considerando su frecuencia general.

Palabra	Frecuencia general
Noticias	34
Ontologías	29
Corpus	27
Meteorológicos	25

Tabla 1. Temas relevantes

En el trabajo de Chuang, Manning y Heer (2012: 75) se menciona que la frecuencia general de las palabras se basa en el cálculo de la prominencia del término, tal como se describe a continuación: para una palabra dada w , se calcula su probabilidad condicional $P(T|w)$: la probabilidad de que la palabra observada w fue generada por tema latente T . También se calcula la probabilidad marginal $P(T)$: la probabilidad de que cualquier palabra w' fue generada por tema T . Se define la distinción de la palabra como la divergencia Kullback-Leibler entre $P(T|w)$ y $P(T)$:

$$\text{Diferencia}(w) = \sum_T P(T|w) \log \frac{P(T|w)}{P(T)}$$

De igual manera, Chuang, Manning y Heer (2012: 75) mencionan que esta formulación describe (en un sentido teórico de la información) cuán informativo es el término específico w para determinar el tema generador, en comparación con un término seleccionado al azar w' . Por ejemplo, si una palabra w ocurre en todos los temas, la observación de la palabra nos dice poco sobre la mezcla temática del documento; por lo tanto, la palabra recibiría una puntuación de distinción baja. La calidad de un término está definida por el producto:

$$\text{Prominencia}(w) = P(w) \times \text{Diferencia}(w)$$

En la *Tabla 2* se muestra una comparación entre los temas identificados por el algoritmo y las palabras clave propuestas por el autor del artículo, con la finalidad de calcular su coeficiente de similitud. Como se puede observar, existen palabras entre ambos conjuntos de términos que son semánticamente similares, como: *Meteorología – Meteorológicos; Ontología – Ontologías; Prensa – Noticias*.

Por el otro lado, la palabra *Bibliominería* no tiene un término similar en los temas obtenidos a través del algoritmo, de lo que se infiere que esa palabra hace referencia a una metodología aplicada en el artículo y su frecuencia y relevancia son mínimas en comparación con otras.

El resultado del coeficiente de similitud con base al índice de Jaccard demuestra que existe un alto porcentaje de semejanza entre ambos conjuntos de temas, lo que indica que el grado de precisión de identificación y extracción de temas del algoritmo es adecuado.

Palabras clave propuestas por el autor	Temas identificados por el algoritmo
Bibliominería	Corpus
Meteorología	Meteorológicos
Ontología	Ontologías
Prensa	Noticias
Coeficiente de similitud	
$J(A, B) = 3 / 4 = 0.75 = 75 \%$	

Tabla 2. Cálculo del coeficiente de similitud

En la *Figura 3* se presenta la visualización de la modelación de temas en forma dinámica por medio de HTML. Esta representación visual fue elaborada mediante el módulo pyLDAvis y tiene la finalidad de responder algunas preguntas básicas sobre el modelado de temas, tales como ¿cuál es el significado de cada tema?, ¿qué tan prevalente es cada tema? y ¿cómo se relacionan los temas entre sí? Distintos componentes visuales responden a cada pregunta, de los cuales algunas herramientas son originales y otras son adaptadas de modelos ya existentes (Sievert y Shirley, 2014: 63).

El panel izquierdo presenta una vista global de los temas y responde las dos últimas preguntas. En esta vista, se trazan los temas como círculos en el plano bidimensional cuyos centros se determinan al calcular la distancia entre temas, y luego mediante el uso de escalas multidimensionales para proyectar las distancias entre temas en dos dimensiones (Sievert y Shirley, 2014: 63).

El panel derecho muestra una gráfica de barras horizontal, que representa los términos individuales que son más útiles para interpretar el tema

actualmente seleccionado a la izquierda, y permite a los usuarios responder la primera pregunta. Un par de barras superpuestas representan tanto la frecuencia de un término dado como la frecuencia específica de un tema correspondiente con un término (Sievert y Shirley, 2014: 63). Esta superposición se puede visualizar en el archivo HTML original.

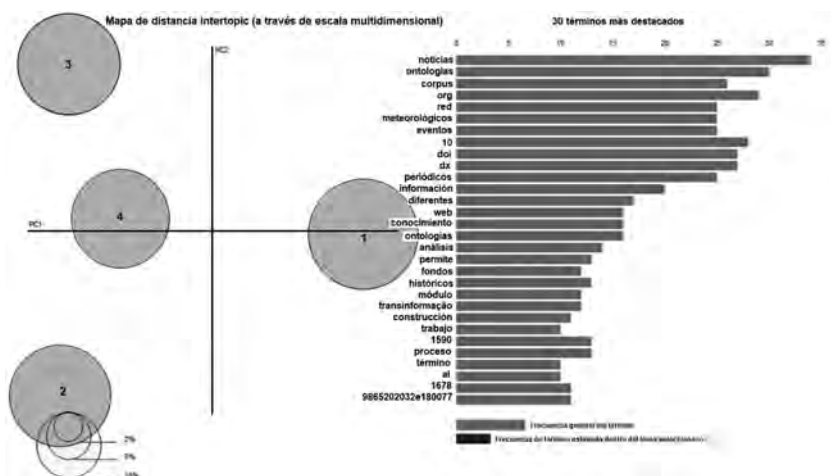


Figura 3. Visualización dinámica de temas

Las limitaciones que tiene esta visualización son que en algunos casos contempla palabras que originalmente debería omitir, como artículos, números o abreviaciones, aun cuando en el algoritmo se eliminan. Sin embargo, no representan un problema para la mayoría de los temas identificados, los cuales son los que se deberían de tomar en cuenta.

En la *Figura 4* se presentan los porcentajes del coeficiente de similitud obtenidos mediante el índice de Jaccard, aplicado a cada uno de los 50 artículos mencionados. La distribución es útil para representar las características de la dinámica de crecimiento y descenso entre los coeficientes, de tal manera que si el porcentaje se acerca a 100 hace referencia a que el total de palabras clave generadas por los autores y los temas identificados automáticamente son similares; por el contrario, si el porcentaje se acerca a 0 estos son diferentes entre sí. De acuerdo con lo mostrado en la gráfica, podemos mencionar que el promedio de similitud de todos los artículos analizados es de $\approx 69\%$.

El promedio general del coeficiente de similitud tras la aplicación del algoritmo en los artículos en español fue de $\approx 66\%$, el promedio en los artículos

en inglés fue de $\approx 72\%$. Con base en esto, podemos mencionar que, aunque el promedio de similitud fue más alto en los artículos en inglés, no indica que el rendimiento del algoritmo fue mejor, ya que existen diversos factores que pueden generar variaciones en los promedios de similitud, como la diferencia entre las palabras clave y los temas obtenidos.

En algunos casos, las palabras clave generadas por los autores estaban constituidas por una cadena de términos, esto causó un inconveniente al realizar la comparación, ya que el algoritmo sólo identifica temas conformados por una palabra (token). Para ello, se consideraron aquellas palabras claves semánticamente similares a los temas obtenidos automáticamente, y viceversa.

Para calcular la eficiencia del algoritmo, se tomó en cuenta el tiempo de ejecución considerando la extensión del documento. La ejecución del algoritmo se probó en una computadora con las siguientes especificaciones básicas: procesador Intel (R) Celeron (R) 1.60 GHz, memoria RAM de 4 GB, sistema operativo de 64 bits y un procesador x64. Para calcular el tiempo de ejecución se utilizaron cuatro documentos en formato PDF con diferente extensión (cantidad de hojas), con la finalidad de medir el tiempo de procesamiento de la información con base en el análisis temático. En las Figuras 5 y 6 se muestran los resultados obtenidos.

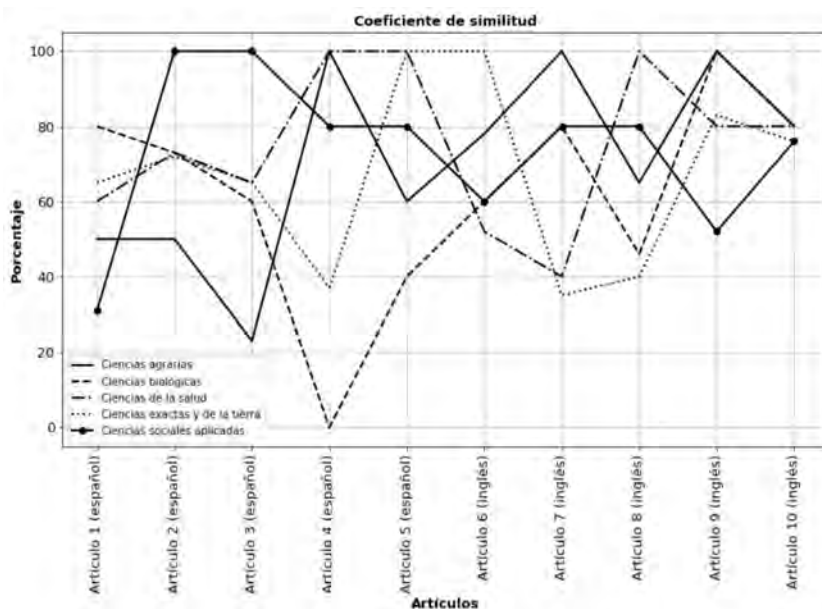


Figura 4. Coeficiente de similitud del corpus de artículos científicos

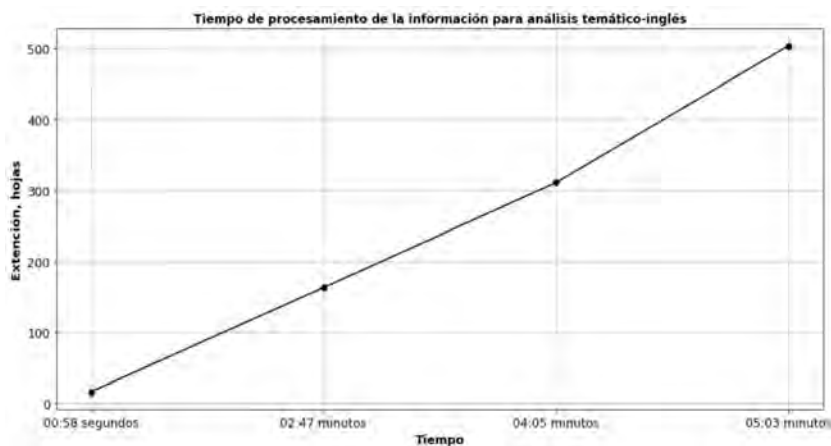


Figura 5. Tiempo de procesamiento de la información para análisis temático-inglés

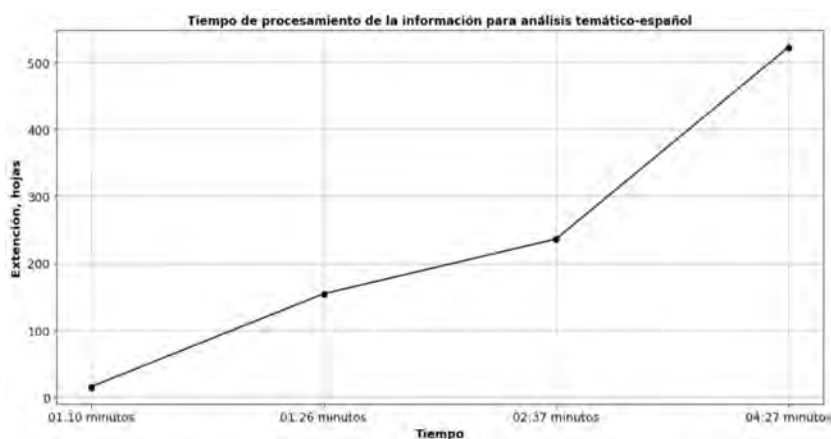


Figura 6. Tiempo de procesamiento de la información para análisis temático-español

Como se puede observar, el tiempo de ejecución del algoritmo referente al procesamiento de la información para el análisis temático de documentos es directamente proporcional a la extensión del documento, lo que indica que al aumentar una variable, en este caso la extensión del documento, también aumenta el tiempo de procesamiento.

De esta forma, podemos mencionar que el promedio de tiempo para procesar un documento con una extensión de entre 10-16 hojas y en idioma inglés o español es de 01:24 minutos y, en el otro extremo, el promedio de tiempo para procesar un documento con una extensión de 500-550 hojas en idioma inglés o español es de 05:05 minutos.

De acuerdo con Delgado y Sosa (1998), la obtención de un libro procesado completamente como producto final en las bibliotecas tradicionales consume en promedio, aproximadamente, más de una hora y media. Los tiempos promedios por cada tarea son los siguientes: catalogar: 24:09 minutos, clasificar e indizar: 46:08 minutos, preparación física del documento: 06:06 minutos, y la creación de fichas bibliográficas: 18:06 minutos.

Con base en lo anterior, consideramos que la utilización de este algoritmo como apoyo en el análisis temático de documentos digitales (indización) podría reducir significativamente el tiempo estimado para la tarea de clasificar e indizar, a aproximadamente menos de la mitad de tiempo.

DISCUSIÓN

Algunos inconvenientes al hacer uso de vocabularios controlados tradicionales como herramientas para el análisis temático de la información son los siguientes:

- a) El control de las formas plural y singular de las palabras por medio de la sinonimia y polisemia puede generar consecuencias negativas en la búsqueda y recuperación de la información, principalmente por la agrupación de diversos términos unificándolos en uno solo.
- b) Algunos encabezamientos o descriptores no reflejan el contenido de un documento específico.
- c) El tiempo que conlleva el proceso de indización es mayor.
- d) La cantidad de información a considerar para la asignación de temas es poca, y a veces no es suficiente para asignar correctamente un descriptor o encabezamiento.
- e) El profesional de la información debe estar familiarizado con el área de conocimiento del documento o comprender los conceptos que se encuentren en su contenido.

Por el contrario, las principales ventajas de la utilización de este algoritmo en el análisis temático son:

- a) Optimización de tiempo por parte del bibliotecario profesional.
- b) Mayor cantidad de información para procesar y en menor tiempo.
- c) Mayor precisión en la identificación de temas relevantes de un documento.
- d) Aunque en este trabajo sólo implementamos el algoritmo en idioma español e inglés, también se puede utilizar para documentos con diferentes idiomas.

- e) La implementación de este algoritmo en las bibliotecas no contempla costos financieros altos.
- f) Las habilidades tecnológicas que requiere el bibliotecario profesional para hacer uso de este algoritmo permiten una mejor adaptabilidad para manejar y gestionar diversas herramientas que puedan automatizar procesos a gran escala dentro de la biblioteca.

La identificación de temas (modelación de tópicos) generalmente se relaciona con las ciencias de la computación, específicamente con el aprendizaje automático y el procesamiento de lenguaje natural; sin embargo, en la actualidad, este tipo de herramientas tecnológicas pueden ser aprovechadas dentro de la bibliotecología y ciencias de la información como métodos de automatización de procesos. La identificación de temas puede ser implementada dentro del proceso de análisis temático en la organización de la información, con la finalidad de identificar temas relevantes en un documento específico.

En varias disciplinas se han desarrollado métodos que utilizan la identificación temática de un *corpus* de datos, generalmente asociado a textos cortos y minería de texto. De igual manera, dentro de la bibliotecología se han desarrollado varios procedimientos similares que tienen como objetivo automatizar el proceso de indización dentro de las unidades de información.

La principal diferenciación del algoritmo descrito en este artículo con los métodos anteriormente desarrollados recae en la utilización del ROC como herramienta de conversión de un documento a un archivo editable que contenga el texto completo de un documento, con la finalidad de facilitar el tratamiento y análisis de la información, así como también la optimización de tiempo y procesos por parte del profesional de la información.

Si bien en algunos casos particulares el rendimiento del algoritmo se vio afectado por la inclusión de palabras vacías como abreviaciones, pronombres, artículos, tal como se mostró en la *Figura 3*, estos no representan mayor inconveniente para seleccionar aquellos temas que reflejen el contenido del documento. De igual manera, un inconveniente del algoritmo es que sólo identifica temas compuestos por una única palabra (token), y generalmente existen áreas temáticas que deben expresarse con dos o más palabras.

Es importante señalar que, en los casos en que el porcentaje de similitud fue bajo en un documento, no significa que los temas identificados automáticamente no hayan sido adecuados, sino que las palabras clave propuestas por los autores generalmente no reflejaban la totalidad del contenido del artículo o su relevancia dentro del *corpus* textual era mínima.

CONCLUSIONES

Presentamos un algoritmo que permite identificar temas relevantes de un documento digital en formato PDF, utilizando el Reconocimiento Óptico de Caracteres y la Asignación Latente de Dirichlet a través del aprendizaje no supervisado, lo que indica que el algoritmo no necesita datos de entrenamiento para poder generar temas basados en el contenido del documento.

La utilización de este algoritmo identifica ≈ 30 temas relevantes de un documento específico en idioma español o inglés, dando oportunidad de seleccionar aquellos que mejor representan el contenido del recurso bibliográfico.

La implementación del algoritmo permite procesar una gran cantidad de información en un tiempo relativamente menor en comparación con un bibliotecario profesional, generando temas que reflejen la totalidad del contenido de un documento.

El promedio del índice de similitud entre los temas generados automáticamente y los propuestos por los autores (palabras clave) del *corpus* de artículos científicos utilizado es de $\approx 69\%$, lo que indica que más de la mitad de los temas generados por el algoritmo fueron semejantes a las palabras clave propuestas por los autores.

Como se mencionó anteriormente, el algoritmo calcula los temas relevantes de un documento con base a una distribución de las palabras que contiene, de esta forma no afectaría de qué área de conocimiento tratase el documento, la información del contenido se procesaría de igual forma teniendo relativamente los mismos resultados de tiempo de ejecución.

La utilización de este algoritmo en el proceso de análisis temático puede solucionar algunos inconvenientes como la optimización del tiempo, la cantidad de información a procesar, la asignación de temas *ad hoc* y la compatibilidad con el idioma del documento.

La implementación de este algoritmo dentro de las bibliotecas no implica costos financieros altos, ya que esta herramienta se puede replicar en una computadora con especificaciones básicas, y al utilizar este tipo de sistemas permite que el bibliotecario profesional obtenga nuevos conocimientos y habilidades que permitan un mejor desempeño en otras áreas, así como mejorar el aprovechamiento de tiempo.

Este algoritmo se puede utilizar para el desarrollo de puntos de acceso alternativos en función al contenido de los textos, así como la creación de vocabularios controlados, como tesauros, taxonomías u ontologías, que representen un dominio de conocimiento específico mediante el análisis del discurso.

Cabe mencionar que este algoritmo puede tener algunas mejoras potenciales relacionadas a la identificación de temas y su estructura en más de una

palabra (token), así como aumentar la cantidad de idiomas que pueda permitir en un documento.

En términos generales, los bibliotecarios profesionales pueden aprovechar este tipo de herramientas tecnológicas para crear sistemas capaces de automatizar procesos dentro de una biblioteca, como catalogar, indizar, clasificar, entre muchos otros, con la finalidad de darle una nueva perspectiva a la biblioteca y por ende a la bibliotecología y ciencias de la información.

REFERENCIAS

- Asociación Española de Normalización y Certificación. 1991. *Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*. https://nanopdf.com/download/metodos-para-el-analisis-de-documentos-de-terminacion-de-su_pdf
- Blei, David. 2012. "Probabilistic topic models". *Communications of the ACM* 55 (4): 77-84.
<https://doi.org/10.1145/2133806.2133826>
- Blei, David, Andrew Ng y Michael Jordan. 2003. "Latent Dirichlet Allocation". *Journal of Machine Learning Research* (3): 993-1022.
<https://web.archive.org/web/20120207011313/http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf>
- Campos, Diego. 2017. "Métricas de similitud para cadenas de texto. Parte III". *Sol-dAI*, 24 de diciembre de 2019.
<https://blog.soldai.com/metricas-conjuntos-emparejamiento-caracteres/>
- Chuang, Jason, Christopher Manning y Jeffrey Heer. 2012. "Termite: Visualization Techniques for Assessing Textual Topic Models". Trabajo presentado en la 12th International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 21-25 de mayo.
<http://vis.stanford.edu/files/2012-Termite-AVI.pdf>
- Contreras, Marcial. 2018. "Aplicación del algoritmo RAKE en la indización de documentos digitales". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 32 (75): 109-123.
<https://doi.org/10.22201/iibi.24488321xe.2018.75.57951>
- Delgado, Nora e Hilda Sosa. 1998. "Evaluación de la eficiencia en bibliotecas". *Investigación bibliotecológica: archivonomía, bibliotecología e información* 12 (24): 57-82.
- Gil Leiva, Isidoro. 1997. "La automatización de la indización: propuesta teórica-metodológica. Aplicación en el área de biblioteconomía y documentación". Tesis de licenciatura, Universidad de Murcia, Departamento de Información y Documentación, España.
<http://hdl.handle.net/10803/10917>
- Islam, Noman, Zeeshan Islam y Nazia Noor. 2017. "A survey on optical character recognition system". *Journal of Information & Communication Technology* 10 (2): 1-4.
<https://arxiv.org/ftp/arxiv/papers/1710/1710.05703.pdf>
- Kosub, Sven. 2019. "A note on the triangle inequality for the Jaccard distance". *Pattern Recognition Letters* 120, 36-38.
<https://doi.org/10.1016/j.patrec.2018.12.007>

- Naumis, Catalina. 2015. "Tendencias en la indización y recuperación temática", en *La información: perspectivas bibliotecológicas y distinciones interdisciplinarias*, coordinado por Jaime Ríos Ortega y César Augusto Ramírez Vélazquez, 223-240. México: UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información. http://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/L105/1/informacion_perspectivas_bibliotecologicas.pdf
- Piepenbrink, Anke y Ajai Gaur. 2017. "Topic models as a novel approach to identify themes in content analysis". *Academy of Management Proceedings* 2017 (1): 11335. <https://doi.org/10.5465/ambpp.2017.141>
- Project Jupyter. 2021. *The Jupyter Notebook*. <https://jupyter.org/>
- Sievert, Carson y Kenneth Shirley. 2014. "LDAvis: A method for visualizing and interpreting topics". Trabajo presentado en Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, Maryland, USA, 27 de junio. <https://doi.org/10.3115/v1/w14-3110>
- Smolaks, Max. 2019. "Intelligent document imaging with natural language processing and optical character recognition". *AI Business*, 16 de agosto de 2019. http://aibusiness.com/document.asp?doc_id=761011

Para citar este texto:

- Polo Bautista, Luis Roberto y Karen Vanessa Martínez Acevedo. 2021. "Algoritmo para el análisis temático de documentos digitales". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 13-31. <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58419>

Anexo. Visualización y descarga del algoritmo

El algoritmo de análisis temático de documentos digitales presentado en este artículo se puede visualizar y descargar completo por medio del siguiente enlace: https://github.com/LuisPoloBautista/Algoritmo_para_el_analisis_tematico_de_documentos_digitaes

Antes de ejecutar el algoritmo, asegúrese de tener en cuenta las recomendaciones mencionadas en el archivo "README".

Servicios accesibles a todos los usuarios en las bibliotecas universitarias españolas: estado de la cuestión

David Domingo Pàmies*

Carina Rey Martín**

Concepción Rodríguez Parada**

Artículo recibido:
9 de marzo de 2021

Artículo aceptado:
16 de junio de 2021

Artículo de investigación

RESUMEN

Hasta el momento, las universidades han centrado la atención a la discapacidad en el ámbito de la docencia mediante la puesta en marcha de programas o planes que contemplan la adquisición de software específico, medidas de ayuda a la docencia y la contratación de personas de acompañamiento, entre otras estrategias. Sin embargo, los servicios universitarios, como, por ejemplo, los ofrecidos por las bibliotecas, no han mostrado el mismo nivel de adaptación. A diferencia de lo que ocurre con las bibliotecas públicas,

* Centre de Recursos per a l'Aprenentatge i la Investigació, Campus Catalunya, Universitat Rovira i Virgili, España david.domingop@urv.cat

** Departament de Biblioteconomia, Documentació i Comunicació Audiovisual, Facultat de Informació y Medios Audiovisuales, Universitat de Barcelona, España carina.rey@ub.edu crodriguezp@ub.edu

la bibliografía centrada en accesibilidad y servicios inclusivos en bibliotecas universitarias o CRAI es muy escasa. Este trabajo pretende paliar esta carencia a partir del estudio bibliográfico y posterior análisis cualitativo de las necesidades de los usuarios con discapacidad y su atención en las bibliotecas públicas, para después revisar las buenas prácticas y los servicios específicos prestados por las universidades y sus respectivos CRAI tal como se difunden en las páginas web de cada una de ellas.

Aunque el marco legal y normativo fija pautas genéricas para garantizar la accesibilidad física y a los servicios bibliotecarios, este artículo pretende concretar dichas pautas para adaptarlas a las circunstancias específicas de las bibliotecas universitarias.

Palabras clave: Accesibilidad; Inclusividad; Discapacidad; Biblioteca Universitaria Inclusiva

Accessible services to all users in Spanish University Libraries: state of the art

David Domingo Pàmies, Carina Rey Martín and Concepción Rodríguez Parada

ABSTRACT

Until now, universities have focused its attention on disability in the field of teaching through the implementation of programs or plans which include, among others, the acquisition of specific software, teaching aid measures and the recruitment of support staff. However, university services, such as those offered by libraries, have not shown the same level of adaptation. Unlike what happens with public libraries, the bibliography focused on accessibility and inclusive services in university libraries is very scarce. This work aims to alleviate this gap through a bibliographical study and subsequent qualitative analysis of the needs of users with disabilities and their care in public libraries, to later review the good practices and specific services provided by the universities and their respective libraries as they are disseminated on the web pages of each of them.

Although the legal and regulatory framework establishes generic guidelines to guarantee physical and

library services accessibility, this article aims to specify those guidelines for adapting them to the specific circumstances of university libraries.

Keywords: Accessibility; Inclusion; Disability; Inclusive University Library

INTRODUCCIÓN

El *Plan de Actuaciones* del CRAI (Centro de Recursos para el Aprendizaje y la Investigación) de la Universitat Rovira i Virgili (Tarragona, España) establece como uno de sus objetivos más importantes llegar a ser un CRAI inclusivo en el medio plazo. Así, el curso 2019-2020 incorporó a su plantilla cinco trabajadores con discapacidad cognitiva, medida promovida por la ALA en el programa “Value of libraries in promoting social justice and inclusion” (ALA, 2020), lo que permitió conocer de primera mano la necesidad de que las bibliotecas sean accesibles para todos. La búsqueda de información al respecto puso de manifiesto la escasez de estudios centrados en la atención a los usuarios con discapacidad en las bibliotecas universitarias, que contrasta con los numerosos trabajos publicados sobre el tema en el ámbito de las bibliotecas públicas.

Las líneas que siguen buscan contribuir a llenar un vacío a partir de la consecución de los objetivos siguientes:

- Construir un marco teórico, por ahora inexistente, sobre la atención de las personas con discapacidad en las bibliotecas universitarias.
- Precisar la terminología específica.
- Identificar ejemplos de buenas prácticas de bibliotecas públicas y universitarias españolas en el campo de la accesibilidad y la inclusividad.

METODOLOGÍA

Dada la falta de bibliografía sobre servicios inclusivos en las bibliotecas universitarias o CRAI, se ha optado, en primer lugar, por hacer un estudio cualitativo de los ofrecidos por las bibliotecas públicas –cuyos resultados serán expuestos en el apartado “Bibliotecas públicas” de este trabajo– porque pueden servir de ejemplo de buenas prácticas dado el tiempo que hace que prestan atención a este colectivo de usuarios; mientras que, en el caso de las

bibliotecas universitarias, se han revisado sus páginas web y también las de sus respectivas universidades para ver qué servicios ofrecen a los usuarios con discapacidad.

MARCO TEÓRICO

Hasta ahora nos hemos referido a conceptos como inclusividad, accesibilidad, discapacidad, etc., es por ello que creemos necesario definirlos para saber exactamente a qué nos referimos en cada caso.

En el preámbulo de la *Convención internacional sobre los derechos de las personas con discapacidad* se precisa que la *discapacidad* es la limitación resultante de la interacción entre las personas con deficiencias y las barreras provocadas por la actitud y el entorno que imposibilitan su participación en igualdad de condiciones respecto a los demás ciudadanos (ONU, 2006). En este sentido, “La discapacidad es una circunstancia personal y un hecho social producto de la interacción de factores tanto individuales (el hecho físico, la materialidad de la discapacidad, que en ocasiones consiste o deriva de un desorden de salud), como sociales (el entorno de la persona que presenta el hecho físico)” (Velarte Pérez *et al.*, 2011: 14-15).

En España, la *Ley General de derechos de las personas con discapacidad* considera *accesibles* los edificios, espacios, recursos, etc., que pueden ser utilizados y disfrutados por todo el mundo con seguridad y comodidad y de la forma más autónoma y natural posible (RD 1/2013, art. 2.k). En el caso de una biblioteca universitaria o CRAI será accesible si permite a los usuarios con discapacidad *acceder, utilizar* sus recursos, *orientarse y comunicarse* en condiciones equivalentes a las del resto de usuarios (Givry, 2010: 273).

Ahora bien, para conseguir un CRAI verdaderamente *inclusivo*, hacerlo accesible es sólo el primer paso. En un CRAI inclusivo, todos los usuarios son tratados con respeto, son valorados por sus habilidades distintivas y tienen igual acceso a los recursos para poder contribuir completamente al éxito de la universidad (ALA, 2017). Para Ciuró (2019: 35) se trata de algo más sencillo aún: “La inclusión es esto: que se sientan parte de nuestra comunidad bibliotecaria y nos ayuden a crecer”.

En lo que se refiere a la *igualdad*, si no se delimita bien, puede esconder la mayor de las desigualdades. Este es el sentido de la *Figura 1*. Es decir, hay que evitar que la aplicación rígida del principio de igualdad acentúe las desigualdades de base hasta hacerlas insuperables. En cambio, si a cada uno se le brinda *equitativamente* un apoyo adaptado a su necesidad, será capaz de superar todas las barreras que se le planteen.

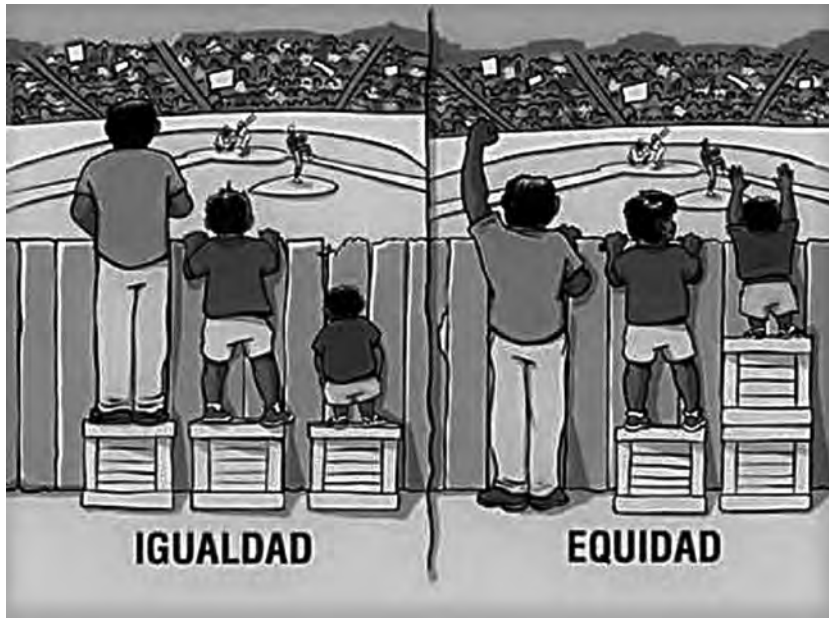


Figura 1. Representación de Igualdad vs. Equidad

Fuente: <https://www.semfy.com/extracto-atencion-comunitaria/>

Otro de los conceptos que se suelen utilizar es lo que la ya citada *Convención* de la ONU llama “ajustes razonables” (ONU, 2006: art. 2.4), que consiste en ir adaptándose a medida que se plantean casos particulares. Dichos “ajustes razonables” son el mínimo que marca la ley. Así ocurre, por ejemplo, en el marco jurídico español (RD 1/2013, art. 18.2); sin embargo, este *mínimo* no es suficiente si se pretende que el CRAI sea el “corazón de la universidad”.

Un estudio de la Fundación Universia (2018) estima que la mitad de los estudiantes con discapacidad de las universidades españolas no la declaran; además, muy a menudo, dichas discapacidades son invisibles por lo que la inclusión es más difícil. Pero que no las veamos no significa que no estén ahí (Missiroli, 2018: 13). No podemos estar esperando que todos los usuarios vengan a explicarnos su estado de salud. Las *Pautas de acceso a bibliotecas para personas con discapacidad* de la IFLA recuerdan que “las personas con discapacidad deben superar barreras no sólo físicas, sino también psicológicas, para ir a la biblioteca y expresar sus necesidades” (Irvall y Nielsen, 2016: 14).

Hacer adaptaciones individualizadas cuando el usuario lo pide o “ajustes razonables”, pues, no parece lo más recomendable. Sería preferible que una biblioteca diseñara sus servicios y recursos pensando en su uso por todos sin

necesidad de adaptación o sólo la mínima necesaria (Nygaard, 2018). En esto consiste el “diseño universal” (ONU, 2006: art. 2.5). Así, antes de buscar soluciones específicas, la primera vía que hay que explorar siempre es la de las soluciones generales que benefician a todos los usuarios. Por ejemplo, el servicio de reserva telemática de documentos, que pueden pedir todos los usuarios, beneficia especialmente a aquellos que no pueden acceder a los estantes, sin tener que solicitar un “servicio especial”. Es, pues, un enfoque radicalmente inclusivo.

Marco legal y normativo

El punto de partida del marco jurídico español es la *Convención internacional sobre los derechos de las personas con discapacidad*, ratificada por España en 2007, que establece las definiciones de “discapacidad”, “diseño universal” y “ajustes razonables” ya expuestas y fija la obligación de asegurar el acceso de las personas con discapacidad a los edificios y a los servicios de información (ONU, 2006: art. 9.1). La obligatoriedad de que las bibliotecas universitarias sean accesibles es recogida específicamente en el artículo 14.6 de la *Ley de la lectura, del libro y de las bibliotecas* (Ley 10/2007), el artículo 8.3 del *Real Decreto de creación, reconocimiento, autorización y creación de universidades* (RD 420/2015), la disposición adicional 24.4 de la *Ley Orgánica de Universidades* (LO 6/2001) y en el artículo 15.2 del *Estatuto del Estudiante Universitario* (RD 1791/2010).

El marco normativo viene regulado por las secciones de la IFLA *Library Services to People with Special Needs Section* (LSN), *Libraries Serving Persons with Print Disabilities Section* (LPD) y por la Agenda 2030 para el Desarrollo Sostenible de la ONU.

La LSN se ocupa de las dificultades de acceso a los servicios de las personas con discapacidades físicas, cognitivas y mentales, además de las que están en hospitales, residencias, cárceles, sin hogar, refugiados, etc. (IFLA, 2019). Entre los documentos publicados por la LSN, destacamos las *Pautas para el Acceso a las Bibliotecas para personas con discapacidad* (Irvall y Nielsen, 2016), así como guías sobre servicios bibliotecarios para personas sordas, con demencia, dislexia y sobre el formato de Lectura Fácil; mientras que la LPD atiende a las personas que no pueden leer un texto impreso o pueden no tener suficiente comprensión lectora (IFLA, 2019). Más centrada en la propuesta de formatos alternativos (braille, DAISY, tipografía grande, etc.) también ha desarrollado directrices sobre el servicio para usuarios de braille y sobre biblioteca para ciegos en la era de la información.

El lema de la contribución de la IFLA para la consecución de los Objetivos de Desarrollo Sostenible de la Agenda 2030 es “Acceso y oportunidades

para todos” (IFLA, 2015), lo que da cuenta de la importancia que concede a la inclusividad de las bibliotecas. La LSN y la LPD son una buena muestra de ello ya que no abordan la discapacidad desde el punto de vista de la deficiencia del individuo (visual, auditiva, motora, etc.), sino del obstáculo que éste encuentra en la biblioteca. Si la discapacidad es la interacción de ambas cosas, el ámbito en el que puede incidir la biblioteca es en la eliminación de sus propias barreras.

Bibliotecas públicas

La búsqueda bibliográfica de ejemplos de buenas prácticas de inclusión y accesibilidad en bibliotecas públicas debe hacerse con cautela. En la gran mayoría de casos, están escritos por personal de la institución destacando los logros pero sin mencionar ningún punto débil, por lo que a veces es difícil distinguir el optimismo de la mera propaganda.

Quizá sea la accesibilidad arquitectónica la que se ha cuidado más hasta la fecha: puertas automáticas, franjas podotáctiles, contraste de color de enchufes y mobiliario, ascensores y lavabos accesibles, plano en relieve, señalización en braille y la instalación de un bucle magnético. Destacan, en este ámbito, la Biblioteca Montserrat Abelló de Barcelona (Carbonell, Matamoros y Peix, 2019) y la Médiathèque José Cabanis de Toulouse (Renaudin, Henard y Andissac, 2015), diseñadas desde el origen para ser accesibles y que han sido puestas como ejemplos a seguir en congresos de la IFLA (Andissac, 2014; Peix Cruz, 2018). Cabe señalar que, según Bonello (2009: 38), las asociaciones de ciegos rechazaron la instalación de franjas podotáctiles en la mediateca de Toulouse porque los ciegos “irían acompañados”.

Respecto a la colección, muchas bibliotecas tienen una sección especial (centro de interés, espacio, *pôle*, etc.) con documentos en formato accesible (braille, audiolibros, letra grande, lectura fácil, lenguaje pictográfico, etc.). Peix Cruz (2014: 47) cree que segregar estos fondos no es inclusivo y recomienda mezclarlos con el resto ya que el catálogo permite recuperarlos por separado, mientras que Gálvez Domingo y López Sánchez (2017: 30) subrayan que situarlos en un fondo especial permite hacer estudios de uso.

El equipamiento utilizado para consultarlos consiste en ordenadores con software de aumento y lectura de pantalla y telelupa. La Médiathèque José Cabanis también ofrece una línea de braille y escáner. Estas instalaciones, sin embargo, pueden tener muy poco uso, especialmente si se colocan en lugares ocultos y no se les da suficiente difusión, así que habrá que concentrar los esfuerzos en la calidad de la mediación humana, servicios y actividades y no sólo en la tecnología (Bonello, 2009: 34-41). En el ámbito de servicios y

actividades, se mencionan el préstamo y lectura a domicilio, guías y folletos en lectura fácil, exposiciones de concienciación, clubes de lectura fácil y actividades en lengua de signos.

Por último, se destacan dos aspectos relacionados con el personal bibliotecario. Por un lado, la necesidad de que se formen en temas de discapacidad como, por ejemplo, que tengan nociones básicas de lengua de signos; por otro, se recomienda que la plantilla esté formada también por trabajadores con discapacidad (visual, auditiva, intelectual, etc.), ya que permite mejorar la percepción del resto del personal y por su implicación con usuarios que presenten alguna discapacidad (Ciuró, 2019).

ESTUDIO DE CASOS: BIBLIOTECAS UNIVERSITARIAS

Como se ha señalado con anterioridad, es difícil encontrar bibliografía con ejemplos prácticos de accesibilidad en bibliotecas universitarias. Ahora bien, la escasa bibliografía publicada no significa que las bibliotecas universitarias no hagan nada en este campo. Empezando por nuestro contexto más inmediato, analizaremos las páginas web de las bibliotecas universitarias catalanas miembros del Consorcio CSUC para conocer los servicios dirigidos a los usuarios con discapacidad y, más concretamente, comparar los ofrecidos por sus bibliotecas.

Tres bibliotecas del total de siete universidades públicas catalanas no incluyen en su página web información precisa sobre los servicios dirigidos a usuarios con discapacidad (*Tabla 1*).

Universidad	Enlace a la página de la biblioteca de servicios para usuarios con discapacidades
Universitat Autònoma de Barcelona (UAB)	https://ddd.uab.cat/record/100014
Universitat de Barcelona (UB)	https://crai.ub.edu/ca/que-ofereix-el-crai/serveis-persones-necessitats-especificques
Universitat de Girona (UdG)	https://biblioteca.udg.edu/ca/compromis-social/inclusio-i-accessibilitat
Universitat de Lleida (UdL)	No tiene
Universitat Politècnica de Catalunya (UPC)	No tiene
Universitat Pompeu Fabra (UPF)	https://www.upf.edu/web/biblioteca-informatica/suport-a-persones-amb-necessitats-especials
Universitat Rovira i Virgili (URV)	No tiene

Tabla 1. Webs sobre servicios para usuarios con discapacidad de las bibliotecas de las universidades públicas catalanas (fecha de consulta: 13/07/2020)

Sin embargo, que no dediquen una sección específica en el sitio web no significa que no ofrezcan servicios a los usuarios con discapacidad. Pueden tenerlos y no publicitarlos olvidando con ello la importancia que dan a esta información los futuros estudiantes con discapacidad en el momento de decidir en qué universidad matricularse.

En la *Tabla 2* se relacionan los servicios de las bibliotecas de las cuatro universidades catalanas que sí disponen de esta información en su web.

Servicio	UAB	UB	UdG	UPF
Sitio web de la biblioteca con información sobre la accesibilidad de los edificios	-	Sí	Sí	Sí
Alarmas de incendio visibles y auditivas	-	-	Sí	-
Plan de evacuación adaptado para personas con discapacidad	-	-	Sí	-
Mesa individual adaptada para usuarios con movilidad reducida	-	-	-	Sí
Fondo en formatos accesibles: braille, audiolibros, etcétera	-	Sí	-	-
Lupa-televisión	En todas	-	En todas	-
Cabinas adaptadas para usuarios con discapacidad	6	1	-	-
Teclado o periféricos adaptados	Cabinas	Cabinas	-	-
Software lector de pantalla (Jaws)	Cabinas	14	En todas	En todas
Software amplificador de pantalla (Zoomtext)	Cabinas	2	En todas	En todas
Software para la dislexia (ClaroRead)	Cabinas	-	-	-
Permiten personalizar las opciones de accesibilidad de un ordenador del CRAI	-	Sí	-	-
Ordenador reservado para usuarios con discapacidad visual	-	-	-	1
Servicio de petición anticipada de documentos	-	-	-	Sí
Servicio de localización y entrega	Sí	Sí	-	-
Préstamo a personas autorizadas	Sí	Sí	-	-
Préstamo adaptado o con duración especial	Sí, el doble	Sí, un tercio más	Sí, no especifica	-
Reserva de puntos de lectura	-	Sí	-	-
Reserva de ordenadores portátiles	-	Sí	-	-
Servicio de digitalización de documentos	-	Sí	-	-

Proporcionan los libros excluidos para ser escaneados por una institución externa	Sí	-	-	-
Formación a medida para usuarios con discapacidad	Sí	-	Sí	-
Número de bibliotecas	9	16	3	3

Tabla 2. Resumen de los servicios para usuarios con discapacidad de las bibliotecas de las universidades públicas catalanas

Fuente: elaboración propia

En tres de ellas encontramos información en su web sobre la accesibilidad de los edificios de la biblioteca. La UdG menciona además que el plan de evacuación ha sido adaptado y las alarmas son visuales y sonoras.

Sólo la UB afirma que dispone de fondos en formatos accesibles como braille, audiolibros, etc., y los enlaza al catálogo. Como suele ocurrir con estas colecciones, en su mayoría son cuentos infantiles por lo que no se ve la relevancia para un estudiante universitario a no ser que estudie para ser maestro.

Tanto la UAB como la UdG afirman tener una telelupa en cada biblioteca. También las tiene la UdL que, a pesar de no disponer de una sección específica en el sitio web sobre servicios para usuarios con discapacidad, sí que menciona su instalación en cada biblioteca y también en la información del Servicio de Apoyo de la Universidad (<http://www.udl.cat/ca/serveis/seu/UdLxtothom/recursos/aids/>).

La UAB informa que cuenta con seis cabinas adaptadas para usuarios con discapacidad y la UB con una, todas ellas equipadas con ordenadores que disponen de teclado adaptado. Las bibliotecas de la UdG y la UPF ofrecen acceso al software amplificador de pantalla Zoomtext y al lector de pantalla Jaws. El CRAI de la UB menciona el programa Jaws en 14 de sus 16 sedes y Zoomtext en dos; las bibliotecas UAB, en cambio, sitúan ambos programas en las cabinas adaptadas, así como el programa Claread, dirigido a personas con dislexia. La UAB condiciona el uso de las cabinas a la inscripción en el Servicio de Apoyo de la Universidad.

El CRAI de la UB ofrece la posibilidad de personalizar las opciones de accesibilidad del sistema operativo de un ordenador en cada una de sus bibliotecas. La UPF menciona un ordenador reservado para usuarios con discapacidad visual.

Sólo la UPF enumera servicios generales que “pueden ser especialmente útiles para las personas con discapacidades físicas o de movilidad”, como la reserva y la renovación telemática de préstamos, aunque, de hecho, son servicios que prestan todas las bibliotecas del CSUC. La UAB y la UB mencionan la posibilidad de que el personal vaya a buscar un libro en las estanterías y

tienen regulado el préstamo a personas autorizadas, así como un préstamo más largo para estos usuarios (el doble en la UAB, un tercio más en la UB). La UdG menciona la posibilidad de prorrogar el préstamo sin especificar por cuánto tiempo.

El CRAI UB contempla la reserva de puntos de lectura y portátiles, así como la posibilidad de reproducir documentos en formatos accesibles, acogiendo a la excepción de la *Ley de Propiedad Intelectual* para personas con discapacidad (RD Leg. 1/1996, art. 31 ter). El Servei de Biblioteques de la UAB, en cambio, permite prestar libros de consulta en sala para escanearlos, pero deja muy claro que el escaneo y la transcripción deben ser realizados por una “institución externa”.

Por último, en el ámbito de la formación, sólo la UAB y la UdG mencionan la posibilidad de formación a medida para los usuarios con discapacidad.

Las únicas referencias a la biblioteca incluidas en la guía oficial de servicios universitarios para estudiantes con discapacidad (Unidiscat, 2019) son las relacionadas con ordenadores y programas como Jaws y Zoomtext adaptados para estudiantes con baja visión y sólo en el caso de las bibliotecas de la UPF y de la UdG. Para el resto de universidades, aunque se mencionan dichos programas en ningún caso se informa de que están disponibles en el CRAI.

Para el análisis del resto de bibliotecas universitarias españolas, hemos tomado las 10 mejores universidades del U-Ranking Volumen (<https://www.u-ranking.es/analisis.php>). La razón para elegirlo es precisamente la principal crítica que se le hace: que prioriza a las grandes universidades sobre las pequeñas. En nuestro caso, pensamos que serían las que tendrían más recursos y, tal vez, por ello podrían ofrecer un mayor número de prestaciones al colectivo de usuarios con discapacidad (*Tabla 3*).

Universidad	Enlace a la página de la biblioteca de servicios para usuarios con discapacidades
Universidad Complutense de Madrid (UCM)	https://biblioteca.ucm.es/serviciosdiscapacitados
Universitat de Barcelona (UB)	https://crai.ub.edu/ca/que-ofereix-el-crai/servis-persones-necessitats-especificques
Universidad de Granada (UGR)	https://biblioteca.ugr.es/pages/biblioteca_ugr/bibliotecas_centros/derecho/biblioteca-accesible
Universidad de Sevilla (US)	https://bib.us.es/utiliza_la_biblioteca/atencion
Universitat de València (UV)	https://www.uv.es/uvweb/servei-biblioteques-documentacio/ca/sbd/accessibilitat/accessibilitat-1286109772452.html

Universidad del País Vasco (UPV)	https://www.ehu.eus/es/web/discapacidad/liburutegia
Universitat Autònoma de Barcelona UAB)	https://ddd.uab.cat/record/100014
Universidad Politécnica de Madrid (UPM)	No tiene
Universitat Politècnica de València (UPV)	No tiene
Universitat Politècnica de Catalunya (UPC)	No tiene

Tabla 3. Webs sobre servicios para usuarios con discapacidad de las 10 primeras universidades del U-Ranking (fecha de consulta: 13/07/2020)

A continuación, analizaremos la información ofrecida por todas ellas (UCM, UGR, US, UV y UPV, excepto UB y UAB que ya han sido consideradas) (*Tabla 4*).

Servicio	UCM	UGR	US	UV	UPV
Sitio web de la biblioteca con información sobre la accesibilidad de los edificios	-	1	-	2	-
Edificio accesible (ascensores, baños, puerta, etc.)	-	1	-	2	-
Plan de evacuación adaptado para personas con discapacidad	-	-	-	-	-
Mesa individual adaptada para usuarios con movilidad reducida	Sí	Sí	Sí	Sí	-
Equipamiento adaptado (atriles, pasapáginas, lopes, etc.)	-	-	-	Sí	-
Lupa-televisión	-	-	-	Sí	Sí
Cabinas adaptadas para usuarios con discapacidad	11	-	-	Sí	4
Ordenador reservado para usuarios con discapacidad visual	-	5	14	Sí	-
Teclado o periféricos adaptados	Sí	-	-	-	Sí
Software lector de pantalla (Jaws)	Sí	Sí	-	-	-
Software lector de pantalla (NVDA)	-	Sí	-	-	-
Software amplificador de pantalla (Zoomtext)	Sí	Sí	-	-	Sí
Software para la dislexia (ClaroRead)	-	-	-	-	-
Lector DAISY	-	Sí	-	-	-
Servicio de petición anticipada de documentos	Sí	-	-	Sí	-
Servicio de localización y entrega	-	-	Sí	Sí	-
Préstamo a personas autorizadas	Sí	-	Sí	Sí	-
Préstamo adaptado o con duración especial	Sí, el doble	-	Sí, el doble	Sí	-
Préstamo por correo	Sí	-	Sí	-	-

Reserva de puntos de lectura	-	Sí	-	-	-
Reserva de ordenadores portátiles	-	Sí	-	-	-
Servicio de digitalización de documentos	-	-	-	-	-
Proporcionan los libros excluidos para ser escaneados por una institución externa	Sí	-	-	-	-
Formación a medida para usuarios con discapacidad	-	-	Sí	-	-
Número de bibliotecas	26	23	17	8	12

Tabla 4. Resumen de los servicios para usuarios con discapacidad de las bibliotecas de las 10 primeras universidades del U-Ranking

Fuente: elaboración propia



Figura 2. Universidad de Granada. Biblioteca de la Facultad de Derecho

Fuente: https://biblioteca.ugr.es/pages/biblioteca_ugr/bibliotecas_centros/derecho/biblioteca-accesible

Ninguna página web informa sobre la accesibilidad de cada una de las bibliotecas de su universidad. Este es el caso de la UGR, que sólo facilita información de la Biblioteca de la Facultad de Derecho, diseñada en origen para ser accesible (puertas automáticas, franjas podotáctiles, ascensores adaptados, etc.) (Figura 2).

Dos bibliotecas de la UV, en concreto la Biblioteca d'Educació María Moliner y la Biblioteca d'Humanitats Joan Reglà, disponen del certificado AENOR de Accesibilidad Universal UNE 170001-2:2007.

Mientras que la mayoría de bibliotecas hacen hincapié en los ordenadores con programas para discapacidad visual como Zoomtext (es el único servicio de este tipo que menciona la UPV), la UV los cita de paso y se centra en equipamientos mucho más baratos y ligeros, pero que facilitan la comodidad de los usuarios: 10 lupas de aumento, 47 atriles, cinco pasapáginas, 15 auriculares con micrófono, etc.

La Universidad Complutense (UCM) tiene un total de 11 “puestos adaptados” (uno o dos por facultad): ordenadores con alguna versión de Jaws o

Zoomtext, y algún periférico adaptado (teclado, *joystick*, etc.). La Universidad de Sevilla (US) menciona 14 puestos informáticos “con ayudas visuales”, sin concretar cuáles. La UGR (*Figura 3*) ofrece un espacio compuesto por cinco ordenadores con Jaws, NVDA, un escáner OCR y un lector DAISY, al que llama SADDIS (Servicio de Apoyo Documental a personas con Discapacidad).



Figura 3. SADDIS de la Universidad de Granada
Fuente: <https://biblioteca.ugr.es/pages/servicios/saddis>

En cuanto al préstamo, las bibliotecas de la UCM y de la US permiten el préstamo por correo, junto con el préstamo delegado. Ambas duplican, también, la duración del tiempo de préstamo.

La UCM también contempla, como la UAB, la posibilidad de retirar un libro excluido para su escaneo, preferentemente en un puesto adaptado de la propia facultad o por una “asociación de personas con discapacidad”. Sin embargo, si la biblioteca lo cree oportuno, puede hacerlo directamente.

Podemos ver, pues, que las medidas adoptadas en las bibliotecas estudiadas no difieren mucho de las encontradas en las bibliotecas catalanas del CSUC. Esto nos hace pensar que los resultados de este estudio son lo suficientemente representativos de la realidad bibliotecaria española.

CONCLUSIONES

Un CRAI inclusivo es aquél en el que todos los usuarios son valorados por sus habilidades distintivas y tienen igual acceso a los recursos. El primer paso para lograrlo es conseguir que el CRAI sea accesible para que las personas con discapacidad puedan acceder a él, orientarse, utilizar sus recursos y comunicarse con el personal.

Las adaptaciones individualizadas a requerimiento del usuario no son pertinentes para un CRAI que pretenda ser el punto de referencia, ágora o corazón del campus y ser capaz de atender a cualquier usuario sin obligarlo a dar a conocer su discapacidad.

En general, la atención a los usuarios con discapacidad está más presente en la bibliografía sobre bibliotecas públicas que en la dedicada a las universitarias, como era de esperar dado el protagonismo de su función social.

La mejora en la percepción social de la discapacidad ha facilitado la incorporación de un número mayor de personas con discapacidad a las universidades, las cuales han facilitado la accesibilidad física de los edificios y han puesto en marcha programas de apoyo a los profesores que tienen que atender a estos estudiantes, dejando en segundo término los servicios universitarios entre los que se encuentran las bibliotecas.

Del estudio de casos efectuado se desprende que en la realidad bibliotecaria española:

- Las medidas de accesibilidad arquitectónica más espectaculares (puertas automáticas, franjas podotáctiles, etc.), aunque sólo se encuentran en equipamientos nuevos, cada vez son más habituales.
- Además de albergar una colección de documentos en braille, audiolibros, etc., las bibliotecas públicas se plantean cómo presentarla (un rincón, un centro de interés, etc.). En las universitarias, en cambio, esta colección es poco relevante y el debate se centra en la digitalización del fondo y quién debe hacerla (si la propia biblioteca, el usuario dentro de la biblioteca, una asociación externa, etc.).
- En las bibliotecas universitarias los equipamientos más comunes son ordenadores con software de aumento (Zoomtext) y lectura de pantalla (Jaws, NVDA) y lupas-televisión, pudiendo ubicarse en las aulas de informática, en cabinas reservadas o algún otro espacio concreto.
- Sólo una biblioteca ofrece una amplia gama de equipamientos sencillos que pueden hacer más cómoda la experiencia de los usuarios con o sin discapacidad, como lupas, atriles, etc.

- El servicio de préstamo se adapta con la prórroga del plazo, el préstamo delegado y, en algunos casos, el préstamo por correo.
- Pocas bibliotecas disponen de servicios genéricos que puedan interesar a los usuarios con discapacidades, como las reservas y la renovación en línea.

Este estudio nos debería permitir, en análisis sucesivos, realizar propuestas para conseguir unas bibliotecas universitarias plenamente inclusivas.

REFERENCIAS

- ALA (American Library Association). 2017. *ODLOS [Office for Diversity, Literacy, and Outreach Services] Glossary of Terms*.
<http://www.ala.org/aboutala/odlos-glossary-terms>
- ALA. 2020. "InserLab Project: Learning and Research Centre (CRAI), Rovira i Virgili University, Tarragona, Spain". *Best practices from world libraries photo gallery 2020*.
https://sites.google.com/view/icc-programs/2020-photo-gallery#h.p_W8vMx-VDxtTK
- Andissac, Marie-Noëlle. 2014. "Des bibliothèques accessibles et inclusives : quelques exemples de bonnes pratiques en France". *IFLA WLIC 2014 - Lyon - Libraries, Citizens, Societies: Confluence for Knowledge*. IFLA.
<http://library.ifla.org/id/eprint/965>
- Bonello, Claire. 2009. "Accessibilité et handicap en bibliothèque". Mémoire d'étude Diplôme de conservateur de bibliothèque, École nationale supérieure des sciences de l'information et des bibliothèques.
<http://www.enssib.fr/bibliotheque-numerique/document-2041>
- Carbonell, Assumpta, Eva Matamoros y Susana Peix. 2019. "L'accessibilitat a les biblioteques públiques: diversitat i inclusió social". *Item: revista de biblioteconomia i documentació* 67: 103-117.
<https://www.raco.cat/index.php/Item/article/view/362839>
- Ciuró, Gemma. 2019. "Aixopluga't a la biblioteca: La @BiblioBlanes, una biblioteca social, inclusiva i integradora". *Item: revista de biblioteconomia i documentació*, 67: 28-40.
<https://www.raco.cat/index.php/Item/article/view/362808>
- Domingo Pàmies, David. 2020. "Propostes per assolir un CRAI inclusiu: serveis accessibles per a tots els usuaris del CRAI de la Universitat Rovira i Virgili". Documento inédito. Tesis de maestría de Gestión y Dirección de Bibliotecas y Servicios de Información, Universidad de Barcelona, Facultad de Información y Medios Audiovisuales.
- Fundación Universia. 2018. *Universidad y discapacidad: IV Estudio sobre el grado de inclusión del sistema universitario español respecto de la realidad de la discapacidad*. Madrid: Fundación Universia.
https://www.fundacionuniversia.net/wp-content/uploads/2019/05/IVEstudio_UniversidadyDiscapacidad_ACC.pdf

- Gálvez Domingo, Irene y Xavier López Sánchez. 2017. "Es fácil ser inclusivo, si sabes cómo: fondo y servicios para personas con diversidad funcional en la Biblioteca Josep Janés de L'Hospitalet de Llobregat". *Boletín de la Asociación Andaluza de Bibliotecarios* 32 (114): 29-43.
<https://www.aab.es/publicaciones/boletín-aab/boletín-114/>
- Givry, Luc. 2010. "L'accessibilité des bibliothèques aux personnes handicapées", en *Bibliothèques d'aujourd'hui : à la conquête de nouveaux espaces*, editado por Marie-Françoise Bisbrouck, 272-280. Paris: Éditions du Cercle de la Librairie.
- IFLA (International Federation of Library Associations). 2015. *Access and Opportunity for All: How Libraries contribute to the United Nations 2030 Agenda*. The Hague: IFLA.
<https://www.ifla.org/publications/node/10546>
- IFLA. 2019. *Libraries Serving Persons with Print Disabilities Section*. 25 de noviembre de 2019.
<https://www.ifla.org/lpd>
- Irvall, Birgitta y Gyda Skat Nielsen. 2016. *Access to libraries for persons with disabilities: checklist*. IFLA Professional Reports, no. 89. The Hague: IFLA.
<https://www.ifla.org/files/assets/hq/publications/professional-report/89.pdf>
- Missiroli, Bélinda. 2018. "Handicap et bibliothèque universitaire : quelle accessibilité pour quel public ?" Mémoire d'étude Diplôme de conservateur de bibliothèque, École nationale supérieure des sciences de l'information et des bibliothèques.
<https://www.enssib.fr/bibliotheque-numerique/documents/68128-handicap-et-bibliotheque-universitaire-quelle-accessibilite-pour-quel-public.pdf>
- Nygaard, Knut M. 2018. "What is Universal Design : Theories, terms and trends", en *IFLA WLIC 2018 – Kuala Lumpur – Transform Libraries, Transform Societies*. IFLA.
<http://library.ifla.org/id/eprint/2250>
- ONU (Naciones Unidas). 2006. *Convención sobre los Derechos de las Personas con Discapacidad y Protocolo Facultativo*. Nueva York: Naciones Unidas.
<https://www.un.org/disabilities/documents/convention/convoptprot-s.pdf>
- Pacios, Ana R. 2015. *De la biblioteca al CRAI: una aproximación al avance del modelo en España*. Traduccions del CBUC, 56. <http://hdl.handle.et/2072/256556>
- Peix Cruz, Susana. 2014. "Bibliotecas inclusivas, bibliotecas para todos". *Boletín de la Asociación Andaluza de Bibliotecarios* 29 (107): 42-49.
<https://www.aab.es/publicaciones/boletín-aab/boletín-107/>
- Peix Cruz, Susana. 2018. "Bibliotecas inclusivas: arquitectura, servicios, fondo documental y formación se fusionan en un proyecto de accesibilidad en las bibliotecas de Cataluña", en *IFLA WLIC 2018 – Kuala Lumpur – Transform Libraries, Transform Societies*. IFLA.
<http://library.ifla.org/2263/1/127-peix-cruz-es.pdf>
- Plécard, Christine. 2010. "La Médiathèque Marguerite-Duras, Paris (20e arrondissement)", en *Bibliothèques d'aujourd'hui : à la conquête de nouveaux espaces*, editado por Marie-Françoise Bisbrouck, 219-22. Paris: Éditions du Cercle de la Librairie.
- Renaudin, Coline, Charlotte Henard y Marie-Noëlle Andissac. 2015. "Favoriser des pratiques professionnelles inclusives: La démarche de la bibliothèque de Toulouse". *Bibliothèque(s): Revue de l'Association des bibliothécaires de France* 80: 13-17.
<https://www.enssib.fr/bibliotheque-numerique/documents/67095-80-bibliotheques-et-inclusion.pdf>

- Unidiscat (Universidad y Discapacidad en Cataluña). 2019. *Universitat i discapacitat: Programes i serveis de les universitats catalanes*. Barcelona: Consell Interuniversitari de Catalunya.
http://universitatsirecerca.gencat.cat/web/.content/16_integracio_dels_estudiants_discapacitats/accions_d_integracio_dels_estudiants_discapacitats_a_la_universitat/documents/Programes-i-Serveis-de-les-universitats-catalanes-2019.pdf
- Velarte Pérez, Cristina, Jesús Faucha Pereda, Ana Pérez Moreno, Carlos Rodríguez Mahou, Margarita Sebastián Herranz, Juan Luis Marroquín, Jesús Martín Blanco, Ma. José Sánchez Lorenzo, Mercedes Hernández Navarro y Francisco Javier Martínez Calvo. 2011. *Bibliotecas accesibles para todos: Pautas para acercar las bibliotecas a las personas con discapacidad y a las personas mayores*. 2a. ed. Madrid: IMSERSO.
<http://hdl.handle.net/10421/5761>

Para citar este texto:

- Domingo Pàmies, David, Carina Rey Martín y Concepción Rodríguez Parada. 2021. “Servicios accesibles a todos los usuarios en las bibliotecas universitarias españolas: estado de la cuestión”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 33-50.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58426>

Evolution of the scholar community in the area of informetrics in Mexico: 1971-2018

María Elena Luna-Morales*
Miguel Ángel Pérez-Angón**
Evelia Luna-Morales*

Artículo recibido:
12 de diciembre de 2020

Artículo aceptado:
16 de junio de 2021

Artículo de investigación

ABSTRACT

We study the evolution of the Mexican community of researchers active in the field of informetrics studies in the period 1971-2018. Bibliometric data was retrieved from ten databases. 938 documents were registered with a total of 2121 authors affiliated to Mexican institutions. However, 42.2% of them have a scarce research production during this period. The production profiles of the full sample are organized according to the publications and citations by Mexican scholars, research lines of the authors, evolution of the number of local and foreign scholars, number

* Coordinación General de Servicios Bibliográficos, Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav-IPN), México elena.5280@gmail.com
eveorama@gmail.com

** Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, (Cinvestav-IPN), México mperez@fis.cinvestav.mx

of active and non-active scholars, and members of the National System of Researchers and scientific disciplines. Our findings indicate that a high proportion (54.5%) of the authors are members of the National System of Researchers. Our sample of authors include scholars from different areas of knowledge, and the most prolific corresponds to professionals of librarianship and information science. There is an increasing trend in their research production, specifically during the period 1990-2018, but the collaboration network is structured with a minimum set of dominant nodes.

Keywords: Informetrics-Mexico; Scientific Production and Impact; Informetrics-Graduate Studies; Researchers in Informetrics

Evolución de la comunidad académica en el área de informetría en México: 1971-2018

María Elena Luna-Morales, Miguel Ángel Pérez-Angón y Evelia Luna-Morales

RESUMEN

Se estudia la evolución de la comunidad mexicana de investigadores activos en el campo de los estudios de informetría en el periodo 1971-2018. Los datos bibliométricos se recuperaron de 10 bases de datos distintas. Se registraron 938 documentos con un total de 2 121 autores afiliados a instituciones mexicanas. Sin embargo, 42.2 % de ellos tiene una producción de investigación escasa durante este periodo. Los perfiles de producción de la muestra completa están organizados de acuerdo a las publicaciones y citas de académicos mexicanos, líneas de investigación de los autores, evolución del número de académicos locales y extranjeros, número de académicos activos y no activos, miembros del Sistema Nacional de Investigadores y disciplinas científicas. Nuestros hallazgos indican que una alta proporción de los autores (54.5 %) son miembros del Sistema Nacional de Investigadores. La muestra de autores incluye académicos de diferentes áreas del conocimiento, y la más prolífica corresponde a profesionales de la bibliotecología y las ciencias de la información. Existe una tendencia creciente en su producción investigadora, es-

pecíficamente en el periodo 1990-2018, pero la red de colaboración se estructura con un conjunto mínimo de nodos dominantes.

Palabras clave: Informetría-México; Producción e Impacto Científico; Graduados en Informetría; Investigadores-Informetría

INTRODUCTION

Informetrics has been described by Egghe and Rousseau (1990) as all metric studies related to information science, including bibliometrics, scientometrics and webometrics. This term was originally introduced by Nacke, as well as Blackert and Siegel, in 1979. Informetrics has been considered as the most complete field (Tague-Sutcliffe, 1996) in the area of knowledge associated to the new disciplines in metric studies such as informetry, patentometry and altmetrics (Peters and Bar-Ilan, 2015; Jovanovic, 2012; Hood and Wilson, 2001). The evolution of all these metric disciplines has been studied in recent years (Hérubel, 1999; Gorbea-Portal, 2016; Ball, 2017).

On the other hand, Indranil N. Sengupta (1992) has stressed that this diversity is a natural consequence of the combination of bibliometry, information and library science, science and technology: all these terms may be considered as synonyms in their reach and applications.

Informetrics studies had a steady increase after they were addressed in 1994 by the International Society for Scientometrics and informetrics and of course, other metric disciplines influenced in this increase. Informetrics has been consolidated as a mature field of research with methods and theoretical models well defined to characterize the analysis associated to libraries, information centers and the scientific activity in general (Gorbea-Portal, 2016). These studies are complemented with the use of social sciences methods and data visualization, as well as network analysis and science maps (Wolfram, 2015).

We performed a data search on the informetrics (bibliometrics, scientometrics, informetrics, webometrics, patentometrics and altmetrics) subject in the WoS databases in the last five years of the period studied. We retrieved about 10,000 files on metric studies, 1.4% of them correspond to Mexican authors. A selection of this sample corresponds to collaborations, communication and scientific policies (Hanel and Mehler, 2019; Uddin, Choudhury, and Hossain, 2019); citation analysis and Hirsch index (Mingers and Leydesdorff,

2015); bibliometric methodologies (Mandelis, 2019); evaluation of scientific journals and impact factors (Collazo-Reyes, 2014); interdisciplinary and trans-disciplinary research (Youngblood and Lahti, 2018); bibliometric indicators, use of new databases and mathematical models (Fischhoff, 2019); theory applications, data visualization, science maps through Pajack and VosViewer software (Ekanayake, Shen, and Kumaraswamy, 2019; Restrepo-Arango and Urbizagástegui-Alvarado, 2017) and finally cybermetrics, webmetrics and almetrics (Haunschild and Bornmann, 2017), which arised by the impulse of social networks. The almetrics has open the way to another type of indicators associated to the development of the Web 2.0, also known as the second generation of Web services (Peralta-González, Frías-Guzmán, and Gregorio-Chaviano, 2015). They pretend to identify the social processes reproduced in the web by the users and different civil organizations (Ayala, 2014).

In Latin America, the social studies on the scientific production were published in the 1970 decade (Krauskopf, Pessot, and Vicuña, 1986) using the regional databases of Clase and Periodica (Alonso-Gamboa, 2003). The evolution of the scientific activity was also analyzed (Pérez-Angón and Torres-Vega, 1998; Almeida-Filho et al., 2003; Herrero-Solana and Ríos-Gómez, 2006) as well as the quality of the local journals (Bonilla-Marín and Pérez-Angón, 1999).

Scientometrics published in 1995 a special volume on the Latin American studies on science and technology (Cortés-Vargas, 2007). In this volume we can appreciate the consolidation of the research groups from Argentina, Brazil and Mexico. The first Mexican publication on informetrics was published in 1971 by Jorge Robles Glenn from the National Autonomous University of Mexico (UNAM). In particular, the performance of some Mexican scientific communities has been analyzed in the areas of ecology (Castillo, 2000), physics (Contreras-Gómez et al., 2015), engineering (Rodríguez-Miramontes and González-Brambila, 2016), medicine (William, 2001), agriculture (Duarte-Malanski, Schiavi, and Dedieu, 2019) and social sciences (Gil-Antón and Contreras-Gómez, 2017; Contreras-Gómez et al., 2020).

The general aim of the present work is to study the evolution of the Mexican community active in research in the field of informetrics studies. We have studied recently (Luna-Morales, Luna-Morales, and Pérez-Angón, 2021) the effect of local and international collaborations on the research production of this community. Our interest in the present work is to characterize this set of scholars according to their institutional affiliation, their geographical distribution and academic formation. We have identified the authors

of 938 documents registered in 10 databases, which include Web of Science (WoS) and Scopus. A high fraction of them (54.5%) is member of the National System of Researchers. Our findings indicate that there is a positive trend of the research production in this area of knowledge, with a high degree of collaboration but that the respective research network is characterized by very few (less than ten) dominant nodes. Accordingly, our research questions can be summarized in two points:

1. How much mature is the research community of Mexican scholars involved in the area of informetrics studies?
2. What are the perspective of this community in terms of the training of new researchers and the generation of strong research networks?

METHODS

We were able to build a bibliometrics database using ten sources of information: Web of Science (WoS) in all Databases (Web of Science Core Collection, Derwent Innovations Index, KCI-Korean Journal Database, Russian Science Citation Index, SciELO Citation Index), Scopus, Google Scholar, Clase, Periodica, Humanindex, Infobila, PubMed, Scielo (<https://scielo.org/es/>) and Latindex. We expected to get a stronger data sample by including also local and regional databases in spite the diversity of fields they include and the normalization they require. We retrieved bibliometric data for the period 1971-2018 in the April-June 2019 window in such a way the our 2018 data was already complete. According to our data search, the first paper on informetrics studies was published by Robles-Glenn (1971). Our citation search was restricted only to WoS and Scopus. WoS included already a portion of the Scielo collection. The other databases do not include reliable citations data yet.

We have included in our data search the features “Topics” in the case of WoS, while for Scopus we used the search feature “all fields”. On the other hand, we used the advanced search in all other cases with two options: by fields and Boolean operators. In particular, for Scholar we saved the data in CVS.

The search strategy that was applied to retrieve publications in informetrics is shown in *Figure 1*, in both English and Spanish, both in singular and plural.

SPANISH TERMS	ENGLISH TERMS	SPANISH TERMS	ENGLISH TERMS
Acoplamiento bibliográfico	Bibliographic coupling	Índice H	Index H
Altimetría	Altimetrics	Informetría	Informetrics
Análisis bibliométrico	Bibliometric analysis	Investigación bibliotecológica	Library research
Arquivometría	Archivometrics	Journal Citation Reports	Journal Citation Reports
Bibliometría	Bibliometrics	Ley de Bradford	Bradford's Law
Ciencia de los datos	Data science	Ley de Lotka	Lotka's Law
Cientimetría	Scientometrics	Ley de Price	Price's Law
Co-citación	Co-citation	Ley de Zipf	Zipf's Law
Coefficiente de colaboración	Collaboration coefficient	Mapas científicos	Scientific maps
Colaboración científica	Scientific collaboration	Obsolescencia de la literatura	Obsolescence of literature
Colecciones, estudios	Collections, studies	Patentometría	Patentometrics
Colegios invisibles	Invisible schools	Producción científica	Scientific production
Comunicación científica	Scientific communication	Producción e impacto científico	Production and scientific impact
Crecimiento de la literatura	Literature growth	Producción e impacto redes autores	Production and impact on author networks
Estudios Métricos	Metric Studies	Redes bibliométricas	Bibliometric networks
Evaluación científica	Scientific evaluation	Redes de Colaboración	Collaboration Networks
Factor de impacto	Impact factor	Scientimetría	Scientometrics
Impacto científico	Scientific impact	Tecnometría	Technometrics
Indicadores bibliométricos	Bibliometric indicators	Vida media	Half life
Indicadores de la ciencia y tecnología	Science and technology indicators	Webmetría	Webmetrics

Figure 1. Terminology applied in the search for research publications in Mexico

Our search strategy was completed by identifying journals included in JCR-2018 that published articles on informetrics subjects, in particular those in the subject areas Information science and library science. We also included PLoS ONE and Interciencia. In both cases, we carried out direct searches in their web sites on information disciplines published by Mexican authors. The respective data was registered in Excel with the same WoS data structure.

In order to avoid duplicated papers, we used specific matches for author's names, titles, publication years and bibliographical data. Finally, our search strategy retrieved 938 publications, 542 of them written in Spanish, 381 in English, 11 in Portuguese and 4 in French.

WoS and Scopus data files were retrieved in Excel and CVS formats. In all other cases, the respective files were registered directly and with the same structure used for the WoS and Scopus files. We examined directly each file in order to keep just those associated to the subject area of informetrics studies. We went through the usual process of Normalization of authors names, institutions, scientific disciplines, journal titles and Federal entities. Finally, our data sample was arranged according to the ten general disciplines promoted by the project Atlas de la Ciencia Mexicana (ACM, 2013): biology, physics, chemistry, mathematics, medicine, engineering, humanities, social sciences, geosciences, and agrosociences.

These disciplines were classified according to the subject areas considered in the (SC) field of WoS. Our search was performed in the process of

selecting members and non-members of NSR in our data sample. This selection used a direct search in the NRS databases and the web pages of their respective institutions. This strategy was used also to identify the author's research areas. *Table 1* includes the indicators used in this search strategy. It was necessary to normalize the data on each institution before including it in an Access database.

Indicators	Description	General objective
Production	Papers on metric studies	Identification of the production by authors, institutions, and federal entity
	Papers by research subject	Identification of the production by research subject / Papers by scientific area (ACM)
	Papers by scientific area (ACM)	Identification by scientific area
Citation impact	Number of citations by paper	Impact factor by author, institution and federal entity
	Number of citations by subject area	Impact factor by subject area
	Number of citations by scientific discipline (ACM)	Impact factor by scientific discipline
NSR/authors	Identification of NSR members	Determination of active/non-active NSR members Determination of active members by NSR level and academic area
	Identification of addresses of authors	Determination of NSR members by federal entity
	Identification of the NSR academic area of authors	Determination of the production by academic area

Table 1. Indicators used to retrieve bibliometric data

Finally, in order to get a well-organized data sample, it was necessary to tolerate double counting with integer values in the fields of name authors and institutions, as well in subject areas and scientific disciplines. We used TL.exe software for co-word mappings of texts (<https://www.leydesdorff.net/>). The author network was generated with Pajek software (<http://mrvar.fdv.uni-lj.si/pajek/>).

RESULTS

The databases of WoS, Scopus and Scholar were pioneer tools in the bibliometric studies published by Mexican scholars. However, some of the first published works in area of knowledge used Clase, Periodica and Latindex (Luna-Morales, Luna-Morales, and Pérez-Angón, 2021). They constituted the first databases with complete bibliometric information in Latin America.

In *Figure 2* we used the whole sample (938 documents) to depict the evolution of the number of documents, and the respective generated citations, published in the area of informetrics studies in the period 1971-2018. There is a scarce production during the first three decades with a positive trend in the most recent period 2000-2018. The number of citations is appreciable since 1990 and it is possible to identify five years with an impressive number of citations: 402 (2007), 281 (2014), 264 (2012), 244 (2013) and 210 (2011). They correspond mostly to four papers published in *Scientometrics*, *Inter-ciencia*, *Journal of the American Society for Information Science and Technology (JASIST)* and *Research Policy* as it is shown in *Appendix* with the list of the papers that generated the larger number of citations in this period.

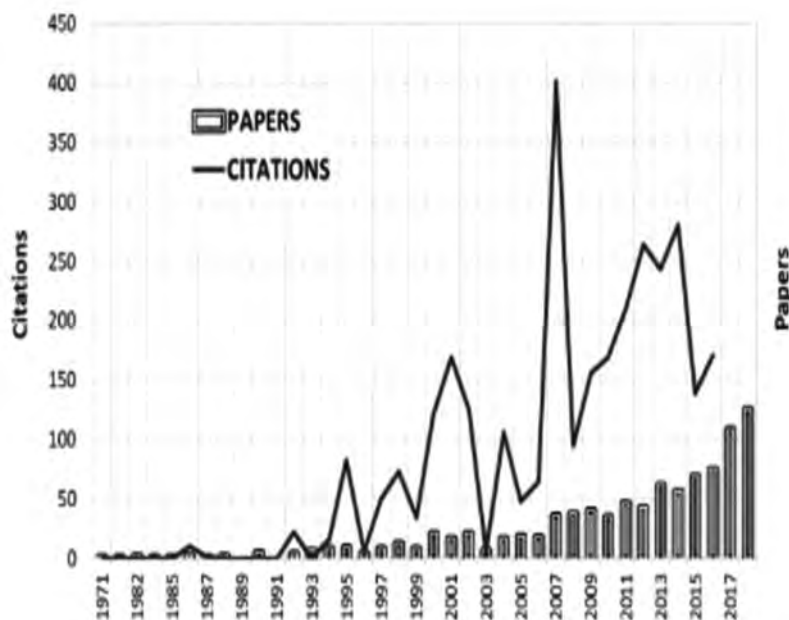


Figure 2. Evolution of the number of publications and citations by Mexican scholars in the field of informetrics (1971-2018)

In *Table 2* we have included the list of the more prolific authors in the field of metric studies of science and technology in Mexico. We were able to identify 1393 active authors in this area of knowledge. There are scholars with different research interests but the most active set of scholars in this field correspond to the area of librarianship.

Table 2 included the 41 most active authors with a minimum of five publications in mainstream journals. Their production corresponds to 22.7% of the total production in this area of knowledge. This set of authors are affiliated to 14 Mexican institutions which also reflects the fact that field of metric studies of science and technology is a very young area of research. There is other group of 187 scholars that published only 2 to 5 papers with 22.2% of the total research production. It is important to notice that the rest of the research production (55.1%) is associated to 1175 scholars with just one publication.

Num.	Authors	Institutions	Papers	Author's research áreas
1	Russell-Barnard, Jane Margaret	UNAM	56	Librarianship
2	Licea de Arenas, Judith	UNAM	46	Librarianship
3	Collazo-Reyes, Francisco	Cinvestav	33	Librarianship
4	Luna-Morales, Maria Elena	Cinvestav	28	Librarianship
5	Macias-Chapula, Cesar Augusto	Secretaría de Salud	24	Librarianship
6	Gorbea-Portal, Salvador	UNAM	23	Librarianship
7	Gonzalez-Brambila, Claudia Noemi	ITAM	21	Engineering
8	Aguado-Lopez, Eduardo	UAEMex	17	Sociology
9	Perez-Angon, Miguel Angel	Cinvestav	17	Physics
10	Valles-Valencia, Javier	UNAM	16	Librarianship
11	Arenas-Vargas, Miguel Angel	UAM	15	Biology
12	Michan-Aguirre, Layla	UNAM	15	Biology
13	Del Rio, Jose Antonio	UNAM	13	Physics
14	Musi-Lechuga, Bertha	UACJ	13	Librarianship
15	Olivas-Avila, Jose Alonso	UACJ	13	Medicine
16	Alonso-Gamboa, Jose Octavio	UNAM	12	Librarianship
17	Carrillo-Calvet, Humberto Andres	UNAM	12	Engineering

18	Tarango-Ortiz, Javier	UACH	11	Librarianship
19	Luna-Morales, Evelia	Cinvestav	10	Librarianship
20	Rogel-Salazar, Rosario	UAEMex	10	Sociology
21	Narvaez-Bertheleot, Nora	UNAM	10	Librarianship
22	Cortes, Hector Daniel	UNAM	9	Physics
23	Ainsworth, Shirley	UNAM	7	Librarianship
24	Cocho, Germinal	UNAM	7	Physics
25	Fuentes-Navarro, Raúl	ITESO	7	Sociology
26	Machin-Mastromatteo, Juan D.	UACH	7	Studies Information
27	Restrepo-Arango, Cristina	ColMex	7	Librarianship
28	Rodríguez-Salvador, Marisela	ITESM	7	Engineering
29	Cetto-Kramis, Ana María	UNAM	7	Physics
30	Becerril-García, Arianna	UAEMex	6	Engineering
31	Roldan-Valadez, Ernesto	Secretaría de Salud	6	Medicine
32	Ayala-Picazo, Micaela	ColMex	5	Librarianship
33	Cantú-Ortiz, Francisco J.	ITESM	5	Engineering
34	García-Mandujano, Esther Ofilia	UNAM	5	Physics
35	García-Gómez, Francisco	IMSS	5	Medicine
36	Gonzalez, Eric	UNAM	5	Librarianship
37	Lieberman, Sofia	UNAM	5	Psychology
38	Mercado-Martínez, Francisco Javier	UDG	5	Medicine
39	Miramontes, Pedro	UNAM	5	Mathematics
40	Reyna-Espinosa, Felipe Rafael	UNAM	5	Librarianship
41	Ríos-Castañeda, Camilo	Instituto Nacional de Neurología y Neurocirugía	5	Medicine

Table 2. List of the most productive authors by institution and research area

Figure 3 also shows that the most recent increase in the Mexican output can be associated with a large number of new journals registered in the WoS and Scopus databases (Collazo-Reyes, 2014), as well as to an increase in the number of scholars active in this research field.

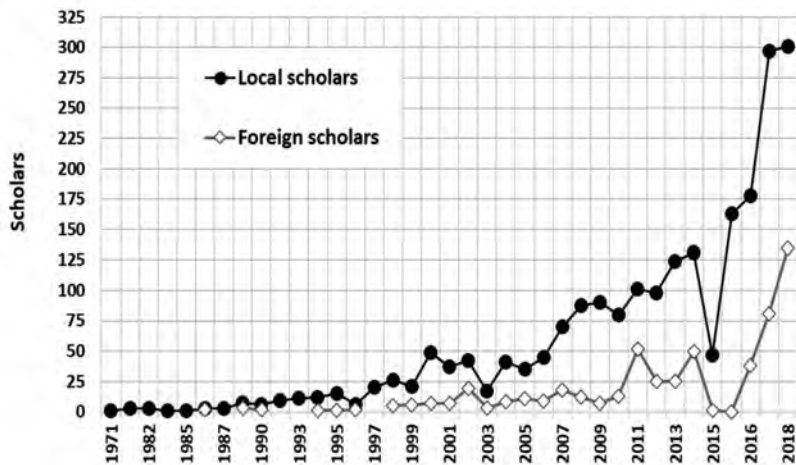


Figure 3. Evolution and comparison of the number of local and foreign scholars in the field of informetrics (1971-2018)

The same figure shows the evolution of the number of local vs. foreign authors in our data sample. It is impressive the dynamics of the number of local authors in this field of research, with an increasing factor of two or three in recent years. However, we found that 42.2% of the whole set of local authors (951) has a scarce research production with just one or two papers published in the analyzed period. We call inactive authors to the scholars identified with this very low research production (Figure 4).

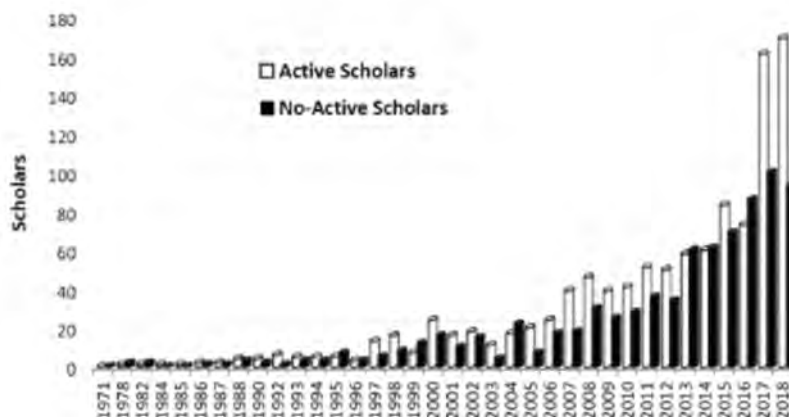


Figure 4. Evolution of the number of active and no-active scholars in informetrics (1971-2018)

In *Figure 5* we present the evolution of the number of authors that are members of the National System of Researchers. We were able to identify a relatively low percentage (54.1%) of NSR members in our data sample, which contribute with 43.7% of the whole production and with 43.4% of the respective number of citations (*Table 2*).

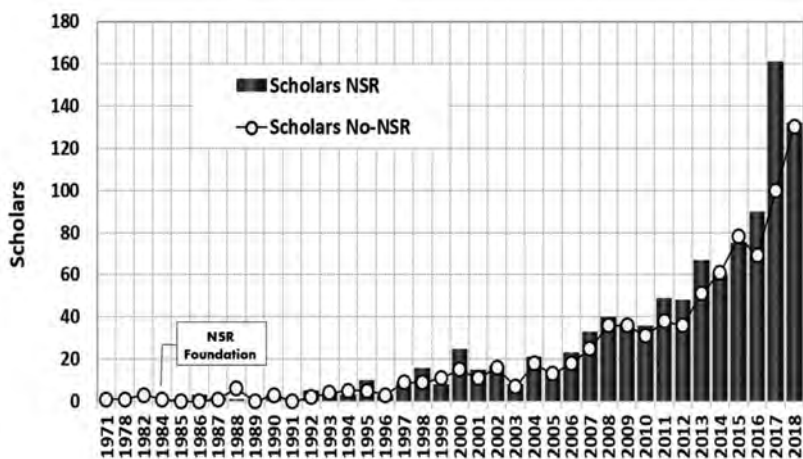


Figure 5. Evolution of the number of scholars which are members or no-members of the National System Research (NSR) in the field of informetrics (1971-2018)

These results are consistent with similar percentages obtained by NSR members in the area of social sciences with respect to the whole research production in this area of knowledge (Contreras-Gómez et al., 2020). We can appreciate that the number of NSR members in our data sample had a strong increase in period 2016-2018. However, this group of NSR authors do not have yet a strong contribution on the whole sample of documents published in the area of informetrics studies. Since NSR was funded in 1984, their first papers appeared in 1985 (*Table 3*). It is important to notice that there is double counting in *Table 2* due to the large number of documents published in collaboration by NSR members and no-members.

Members NSR				No-Members NSR			
Years	Num. Scholars	Num. Papers	Num. Citations	Years	Num. Scholars	Num. Papers	Num. Citations
1971-1975	0	0	0	1971-1975	1	1	1
1976-1980	0	0	0	1976-1980	1	1	0
1981-1985	1	1	0	1981-1985	4	4	0

1986-1990	9	9	11	1986-1990	10	10	43
1991-1995	22	30	93	1991-1995	16	17	71
1996-2000	61	73	257	1996-2000	47	49	77
2001-2005	75	95	494	2001-2005	65	73	241
2006-2010	169	201	733	2006-2010	146	161	651
2011-2015	299	330	873	2011-2015	264	287	778
2016-2018	386	455	295	2016-2018	299	325	256
TOTAL	1022	1194	2756	TOTAL	853	928	2118

Table 3. Relation of the number of publications and citations by members and no-members of NSR in the field of informetrics (1971-2018)

In *Table 4* we included the distribution of the number of documents registered in our data sample according to author's research areas. The most frequent areas included in our data sample correspond to librarianship and medicine, with 27.2% (17.6%) and 21.6% (28.7%) of the documents (citations), respectively. In the following places we found also that engineering and physics have a substantial contribution to our data sample, with 7.4% (10.0%) and 5.4% (10.8%) of the documents (citations).

The large number of documents published by scholars with specialties different from librarianship is associated to their interest in the development of their own research communities, and to the interest in applying some methodologies generated by members of their communities like network analysis, visualization systems and mining data.

Num.	Authors' research areas	Papers	Citations	% Papers	% Citations
1	Librarianship	555	1564	27.2	17.6
2	Medicine	441	2557	21.6	28.7
3	Engineering	151	894	7.4	10.0
4	Physics	110	962	5.4	10.8
5	Biology	76	424	3.7	4.8
6	Psicology	51	154	2.5	1.7
7	Sociology	51	143	2.5	1.6
8	Mathematics	49	383	2.4	4.3
9	Education	48	62	2.4	0.7
10	Economy	47	217	2.3	2.4
11	Ecology	43	90	2.1	1.0

12	Social Studies	43	124	2.1	1.4
13	Administration	42	160	2.1	1.8
14	Agronomy	39	102	1.9	1.1
15	Chemistry	29	223	1.4	2.5
16	Nursing	27	18	1.3	0.2
17	Veterinary and zootechnics	24	9	1.2	0.1
18	Artificial intelligence	21	169	1.0	1.9
19	Anthropology	16	133	0.8	1.5
20	Vegetal biology	16	107	0.8	1.2
21	Science Studies	15	32	0.7	0.4
22	Aquaculture	14	18	0.7	0.2
23	Social communication	13	31	0.6	0.3
24	Ecosystems	13	6	0.6	0.1
25	Astronomy	12	4	0.6	0.0
26	Technologies	11	79	0.5	0.9
27	Geophysics	11	64	0.5	0.7
28	Political Sciences	11	5	0.5	0.1
29	Journal editors	11	5	0.5	0.1
30	Manufacture	10	73	0.5	0.8
31	Design	10	43	0.5	0.5
32	Scientific communication	10	28	0.5	0.3
33	Science, Technology and Society	10	11	0.5	0.1
34	Pedagogy	10	2	0.5	0.0

Table 4. Distribution of the research areas involved in the production of Mexican scholars in the field of informetrics (1971-2018)

In *Table 5*, there are some scientific disciplines, like physics, medicine and engineering, that have higher percentages of citations as compared with their respective contributions to the number of publications and scholars.

The case of humanities deserves a special comment. This area of knowledge has contributed to the Mexican scientific production with a low number of publications and citations (ACM, 2013). Some of the leading researchers in librarianship are members of NSR in the área V (humanities); as a consequence, they have contributed with an extraordinary number of publications and citations.

Num. Scholars	Scientific Discipline	Papers	Cita-tions	% Papers	% Citations
8	Earth sciences	8	17	0.3	0.2
11	Humanities	12	151	0.4	1.7
21	Chemical sciences	29	223	1.1	2.5
40	Agriculture	43	122	1.6	1.4
41	Mathematics	50	291	1.9	3.3
102	Physical sciences	172	1135	6.4	12.9
190	Biological Sciences	212	512	7.9	5.8
222	Engineering	315	1289	11.8	14.6
551	Medicine and Health Sciences	641	2595	24.0	29.5
661	Social Sciences and Behavioral Sciences	1193	2474	44.6	28.1
TOTAL		2675	8809	100.0	100.0

Table 5. Distributions of the number of scholars, publications and citations, according to the scientific disciplines (1971-2018)

The geographical distribution of authors is presented in *Tables 6* and *7*. There is a very high concentration of scholars in Mexico City (CDMX): 52.2% that produce 59.4% (55.3%) of the published documents (citations). Only four of the other federal entities (Jalisco, Edo. de México, Nuevo León, Morelos) have percentages higher than 4%. This is a rather dramatic concentration of the research activity in CDMX which has been also observed by other studies in social sciences and humanities (Contreras-Gómez et al., 2020). The same situation is reflected in *Table 7* for the authors data in terms of NSR membership.

Geographical entity	Num. Scholars	% Scholars	Papers	% Papers	Citations	% Citations
CDMX	724	52.2	1257	59.4	3913	55.3
Jalisco	91	6.6	106	5.0	538	7.6
Estado de México	82	5.9	107	5.1	160	2.3
Nuevo León	58	4.2	80	3.8	477	6.7

Morelos	57	4.1	101	4.8	559	7.9
Baja California	50	3.6	52	2.5	157	2.2
Puebla	49	3.5	63	3.0	140	2.0
Chihuahua	37	2.7	88	4.2	404	5.7
Tamaulipas	24	1.7	26	1.2	22	0.3
Veracruz	22	1.6	25	1.2	85	1.2
Yucatán	21	1.5	21	1.0	27	0.4
Michoacán	20	1.4	24	1.1	308	4.3
Guanajuato	17	1.2	17	0.8	39	0.6
Sinaloa	17	1.2	17	0.8	28	0.4
Sonora	16	1.2	18	0.9	27	0.4
Querétaro	15	1.1	15	0.7	34	0.5
San Luis Potosí	12	0.9	18	0.9	17	0.2
Tabasco	10	0.7	10	0.5	3	0.0
Zacatecas	10	0.7	16	0.8	19	0.3
Chiapas	9	0.6	9	0.4	2	0.0
Colima	9	0.6	9	0.4	22	0.3
Coahuila	7	0.5	7	0.3	6	0.1
Hidalgo	7	0.5	7	0.3	14	0.2
Durango	5	0.4	5	0.2	29	0.4
Quintana Roo	4	0.3	4	0.2	24	0.3
Tlaxcala	3	0.2	4	0.2	19	0.3
Nayarit	3	0.2	3	0.1	1	0.0
Aguascalientes	2	0.1	2	0.1	3	0.0
Baja California Sur	2	0.1	2	0.1	0	0.0
Campeche	2	0.1	2	0.1	1	0.0
Guerrero	2	0.1	2	0.1	3	0.0
TOTAL	1387	100	2117	100	7081	100

Table 6. Distribution of the scholars by geographical entity in the field of metric studies (1971-2018)

Members NSR			No-Members NSR		
Geographical entity	Papers	% Papers	Geographical entity	Papers	% Papers
CDMX	672	56.4	CDMX	585	63.2
Estado de México	63	5.3	Jalisco	47	5.1
Nuevo León	61	5.1	Morelos	43	4.6
Jalisco	60	5.0	Chihuahua	40	4.3
Baja California	50	4.2	Puebla	38	4.1
Chihuahua	47	3.9	Estado de México	33	3.6
Puebla	45	3.8	Nuevo León	20	2.2
Morelos	32	2.7	Baja California	18	1.9
Michoacán	16	1.3	Guanajuato	10	1.1
Veracruz	15	1.3	Querétaro	10	1.1
Yucatán	15	1.3	Tamaulipas	10	1.1
Querétaro	13	1.1	Veracruz	9	1.0
San Luis Potosí	12	1.0	Sinaloa	8	0.9
Tabasco	12	1.0	Tabasco	8	0.9
Tamaulipas	12	1.0	Michoacán	5	0.5
Sonora	11	0.9	San Luis Potosí	5	0.5
Sinaloa	10	0.8	Zacatecas	5	0.5
Zacatecas	10	0.8	Chiapas	4	0.4
Chiapas	6	0.5	Yucatán	4	0.4
Coahuila	5	0.4	Durango	3	0.3
Hidalgo	5	0.4	Hidalgo	3	0.3
Colima	4	0.3	Nayarit	3	0.3
Guanajuato	4	0.3	Quintana Roo	3	0.3
Aguascalientes	2	0.2	Campeche	2	0.2
Campeche	2	0.2	Coahuila	2	0.2
Durango	2	0.2	Colima	2	0.2
Quintana Roo	2	0.2	Guerrero	2	0.2
Tlaxcala	2	0.2	Sonora	2	0.2
Baja California Sur	1	0.1	Aguascalientes	1	0.1

			Baja California Sur	1	0.1
TOTAL	1191	100	TOTAL	926	100

Table 7. Geographical distribution of the number of publications by NSR members in the field of metric studies (1971-2018)

In *Figure 6* we have depicted the collaboration network obtained from our data sample in the informetrics studies. We can appreciate two main sub-networks associated to Jane Margaret Russell Barnard and Judith Licea de Arenas, two of the pioneers in this research field. They are affiliated to the National University of Mexico (UNAM). Most of their collaborations shown in *Figure 6* involve their own graduate students. Another aspect of interest in *Figure 6* is the very low (less than ten) dominant nodes, which induce a large vulnerability of the network. We used the Leydesdorff and Pajack softwares in order to construct this collaboration network.

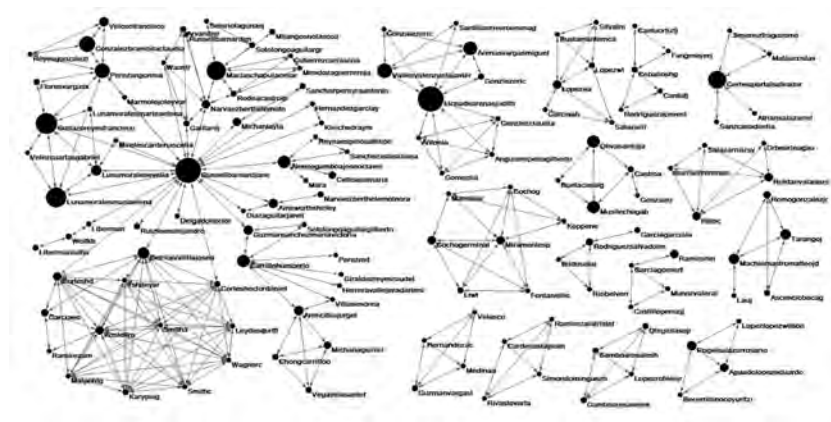


Figure 6. Co-authorship network of Mexican scholars in the field of informetrics (1971-2018)

There are eight training institutions with graduate programs in the field of informetrics (*Table 8*). Most of these programas require presencial participation of students. UNAM was the pioneer institution in these graduate programs in Mexico with both master and PhD programs. For this reason, three of the dominant nodes in the collaboration network shown in *Figure 6* correspond to UNAM faculty members (Russell Barnard, Licea de Arenas, and Gorbea Portal). It has been pointed out also that the training of new researchers has a positive effect on the development of strong research groups.

Num.	Institutions	Graduate programs	Begin Year	Learning Unit Character
1	Universidad Nacional Autónoma de México (UNAM)	Master in librarianship and Information Science	1998	Virtual
		Master in librarianship and Information Science	2000	Presential
		PhD in librarianship and Information Science	2006	Virtual
2	Universidad Autónoma Metropolitana (UAM-X) – Universidad de la Habana – Universidad de Murcia, España	Master/PhD in Information Magnagement	2006	Virtual
3	Instituto Tecnológico de Estudios Superiores Monterrey (Campus Hidalgo)	Master, Information Science and Knowledge Management	1999	Virtual
4	Centro de Investigación y de Estudios Avanzados (Cinvestav)	Transdisciplinary PhD, Science, Technology and Society	2009	Presential
5	El Colegio de México (COLMEX)	Master in librarianship	2010	Presential
6	Universidad Autónoma de Ciudad Juárez (UACJ)	Master/PhD in Information Magnagement	2010	Virtual
7	Universidad Autónoma de Chihuahua (UACH)	Master, librarianship and Information Science	2013	Presential
8	Universidad Autónoma de San Luis Potosí (UASLP)	Master in Information Science and Documentation	2015	Presential

Table 8. Relation of graduate programs in the field of informetrics

DISCUSSION AND CONCLUDING REMARKS

The present analysis of the research production in informetrics in Mexico finds a young community of scholars with about 40 years of activity. This community originated in the first training programs on bibliothecology and information science. UNAM was pioneer in both training programs and research projects. Our analysis identified a strong research network of collaboration but with rather few dominant nodes (Luna-Morales, Luna-Morales, and Pérez-Angón, 2021). We were able to identify five scholars which had a clear influence in the development of this community: Jane

Margaret Russell Barnard y Salvador Gorbea-Portal (Instituto de Investigaciones Bibliotecológicas, UNAM), Judith Licea de Arenas (Facultad de Filosofía y Letras, UNAM), and Eduardo Aguado López and Rosario Rogel Salazar (Universidad Autónoma del Estado de México).

While most of the research production in this area of knowledge is published in mainstream journals involving collaborations of several authors (Dorta-González and Santana-Jiménez, 2019), in the Mexican case most of the research production is published in regional journals in Spanish with few international collaborations, with just 17% of the publications involving collaborations with authors beyond the Latin American region (Luna-Morales, Luna-Morales, and Pérez-Angón, 2021). This fact is also a consequence of the large number of local authors in the production of this area of knowledge.

The maturity of a research community depends on the effective implementation of several factors: academic and experimental infrastructure, active leadership, collaborative work and the formation of new researchers through qualified graduate programs (Durand-Villalobos, 2017). Even though the field of informetrics in Latin America and Mexico has been active for about 50 years (Kreimer and Vessuri, 2018), we could identify a positive trend of the Mexican research production starting the period 2000-2018.

Our findings indicate that this research trend is a consequence of three facts: (1) an increasing number of institutions with undergraduate programs in librarianship; (2) an increase in the number of scholars involved in this area of knowledge, which were trained mostly in foreign institutions; (3) the consolidation of several research groups with experienced members of NSR. The formal requirement established for NSR members to maintain a continuous research activity has also contributed to the steady increase in the research production in the field of metric studies of science and technology.

In the present work, we have characterized the scholar community that is active in research in the field of informetrics in Mexico. Its research output was retrieved from ten bibliometric databases. Our findings indicate a positive trend in this research production just in the period 2000-2018 in both in published documents and the respective number of generated citations. The scholars working in the area of librarianship have the largest contribution in the number of published documents (27.2%) but not in the number of generated citations (17.6%). The scholars working in the area of medicine had the largest fraction of generated citations (28.7%) with a lower number of published papers (21.6%). It is natural that the librarianship community is the most active in this research area since its work is directly related to the metric studies of science and technology. Salvador Gorbea-Portal (2013) has pointed out that metric studies of the scientific information constitute a

new research field: librarianship. It is a multidisciplinary discipline that has involved scholars with different research interests, like engineers, physicists, mathematicians and others.

This area of knowledge has still an intense concentration of research activity in the Mexico City metropolitan area: 52.2% of the researchers, 59.4% of the Mexican production and 55.3% of the total number of citations.

On the other hand, there a substantial number of NSR members in our data sample (54.5%), the active researchers in this area of knowledge represent a healthy variety of scientific disciplines. We were able to identify a strong network of collaboration among the members of this community. However, there is weak point in the structure of this network: there is a very few dominant nodes (less then ten).

We would like to stress that, as far we know, the present work is the first study that addresses the evolution and characterization of a local community of scholars in Latin America involved in research and training in the area of informetrics studies. Finally, we hope that the present study will be continued in future work by taking into account the following pints: analysis of the performance of new researchers coming out from the seven graduate programs identified in the present work; and the age distribution of the members of this community in order to determine the perspective of new research groups located outside the Mexico City metropolitan area. It will be interesting also to search for the production in other databases and journals in this research field which are not included in the databases analyzed.

Acknowledgments

We appreciate the support of SNI and Conacyt (proyect A1-S-9013).

REFERENCES

- ACM (Atlas de la Ciencia Mexicana). 2013. *Indicadores por área del conocimiento 1980-2013*. México: Academia Mexicana de Ciencias.
<http://atlasdelacienciamexicana.org/es/index-es.shtml>
- Almeida-Filho, Naomar, Ichiro Kawachi, Alberto Pellegrini-Filho, and J. Norberto W. Dachs. 2003. "Research on health inequalities in Latin America and the Caribbean: bibliometric analysis (1971-2000) and descriptive content analysis (1971-1995)." *American Journal of Public Health* 93 (12): 2037-2043.
<https://doi.org/10.2105/ajph.93.12.2037>
- Alonso-Gamboa, José Octavio. 2003. "Selección de revistas latinoamericanas en bases de datos: criterios utilizados en Clase y Periódica." *Biblioteca Universitaria* 6 (1): 9-21.
<https://www.redalyc.org/pdf/285/28560103.pdf>

- Ayala, Teresa. 2014. "Redes sociales, poder y participación ciudadana." *Revista Austral de Ciencias Sociales* 26: 23-48.
<https://www.redalyc.org/pdf/459/45931862002.pdf>
- Ball, Rafael. 2017. *An Introduction to Bibliometrics*. Germany: Elsevier, Chandos Publishing.
- Bonilla-Marín, Marcial and Miguel Ángel Pérez-Angón. 1999. "Revistas mexicanas de investigación científica y tecnológica." *Interciencia* 24 (2): 102-106.
- Castillo, Alicia. 2000. "Communication and utilization of science in developing countries." *Science Communication* 22 (11): 46-72.
<https://doi.org/10.1177/1075547000022001004>
- Collazo-Reyes, Francisco. 2014. "Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country." *Scientometrics* 98 (1): 197-209.
<https://doi.org/10.1007/s11192-013-1036-2>
- Contreras-Gómez, Leobardo Eduardo, José Luis Olivares-Vázquez, Guadalupe Palacios-Núñez, Rafael Masrmolejo-Leyva, Claudia Noemí González Brambila, Miguel Ángel Pérez Angón, and Manuel Gil-Antón. 2020. "Desconcentración del Sistema Nacional de Investigadores: geografía y estratificación. El caso de las ciencias sociales (2020-2018)." *Revista de la Educación Superior* 49 (193): 83-106.
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-27602020000100083
- Contreras-Gómez, Leobardo Eduardo, Rafael Baquero-Parra, Eduardo Robles-Belmont, and Miguel Ángel Pérez-Angón. 2015. "Patrones de movilidad de los físicos mexicanos en el Sistema Nacional de Investigadores." *Interciencia* 40 (8): 525-532.
- Cortés-Vargas, Daniel. 2007. "Medir la producción científica de los investigadores universitarios: la bibliometría y sus límites." *Revista de la Educación Superior* 36 (142): 43-65.
<http://www.scielo.org.mx/pdf/resu/v36n142/v36n142a3.pdf>
- Dorta-González, Pablo and Yolanda Santana-Jiménez. 2019. "Characterizing the highly cited articles: A large-scale bibliometric analysis of the top 1% most cited research." *Malaysian Journal of Library & Information Science* 24 (2): 23-39.
- Duarte-Malanski, Priscila, Sandra Schiavi, and Benoit Dedieu. 2019. "Characteristics of work in agriculture scientific communities. A bibliometric review." *Agronomy for Sustainable Development* 39 (36).
<https://doi.org/10.1007/s13593-019-0582-2>
- Durand-Villalobos, Juan Pablo. 2017. "Grupos de investigación consolidados en la Universidad de Sonora." *Revista Mexicana de Investigación Educativa* 22 (75): 1143-1167.
- Egghe, Leo and Ronald Rousseau. 1990. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. Amsterdam: Elsevier.
- Ekanayake, E. M. A. C., Geoffrey Shen, and Mohan M. Kumaraswamy. 2019. "Mapping the knowledge domains of value management: a bibliometric approach." *Engineering Construction and Architectural Management* 26 (3): 499-514.
<https://doi.org/10.1108/ECAM-06-2018-0252>
- Fischhoff, Baruch. 2019. "Evaluating science communication." *Proceedings of the National Academy of Sciences of the United States of America* 116 (16): 7670-7675.

- Gil-Antón, Manuel and Leobardo Eduardo Contreras-Gómez. 2017. "The National System of Researchers: ¿mirror or model?" *Revista de la Educación Superior* 46 (184): 1-19.
- Gorbea-Portal, Salvador. 2013. "Tendencias transdisciplinarias en los estudios métricos de la información y su relación con la gestión de la información y del conocimiento." *Perspectivas em Gestão & Conhecimento, João Pessoa* 3 (1): 13-27.
- Gorbea-Portal, Salvador. 2016. "Una nueva perspectiva teórica de la bibliometría basada en su dimensión histórica y sus referentes temporales." *Investigación Bibliotecológica* 30 (70): 11-16.
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2016000300011
- Hanel, Paul H. P. and David M. A. Mehler. 2019. "Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication." *Public Understanding of Science* 28 (4): 468-485.
<https://doi.org/10.1177/0963662519834193>
- Haunschild, Robin and Lutz Bornmann. 2017. "How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data." *Scientometrics* 110 (3): 1209-1216.
- Herrero-Solana, Víctor and Claudia Ríos-Gómez. 2006. "Producción latinoamericana en biblioteconomía y documentación en el Social Science Citation Index (SSCI) 1966-2003." *Information Research* 11 (2): 247-264.
- Hérubel, Jean Pierre V. M. 1999. "Historical Bibliometrics: Its Purpose and Significance to the History of Disciplines." *Libraries and Culture* 34 (4): 380-388.
- Hood, William W. and Conception S. Wilson. 2001. "The literature of bibliometrics, scientometrics, and informetrics." *Scientometrics* 52: 290-303.
<https://doi.org/10.1023/A:1017919924342>
- Jovanovic, Milos. 2012. "A short history of early bibliometrics." *NFD Information-Wissenschaft und Praxis* 63 (2): 71-80.
- Krauskopf, M., R. Pessot, and R. Vicuña. 1986. "Science in Latin America how much and along what lines?" *Scientometrics* 10 (3-4): 199-206.
<https://doi.org/10.1007/BF02026041>
- Kreimer, Pablo and Hebe Vessuri. 2018. "Latin American science, technology and society: a historical and reflexible approach." *Tapuya: Latin American Science, Technology and Society* 1 (1): 17-37.
<https://doi.org/10.1080/25729861.2017.1368622>
- Luna-Morales, María Elena, Evelia Luna-Morales, and Miguel Ángel Pérez-Angón. 2021. "Influence of the international collaboration in the field of metric studies of science and technology: the case of Mexico (1971-2018)." *Scientometrics* 126: 2485-2511.
<https://link.springer.com/article/10.1007/s11192-020-03522-5>
- Mandelis, Andreas. 2019. "Review of Scientific Instruments New Products." *Review of Scientific Instruments* 90 (3): 039501.
- Mingers, John and Loet Leydesdorff. 2015. "A Review of Theory and Practice in Scientometrics." *European Journal of Operational Research* 246 (1): 1-19.
<https://doi.org/10.1016/j.ejor.2015.04.002>
- Peralta-González, María Josefa, Maylin Frías-Guzmán, and Orlando Gregorio-Chaviano. 2015. "Criterios, clasificaciones y tendencias de los indicadores bibliométricos en la evaluación de la ciencia." *Revista Cubana de Información en Ciencias de la Salud* 26 (3): 290-309.

- Pérez-Angón, Miguel Ángel and Gabino Torres-Vega. 1998. "Perspectiva de la física mexicana." *Interiencia* 23: 163.
- Peters, Isabella and Judith Bar-Ilan. 2015. "Informetrics, bibliometrics, altmetrics: ¿What is it all about?" *Proceedings of the American Society for Information Science and Technology* (April 24).
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101015>
- Restrepo-Arango, Cristina and Rubén Urbizagástegui-Alvarado. 2017. "Red de co-palabras en la bibliometría mexicana." *Investigación Bibliotecológica* 31 (73): 17-45.
<http://www.scielo.org.mx/pdf/ib/v31n73/2448-8321-ib-31-73-00017.pdf>
- Robles-Glenn, Jorge. 1971. "La investigación mexicana y los índices extranjeros de información". *Anuario de Bibliotecología, Archivología e Informática* 3 (época 2): 47-100.
- Rodríguez-Miramontes, Jorge and Claudia Noemi González-Brambila. 2016. "The effect of external collaboration on research output in engineering." *Scientometrics* 109 (2): 661-675.
<https://doi.org/10.1007/s11192-016-2054-7>
- Sengupta, Indranil N. 1992. "Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview." *Libri* 42 (2): 75-98.
<https://doi.org/10.1515/libr.1992.42.2.75>
- Tague-Sutcliffe, Jean M. 1996. "Some perspectives on the evaluation of information retrieval systems." *Journal of the American Society for Information Science* 47: 1-3.
- Uddin, Shahadat, Nazim Choudhury, and Md Ekramul Hossain. 2019. "A research framework to explore knowledge evolution and scholarly quantification of collaborative research." *Scientometrics* 119 (2): 789-803.
<https://doi.org/10.1007/s11192-019-03057-4>
- William, Evans. 2001. "Mapping Mainstream and Fringe Medicine on the Internet." *Science Communication* 22 (3): 292-299.
- Wolfram, Dietmar. 2015. "The symbiotic relationship between information retrieval and informetrics." *Scientometrics* 102: 2201-2214.
<https://doi.org/10.1007/s11192-014-1479-0>
- Youngblood, Mason and David Lahti. 2018. "A bibliometric analysis of the interdisciplinary field of cultural evolution." *Palgrave communications* 4 (120): 1-9.
<https://www.nature.com/articles/s41599-018-0175-8>

Para citar este texto:

- Luna-Morales, María Elena, Miguel Ángel Pérez-Angón y Evelia Luna-Morales. 2021. "Evolution of the scholar community in the area of informetrics in Mexico: 1971-2018". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 51-78.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58386>

Appendix

Relation of the most cited papers by Mexican scholars in the field of metric studies (1971-2018)

Year	Authors	Title	Source	Document Type	Citations	*Database
2007	Gonzalez-Brambila, C.; Veloso, F. M.	"The determinants of research output and impact: A study of Mexican researchers"	<i>Research Policy</i> 36 (7): 1035-1051	Article	105	WoS
2013	Nagata, J. M.; Hernández-Ramos, I.; Kurup, A. S.; Albrecht, D.; Vivas-Torrealba, C.; Franco-Paredes, C.	"Social determinants of health and seasonal influenza vaccination in adults ≥65 years: A systematic review of qualitative and quantitative data"	<i>BMC Public Health</i> 13 (388)	Review	99	Scopus
2012	Barreto, S. M.; Miranda, J. J.; Figueroa, J. P.; Schmidt, M. I.; Munoz, S.; Kuri-Morales, P. P.; Silva, J. B.	"Epidemiology in Latin America and the Caribbean: current situation and challenges"	<i>International Journal of Epidemiology</i> 41 (2): 557-571	Article	84	WoS
2007	Hamel, R. E.	"The dominance of English in the international scientific periodical literature and the future of language use in science"	<i>Aila Review</i> 20 (1): 53-71	Article	83	Scopus
2007	Lowe, R. A.; Gonzalez-Brambila, C.	"Faculty entrepreneurs and research productivity"	<i>Journal of Technology Transfer</i> 32: 173-194	Article	77	WoS
2009	Martínez-Mekler, G.; Martínez, R. A.; Del Río, M.-B.; Mansilla, R.; Miramontes, P.; Cocho, G.	"Universality of rank-ordering distributions in the arts and sciences"	<i>Plos One</i> 4 (3): e4791	Article	73	Scopus

2007	Mansilla, R.; Koppen, E.; Cocho, G.; Miramontes, P.	"On the behavior of journal impact factor rank-order distribution"	<i>Journal of Informetrics</i> 1 (2): 155- 160	Article	63	WoS
2014	Aleman-Nava, G. S.; Casiano-Flores, V. H.; Carde- nas-Chavez, D. L.; Diaz-Chavez, R.; Scarlet, N.; Mahlknecht, J.; Dallemand, J. F.; Parra, R.	"Renewable energy research progress in Mexi- co: A review" (ITESM, estudio de la producción Tecnología del agua)	<i>Renewable & Sustainable Energy Reviews</i> 32: 140-153	Review	62	WoS
2004	Morales-Gonzales, J. M.; Benito-Leon, J.; Rivera-Navarro, J.; Mitchell, A. J.	"A systematic approach to analyse heal- th-related quality of life in multiple sclerosis: the GEDMA study" (Universidad de Veracruz, Estudio de esclerosis múltiple, metodo- logía cuantitativa y cualitativa)	<i>Multiple Sclerosis Journal</i> 10 (1): 47-54	Article	62	WoS
2001	Kostoff, R. N.; Del Rio, J. A.; Humenik, J. A.; Garcia, E. O.; Ramirez, A. M.	"Citation mining: Integrating text mining and bibliometrics for research user profiling"	<i>JASIST</i> 52 (13): 1148-1156	Article	57	WoS
2011	Arechavala Vargas, R.	"Las univer- sidades y el desarrollo de la investigación científica y tecno- lógica en México: Una agenda de investigación"	<i>Revista de la Educación Superior</i> 40 (158): 41-57	Article	54	Scielo
2000	Ramirez, A. M.; Garcia, E. O.; Del Rio, J. A.	"Renormalized impact factor"	<i>Scientome- trics</i> 47 (3-9): 7	Article	49	WoS

2012	Kaplan, W. A.; Ritz, L. S.; Vitello, M.; Wirtz, V. J.	"Policies to promote use of generic medicines in low and middle income countries: A review of published literature, 2000-2010"	<i>Health Policy</i> 106 (3): 211-224	Review	47	WoS
2001	Ingwersen, P.; Larsen, B.; Rousseau, R.; Russell, J.	"The publication-citation matrix and its derived quantities"	<i>Chinese Science Bulletin</i> 46: 524-528	Editorial Material	46	WoS
2014	Vessuri, H.; Guédon, J.-C.; Cetto, A. M.	"Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development"	<i>Current Sociology</i> 62 (5): 647-665	Article	41	Scopus
2016	Cadenas, E.; Rivera, W.; Campos-Amezcuca, R.; Heard, C.	"Wind speed prediction using a univariate ARIMA model and a multivariate NARX model" (Es un estudio de impacto de variables, varias instituciones de México)	<i>Energies</i> 9 (2):109-	Article	40	Scopus
2013	Gonzalez-Brambila, C. N.; Veloso, F. M.; Krackhardt, D.	"The impact of network embeddedness on research output"	<i>Research Policy</i> 42 (9): 1555-1567	Article	39	WoS
2010	Olivas-Avila, J. A.; Musi-Lechuga, B.	"Analysis of the production of psychology professors in Spain in journal articles of the Web of Science"	<i>Psicothema</i> 23 (2): 267-273	Article	36	WoS

2009	Musi-Lechuga B., Olivas-Ávila J.A., Buela-Casal G.	"Producción científica de los programas de Doctorado en Psicología Clínica y de la Salud de España"	<i>International Journal of Clinical and Health Psychology</i> 9 (1): 161-173	Article	36	Scopus
2000	Urzua, C. M.	"A simple and efficient test for Zipf's law"	<i>Economics Letters</i> 66 (3): 257-260	Article	35	WoS
2014	Collazo-Reyes, F.	"Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country"	<i>Scientometrics</i> 98: 197-209	Article	34	WoS
1995	Lozoya, X.; Rivera-Arce, E.; Dominguez, F.; Arellano, M. L.; Muñoz, O.	"Archives of medical research: An historical and subject coverage overview"	<i>Archives of Medical Research</i> 26: S1-S5	Article	33	WoS
2010	Li, W. T.; Miramontes, P.; Cocho, G.	"Fitting Ranked Linguistic Data with Two-Parameter Functions"	<i>Entropy</i> 12 (7): 1743-1764	Article	31	WoS
2008	Naumis G. G.; Cocho G.	"Tail universalities in rank distributions as an algebraic problem: The beta-like function"	<i>Physica A</i> 387 (1): 84-96	Article	30	Scopus

* This column includes the databases that were used to retrieve articles and number of citations

Lectura y conocimiento en Bibliotecología

Héctor Guillermo Alfaro López*

Artículo recibido:
3 de diciembre de 2020

Artículo aceptado:
23 de agosto de 2021

Artículo de investigación

RESUMEN

Se busca comprender cómo se despliega cognoscitivamente la lectura a través del campo bibliotecológico, principalmente en la educación y la investigación, para lo cual se implementan los conceptos *mediadores*, *intermediadores* y *traducción* del sociólogo Bruno Latour. Los intermediadores consisten en las acciones que transportan contenidos sin que sufran cambios; así, la lectura concebida como intermediador ha predominado en el campo bibliotecológico y además está determinada por la objetividad. Los mediadores transforman la significación de sus contenidos al pasar por la traducción. La lectura como mediador conlleva la transformación de la información en conocimiento

* Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México, México
galfaro@unam.mx

diferente o nuevo al ser traducido por mediación del planteamiento de problemas y preguntas al texto, lo cual propicia una lectura signada por la subjetividad y que puede constituirse en factor de transformación cognoscitiva del campo bibliotecológico.

Palabras clave: Lectura; Conocimiento; Mediator; Intermediador

Reading and knowledge in Library Science

Héctor Guillermo Alfaro López

ABSTRACT

The intention here is understanding how reading develops cognitively through the Library Science field, mainly in its education and research dimensions, resorting to the following concepts of sociologist Bruno Latour: *mediators*, *intermediaries* and *translation*. The *intermediaries* consist of the actions that transport content without undergoing changes, so that the reading (conceived as an intermediary) has predominated in the Library Science field and besides is also determined by objectivity. The *mediators* transform the meaning of its contents by undergoing *translation*. Reading as a mediator entails the transformation of information into different or new knowledge as it is translated, through the posing of problems and questions to the text, all of which encourages a reading marked by subjectivity, which can become a transforming factor in the Library Science field.

Keywords: Reading; Knowledge; Mediator; Intermediary

INTRODUCCIÓN

La temática de la lectura sólo hasta un momento relativamente reciente ha ocupado un lugar preponderante y consolidado entre los temas tradicionales y canonizados de la Bibliotecología. Eso ha permitido avizorar la extrema complejidad de esta práctica, que de ninguna manera se reduce a una mera descodificación de la textualidad, con lo cual pone de manifiesto

su posición estratégica dentro del conocimiento bibliotecológico: toda la actividad bibliotecológica y bibliotecaria tiene un objetivo, lo que le da su sentido, satisfacer las necesidades de información de la sociedad; pero de manera individual y concreta las personas acceden a esa información que se les brinda por mediación de la lectura.

Tal posicionamiento de la temática de la lectura ha estimulado una vasta y variada producción investigativa con lo que se ha ampliado el conocimiento de esta práctica dentro de la disciplina bibliotecológica; sin embargo, en la vastedad de esta producción de indagación de diversa índole sobre la lectura, queda de manifiesto una desproporción entre la gran mayoría de investigaciones que se deslizan sobre la superficie del tema y aquellas, escasas, que buscan dar un soporte más sólido, fundando conceptual y teóricamente la temática de la práctica de la lectura dentro del campo bibliotecológico.

Es previsible que en la medida en que más investigaciones se elaboren desde esta última perspectiva, contribuirán a su vez a dar fundamento bibliotecológico a este tema crucial, con lo que por añadidura se contribuye a la fundamentación científica del campo bibliotecológico. De ahí la pertinencia, en este caso, de abordar la relación de lectura y conocimiento en Bibliotecología a partir de una serie de conceptos que explican cómo se ha desplegado la lectura en los procesos cognoscitivos del campo bibliotecológico a lo largo de su desenvolvimiento histórico.

MÉTODO

Al iniciar el recorrido de esta indagación es preciso dejar establecido que se plantea desde un enfoque claramente epistemológico. Desde este enfoque se enuncia el problema que orienta la reflexión: ¿cuál es el rol de la práctica de la lectura en los procesos cognoscitivos en el campo bibliotecológico? Para dar respuesta a esta cuestión, por método, se establece primeramente el contexto: el campo bibliotecológico que, en cuanto campo de conocimiento, se encuentra integrado por un “conjunto de regiones” que conceptualmente pueden definirse como *prácticas globales*, como son investigación, biblioteca, educación bibliotecológica, asociaciones de bibliotecarios y publicaciones, entre otras.

A través de todas las prácticas globales circula y se manifiesta la práctica de la lectura de manera específica, de acuerdo a las características de cada una de las regiones mencionadas. Mas, es en la práctica global de la educación donde la lectura asume una presencia *iniciática*, que será el soporte de su desenvolvimiento a través de las otras regiones del campo. La base que

sustenta el proceso educativo bibliotecológico son los *saberes*, que encuentran cauce y plasmación en los diversos soportes y formatos de los textos, pero que aquí serán conjuntados y subsumidos en la categoría genérica de *libro*. A continuación se explica cómo se correlaciona la lectura en la educación con la lectura en la investigación bibliotecológica.

Conceptos

En el orden metódico, conforme se da explicación del desenvolvimiento de la práctica de la lectura en el campo bibliotecológico, se van exponiendo e instrumentalizando los conceptos centrales que dan explicación de las distinciones que asume la lectura a lo largo de ese desenvolvimiento.

Los conceptos que se implementan han sido tomados de la TAR, teoría del actor red, desarrollada por el sociólogo francés Bruno Latour, así como de la hermenéutica de Hans-Georg Gadamer y de la pedagogía de Paulo Freire; aunque estas dos últimas no se encuentren abiertamente ante los reflectores son cruciales para la comprensión del uso de los conceptos elegidos de la TAR. Es de subrayar que, para evitar una interpretación sesgada o limitada respecto al uso que aquí se hace de los conceptos de las mencionadas teorías, el proceder que se ha seguido es crítico-creativo. Lo que significa que se no trata, por ejemplo, de hacer un estudio exhaustivo de la TAR ni de los conceptos en ella desarrollados (que contiene más conceptos de los que aquí se implementan), lo que estaría totalmente fuera de la concepción y objetivo de este artículo. Un proceder crítico-creativo se refiere a la resignificación creativa de los conceptos. Por poner un ejemplo: el sociólogo francés Pierre Bourdieu tomó de Marx el concepto de *capital*, de sentido económico, y de manera crítico-creativa le dio un nuevo contenido epistemológico, resignificándolo al ampliar el rango explicativo del concepto, de manera que para él el capital pasaba a ser también *capital simbólico* o *político* o *social*. Los conceptos no son entidades fijas ni canonizadas ya que pueden correr el riesgo de esclerotizarse. Siguiendo ese procedimiento, aquí se hacen interactuar los conceptos tomados de la TAR con los elementos aportados por las teorías de Gadamer y Freire. Y es con esa conjunción de elementos conceptuales que *debe ser* comprendido el desarrollo del tema aquí tratado, el cual, en cuanto tal, es un escorzo de una teoría más amplia que desde años atrás se ha venido elaborando, la teoría de la constitución y autonomía del campo bibliotecológico (Alfaro López, 2010; 2018; 2016), cuyos conceptos medulares son el basamento sobre el que se despliega la argumentación de este artículo.

PRESENTACIÓN

Michel Foucault explicó que los saberes son cúmulos de conocimientos ya que son creación de una práctica discursiva y, en cuanto tales, son condición de posibilidad de la ciencia. Pero el saber no se reduce a la ciencia (Foucault, 2009: 9). De manera más explícita: el saber es sobre lo que se habla en la práctica discursiva que, para nuestro caso, es todo aquello que se refiere a la bibliotecología; son los objetos propios del campo y que se les ha dado estatuto científico; es lo referido al campo y los discursos que sobre los objetos se expresan en él, esto es, en las diversas regiones que lo componen; son también la conjunción de enunciados y proposiciones, conceptos y teorías que se aplican en el conocimiento bibliotecológico y, por último, es la utilización de las posibilidades que ofrece el discurso, en nuestro caso, bibliotecológico.

Así es como a lo largo de su proceso formativo los futuros bibliotecólogos reciben el saber propio de su campo que ha quedado plasmado en los libros. En tales libros el saber que circula en el campo bibliotecológico encuentra concreción y con ello la configuración de ciencia, en versión estándar y canónica. En su famoso libro *La estructura de las revoluciones científicas*, Kuhn (2013) da razón de cómo ese tipo de textos son la instancia a través de la que se forman las sucesivas generaciones de científicos, aprendiendo los rudimentos de las ciencias, pero con ello dan también una imagen de la ciencia empobrecida, poco adecuada para su complejidad. Esta clase de libros esquematiza y simplifica el saber, dando una versión estabilizada y uniforme de la ciencia. Es por mediación de la lectura que el saber, constituido en ciencia, puede convertirse en conocimiento; lo cual significa que más allá de los procesos cognoscitivos que acompañan a la lectura, en su producción de conocimiento en sí misma, conlleva elementos que permiten esa generación de conocimiento; y no es, por consiguiente, sólo una técnica neutra que únicamente vehiculiza el saber de la ciencia.

La lectura llevada a cabo en educación bibliotecológica significa el primer acceso orgánico y sistemático al conocimiento bibliotecológico que signará las pautas de lectura normalizada de los sujetos posicionados en las diversas regiones del campo bibliotecológico en su fase de constitución. Esto conduce preguntarse ¿cómo es que esta clase de lectura circula entre los sujetos y las distintas regiones?, ¿cuáles son sus implicaciones en el campo bibliotecológico? El hecho de plantear tales cuestiones nos pone en el lugar donde se bifurcan los caminos entre los *hechos del devenir histórico* (pasado-presente) y los *posibles del porvenir* del campo bibliotecológico y, con ello, los desplazamientos pasado-presente-futuro de la práctica de lectura. Para dar respuesta a las preguntas planteadas se seguirá primero la vía de los hechos que vienen

del pasado y desembocan actualizándose en el presente, para después seguir el camino de los posibles en el porvenir.

Para seguir el desplazamiento de la lectura a través del campo bibliotecológico se implementan algunos de los conceptos de la sociología de Bruno Latour debido a su carácter relacional y fluido: *intermediario* y *mediador-traducción*. Este sociólogo francés define sus conceptos a partir de su concreción dinámica explicando que un intermediario es una entidad que transporta significado, pero no genera transformación, lo cual resulta predecible respecto a los datos que vehiculiza, por lo que no importa lo complicado que pueda ser y por ello puede representar algo o nada, eso lo hace olvidable. Por la parte del mediador éste es heterogéneo en su cualidad de transportar múltiples significados, sus datos de entrada y salida son impredecibles; transfigura y cambia el significado de los elementos con que se expresa su capacidad para llevar a cabo la *traducción* (concepto también central de Bruno Latour, del que se dará explicación y hará uso más adelante), esto es, la transformación de aquello que es transportado. En palabras del propio Latour:

Un intermediario, en mi vocabulario, es lo que transporta significado o fuerza sin transformación: definir sus datos de entrada basta para definir sus datos de salida. Para todo propósito práctico un intermediario puede considerarse no sólo una caja negra sino también una caja negra que funciona como una unidad, aunque internamente esté compuesta de muchas partes. Los mediadores, en cambio, no pueden considerarse sólo uno; pueden funcionar como uno, nada, varios o infinito. Sus datos de entrada nunca predicen bien los de salida; su especificidad debe tomarse en cuenta cada vez. Los mediadores transforman, traducen, distorsionan y modifican el significado o los elementos que se supone que deben transportar. No importa lo complicado que sea un intermediario, puede representar, para todo propósito práctico, una unidad o incluso nada porque puede ser fácilmente olvidado. Por simple que pueda parecer un mediador, puede volverse complejo; puede llevar en múltiples direcciones que modificarán todas las descripciones contradictorias atribuidas a su rol. (Latour, 2008: 63)

Queda así de manifiesto su complejidad multidireccional. Los términos de intermediario y mediador, en cuanto conceptos de tipo relacional que expresan movimiento vehiculizador de contenidos, relacionan tanto a sujetos, acciones y objetos que según el flujo de su movimiento por momentos se presentan en un primer plano haciéndose abiertamente visibles, para un instante después retroceder a un segundo plano, interrumpiendo la transmisión de datos de que son portadores (Latour, 2008: 48-67). Los susodichos conceptos marcan el punto donde los caminos se separan, como se había enunciado líneas atrás, entre los intermediarios que nos dan explicación del desplazamiento de la lectura en el

pasado y el presente, y los mediadores que nos proyectan a las posibilidades de la lectura futura en el campo bibliotecológico.

Veamos en primera instancia la trayectoria que sigue el intermediario: la práctica de la lectura es un intermediario que se relaciona con el objeto libro, que en sí mismo es un intermediario. La lectura en cuanto intermediario transporta significado pero sin transformación; el carácter de los propios datos con que se le da inicio definen los datos posteriores de salida de ella. Esto es debido a que el libro, con el que se lleva una relación de intermediario, contiene el *saber normalizado* que sustenta a la ciencia bibliotecológica asimismo normalizada (en terminología de Thomas Kuhn), la cual tiene un marcado sustrato técnico que también es un factor que signa con su impronta a la lectura. Al leer el libro se desenvuelven como intermediarios lectura-libro, pero no se da entre ambos la traducción (como explica Bruno Latour), no produce modificación de significado o de los elementos que se supone deben transportar ambos para producir conocimientos nuevos o diferentes.

En otras palabras, los saberes con los que de antemano cuenta el lector para llevar a cabo la lectura no sufren cambio, traducción, para dar lugar a un conocimiento bibliotecológico diferente o nuevo. Sólo se reiteran, transportan, los conocimientos previamente establecidos y estabilizados (técnicamente), propios de un capital de conocimiento constituido y estatuido en el campo bibliotecológico. La lectura como intermediario se despliega a través del campo; habiendo adquirido ese estatuto originariamente en la educación, se manifiesta de esa manera en el resto de las regiones del campo. En la región de investigación la lectura y los libros adquieren su pleno estatuto cognoscitivo, puesto que transportan significado, conocimiento, sin transformación; pero con tal estatuto (intermediario) se contribuye a legitimar epistemológicamente el conocimiento ya establecido y consolidado, con su basamento técnico. Con ello pone en evidencia que es un conocimiento objetivado y estático, que se corresponde ceñidamente con una lectura en cuanto intermediario, llevada a cabo por un sujeto normalizado y objetivado.

La lectura de un texto de ciencia enfrenta a una información que se pretende es el resultado de una investigación precisa y rigurosa en observación y experimentación. Y, además, se encuentra escrito en una prosa austera, encorsetada, sin puntos de fuga hacia la subjetividad, donde los conceptos obturan el paso a las metáforas y las imágenes: escritura donde señorea la primera persona del plural. El lector, por su parte, tiene que mediar su lectura con los códigos que le brinda un texto de semejante índole para hacérselo legible. Ello implica que ha de contar con elementos previos del saber de la ciencia y aceptar las señalizaciones de objetividad a lo largo de la ruta que marca el texto. Pero la lectura del texto científico no se reduce a la esfera de

comprensión del contenido, sino que también retroactúa en otro nivel sobre el lector: lo transfigura, constituyéndolo como un lector objetivado, normalizado de acuerdo al orden racionalista establecido por el texto, que a su vez es resultado de una ciencia normalizada bajo el diseño de objetividad y que se encuentra en condición de llevar a cabo la lectura de libros de su ciencia como intermediario dentro del campo de conocimiento.

Una vez llegados a este límite es pertinente retornar al punto en que los caminos se separaban, para ahora seguir la otra senda, de los mediadores; *camínemos hacia el horizonte de los posibles*. Recordemos cómo Bruno Latour caracteriza a los mediadores: “[...] sus datos de entrada nunca predicen bien los de salida; su especificidad debe de tomarse en cuenta cada vez. Los mediadores transforman, traducen, distorsionan y modifican el significado con los elementos que se suponen deben transportar” (Latour, 2008: 63). La práctica de la lectura comprendida como mediador es aquella que al ser un proceso de relación con un libro-texto brinda una información: hay datos de entrada pero no pueden predecirse bien los datos o más exactamente los conocimientos que se producen de salida una vez que ya se ha realizado el acto de lectura, ello en razón de que se da la traducción entre la lectura y el libro transformando el significado que se supone transporta el texto, a diferencia de la lectura en cuanto intermediario que implica continuidad, pero ésta aparentemente sencilla interpretación de la lectura como mediador conlleva una gran complejidad ya que en ella, además, se ventilan arduos procesos cognitivos, que a su vez entrañan procesos de transfiguración del propio campo bibliotecológico en su conjunto, por lo que se requiere dar precisa explicación de ello.

El concepto clave en el despliegue de los mediadores es el de traducción, ya que expresa el momento disruptivo que altera o cambia el movimiento, la acción, dándoles una orientación cualitativa distinta a los contenidos, en este caso de la lectura y el libro. Qué viene a significar la traducción dentro de la lectura en cuanto mediador: el planteamiento de problemas y su enunciación en preguntas. Todo lo contrario a una lectura que reitera, sin cuestionar, una información establecida y canonizada, y que en cuanto tal, funge como reafirmación de semejante canonización. La práctica de la lectura, mediador, al desplegarse sobre el libro plantea y/o encuentra problemas en el texto por conducto de las preguntas. Pero para que la lectura transite por el espíritu de la pregunta, la cual es vehiculizada en la traducción, hay condicionantes que respaldan a la lectura: los saberes.

A diferencia del saber que viene a ser el significado (elementos) que vehiculiza la lectura como intermediario, cuyo carácter es especializado y reiterativo de un conocimiento establecido, la lectura en cuanto mediador ha de estar integrada por saberes múltiples y cambiantes. El bibliotecólogo ha

de estar respaldado en la lectura de sus textos bibliotecológicos por libros de distintas disciplinas: filosofía, literatura, historia, sociología, etc. Bagaje cognitivo cultural que al fusionarse con el saber bibliotecológico permite su ampliación de visión respecto a la realidad social sobre la que actúa. Pero, más aún, el aporte de tales disciplinas propicia la manifestación de la imaginación y la creatividad que pueden galvanizar la práctica de la lectura. De hecho ese impulso imaginativo y creativo es el que se llega a desenvolver en aquella lectura que plantea problemas y preguntas críticas innovadoras, revulsivas al texto.

De esta manera se puede establecer, como plantea Hans-Georg Gadamer, un diálogo con el texto, un diálogo en el que el constante plantear preguntas es el pasaporte infaltable para recorrer todos los territorios del libro (Gadamer, 1993). Las preguntas sacan de ese mutismo en que la lectura como intermediario deja al texto, para hacer que éste asimismo nos interpele y con ello que cuestionemos la certeza de nuestros saberes. Por otra parte, es de indicar que una lectura de tal índole implica que las preguntas estén marcadas por el riesgo; como explican Paulo Freire y Antonio Faundez: al estimular la invención rompen con las formas establecidas, burocráticas y seguras de conocimiento (Freire y Faundez 2013), propiciando la generación de conocimientos nuevos, diferentes, alternativos.

Ahora bien, conforme la lectura-mediador se despliega, una presencia la acompaña en las sucesivas traducciones por las que cruza: la *subjetividad*. Esa presencia que fue exiliada, incluso, anatémizada desde los orígenes de la ciencia occidental moderna, y que como desvaída sombra fue arrinconada en los suburbios de expresiones de saber que se estimaba no ofrecían la verdad, como las Humanidades, las Artes y las Ciencias Sociales. La objetividad se erigió en el inquisidor que condenó al exilio a la subjetividad. Pero el proyecto histórico de conformación de la objetividad comenzó a desgastarse en las décadas finales del siglo pasado, lo que implicó la exhumación y el retorno de la subjetividad al escenario principal del conocimiento, mas ya no con el antiguo atuendo, sino con un rostro renovado: no es una entidad estática ni exclusivamente recluida en el espacio de los sentidos sino, por el contrario, como la caracteriza Michel Foucault, es moviente mutable y puede producir un conocimiento también verdadero.

Es claro que la práctica lectora, mediador, es llevada a cabo por un lector constituido en cuanto subjetividad, pero a su vez esta clase de lectura es instituyente de subjetividad: es un movimiento recursivo donde simultáneamente ambos son causante y causado. La fuerza motriz de este movimiento recursivo es la imaginación y la creatividad producto de la fusión e interacción de saberes diversos, lo que termina por resquebrajar la figura monolítica y pétrea del lector normalizado y objetivado, que pone en práctica la lectura

como intermediario y que resulta ser racionalista y descorporizado, reproductor y legitimador de un capital de conocimiento, establecido y fundado técnicamente. Por otra parte, para que la lectura-mediador muestre todo su potencial transformador, ha de comprenderse a través de su desenvolvimiento en la región de investigación del campo bibliotecológico.

Cuando se hacía el seguimiento de la lectura como intermediación, partiendo de la región de educación bibliotecológica, se explicaba cómo esa trayectoria de lectura en la investigación se convierte en legitimadora epistemológica de un conocimiento establecido y canónico, en el que el cuestionamiento no tiene una posición reconocida. Por el contrario, la práctica de lectura, mediador, con sus consiguientes traducciones a través de problemas y preguntas, rompe a su paso las barreras que impone un conocimiento estático, reiterativo y técnico para desembocar en el océano de un conocimiento diferente e innovador. El proceso de investigación, que tiene su antesala en la lectura de los libros que contienen la ciencia bibliotecológica, al estar marcado por la lectura como mediador conlleva la problematización perenne, la puesta en marcha de preguntas. Esto da lugar a una *reflexividad imaginativa* y *creativa* que, inclusive, puede hacer uso extensivo e intensivo de la *construcción conceptual y teórica de objetos y prácticas* para fundamentarlos bibliotecológicamente (Alfaro López, 2018). Al tener esta ruta de lectura el respaldo de los saberes de otras disciplinas ofrece elementos cognoscitivos de apoyo para llevar a cabo crítica y creativamente dicha construcción epistemológica.

Como mera acotación, la lectura, mediador, confluye a su vez en otro mediador: la escritura, lo que conlleva que igualmente se despliegue a través de varias instancias de traducción. La escritura como mediador implica la concreción de investigación crítica conducida por la subjetividad cognoscente. Escribir bajo la directriz investigativa ha de conllevar mediaciones, supuestos y desenvolvimientos reflexivos y críticos, que son galvanizados por el planteamiento de problemas y preguntas: la escritura de semejante índole no ha de ser una mera redacción y reiteración de saberes y conocimientos ya establecidos. Lectura y escritura en cuanto mediadores son los *posibles* que pueden ser un agente de impulso para que el campo bibliotecológico transite de su fase de constitución hacia su fase de autonomía, de plena cientificidad.

CONCLUSIÓN

La práctica de lectura tiene una notable preponderancia dentro de un campo de conocimiento debido a su posición distintiva y estratégica: es la instancia a través de la cual se tiene acceso a los saberes (información) específicos, en

nuestro caso, propios y definitorios del campo bibliotecológico; asimismo, la lectura en cuanto mediador contribuye para llevar a cabo la transformación de los saberes en conocimiento y sobre todo de manera dirigida cuando está surcada por problemas y su enunciación en preguntas, puesto que son factores que estimulan la creatividad. Creatividad que se convierte en impulso hacia la generación de conocimiento diferente o innovador cuando la lectura de los saberes bibliotecológicos se encuentra, además, respaldada por saberes provenientes de diversas disciplinas. Esto pone de manifiesto la especificidad y complejidad de los procesos cognoscitivos que se realizan en las distintas regiones del campo, en especial en la práctica global de investigación. Para hacer legibles tales procesos, que se realizan por medio de la lectura, se requiere un abordaje de índole conceptual y teórico. Los conceptos ofrecen una visión abstracta y concatenada de procesos empíricos, lo que permite una comprensión unificada de ellos.

La teoría, al ser un sistema estructurado de conceptos, se configura como una concepción y conocimiento multidimensional de la realidad, de la cual no debe estar exenta la imaginación creativa. Con ello se abre también la vía para emprender la fundamentación de la práctica de la lectura como objeto de conocimiento bibliotecológico, y cómo tal práctica, a su vez, puede convertirse en un factor, incluso, decisivo para la transformación del campo bibliotecológico.

Agradecimientos

Agradezco al Programa de Apoyos para la Superación del Personal Académico de la UNAM (PASPA), por el apoyo otorgado durante la estancia de investigación en la Escuela Interamericana de Bibliotecología de la Universidad de Antioquia, Medellín Colombia, donde se llevó a cabo el presente trabajo.

REFERENCIAS

- Alfaro López, Héctor Guillermo. 2010. *Estudios epistemológicos de bibliotecología*. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas.
- Alfaro López, Héctor Guillermo. 2016. "Problemas en la construcción de la imagen y la lectura de imagen como objetos de conocimiento en el campo bibliotecológico", en *Hacia la construcción de la imagen y su lectura como objetos de conocimiento en Bibliotecología*, coordinado por Héctor Guillermo Alfaro López y Graciela Leticia Raya Alonso, 1-30. México: UNAM-IIBI.
- Alfaro López, Héctor Guillermo. 2018. *Construcción epistemológica de la imagen y la lectura de imagen como objetos de conocimiento en el campo bibliotecológico*. México: UNAM, Instituto de Investigaciones Bibliotecológicas y de la Información.

- Foucault, Michel. 2009. *La arqueología del saber*. México: Siglo XXI.
- Freire, Paulo y Antonio Faundez. 2013. *Por una pedagogía de la pregunta. Crítica a una educación basada en preguntas inexistentes*. Buenos Aires, Argentina: Siglo XXI.
- Gadamer, Hans-George. 1993. *Verdad y Método I*. Salamanca, España: Sígueme.
- Kuhn, Thomas S. 2013. *La estructura de las revoluciones científicas*. México: FCE.
- Latour, Bruno. 2008. *Reensamblar lo social. Una introducción a la teoría del actor red*. Buenos Aires, Argentina: Manantial.

Para citar este texto:

- Alfaro López, Héctor Guillermo. 2021. "Lectura y conocimiento en Bibliotecología". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 79-90.
- <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58381>

Repercusión altmétrica y tipo de acceso en artículos peruanos de Ciencias Sociales

Joel Alhuay-Quispe*

Lourdes Bautista-Ynofuente**

Artículo recibido:

7 de mayo de 2021

Artículo aceptado:

31 de agosto de 2021

Artículo de investigación

RESUMEN

El artículo analiza el impacto en medios sociales de artículos peruanos del área temática de las Ciencias Sociales a través de la repercusión mediática en fuentes altmétricas y el tipo de acceso en el que se publican. Se analizan 1 186 artículos con filiación de instituciones peruanas recuperados de la base de datos Scopus (2011-2020), para lo cual se emplearon seis (6) indicadores altmétricos de Altmetric y citas de *Dimensions*. El método de análisis compara la repercusión altmétrica en relación con el tipo de acceso: abierto o restringido, en cinco áreas temáticas de las Ciencias Sociales: Artes y Humanidades, Negocios

* Facultad de Ingeniería, Universidad Privada San Juan Bautista, Perú
joel.alhuay@upsjb.edu.pe

** Universidad Nacional Mayor de San Marcos, Perú
lourdes.bautista1@unmsm.edu.pe

y Contabilidad, Ciencias de la Decisión, Economía y Finanzas, y Psicología. Los resultados muestran una alta correlación entre lecturas en Mendeley y citas de *Dimensions* ($r = 0.797$), y que existe una diferencia significativa por el tipo de acceso a la publicación en las menciones provenientes de Facebook ($p = .000$) y Twitter ($p = .005$). El estudio concluye que existe una mayor tendencia por parte de los autores peruanos que abarcan temas de Psicología en publicar sus artículos en acceso abierto, otorgándoles mayor visibilidad en Mendeley y *Dimensions*.

Palabras clave: Producción Científica; Indicadores Alométricos; Acceso Abierto; Ciencias Sociales

Altmetric attention and access type in Peruvian articles of Social Sciences

Joel Albuay-Quispe and Lourdes Bautista-Ynofuente

ABSTRACT

This article aims to identify the impact in altmetric sources and type of publishing access in Social Sciences articles. It analyses 1,186 articles affiliated to Peruvian institutions retrieved from Scopus (2011 to 2020) using six altmetric indicators from Altmetric and *Dimensions* citations. The method used compares altmetric impact in relation to open or restricted access and subject areas: Arts and Humanities, Business and Accounting, Decision Sciences, Economics and Finance, and Psychology. Results show a high correlation between Mendeley reads and citations of *Dimensions* ($r = 0.797$), and besides there is a significant difference according to access type in mentions from Facebook ($p = .000$) and Twitter ($p = .005$). In conclusion, Peruvian Psychology authors publish mostly in some open access route and publications obtain greater visibility in Mendeley and *Dimensions* when they are published through green route.

Keywords: Scientific Production; Altmetric Indicators; Open Access; Social Sciences

INTRODUCCIÓN

La masificación de plataformas de *social media* ha transformado la manera de comunicar la producción científica, así como la aparición de las denominadas métricas alternativas o altmétricas ha variado la concepción tradicional de la evaluación del impacto de las publicaciones en las Ciencias Sociales como en las demás áreas del conocimiento.

La evaluación o medición de los resultados obtenidos por un investigador y su equipo de estudio se lleva a cabo de modos distintos o complementarios: a) *peer-review* o evaluación por pares, y b) por valoración *a posteriori* de los resultados de la investigación (Peña-Rey Lorenzo, 2004).

El acceso abierto (*open access* en inglés) se refiere al acceso inmediato, permanente y gratuito, en línea y a texto completo de todos los artículos de revistas de investigación arbitradas (Harnad, 2005). Para Piwowar *et al.* (2018), el acceso abierto se define como el acceso libre para leer en línea, ya sea en el sitio web del editor o en un repositorio, por lo que todos los artículos que no cumplen esta definición son de acceso cerrado (restringido). La vía de publicación de acceso abierto (OA) surge en notoria oposición a la mercantilización del conocimiento y de la comunicación científica, posteriormente apropiado por los grandes sellos editoriales y revistas de acceso restringido (por suscripción) como modelo de negocio en proyección a su persistencia en el tiempo. El movimiento OA adquiere su naturaleza internacional a partir de tres grandes hitos: las declaraciones de Budapest (2002), Bethesda (2003), y Berlín (2003) (Harnad, 2005). En consecuencia, iniciativas supranacionales como la de la Unión Europea buscan acelerar la transición de los sistemas tradicionales de publicación cerrada hacia los de acceso abierto (Ellers, Crowther y Harvey, 2017), eliminando así el costo de suscripción para la lectura (i.e. descarga) de los contenidos. Una de las modalidades de suscripción más comunes es la institucional, la cual consiste en que una entidad realice compras de colecciones de títulos y recursos electrónicos mediante su biblioteca o quien hiciera labores de gestor de información. Por otro lado, también está la opción de acceder de manera individual mediante el pago por artículo proveniente de los recursos propios del investigador o las fuentes de financiamiento de sus proyectos.

Las revistas inicialmente publicaban sus contenidos mediante suscripción; hoy en día, se encuentran apostando por el acceso abierto o modalidad mixta. De este modo, el movimiento Open Access nace como ruta alterna de comunicación de la producción científica ante el predominio de los canales comerciales de publicación académica, permitiendo la disponibilidad del texto completo para lectura de manuscritos: la ruta *verde*, a través de repositorios

digitales, y la ruta *dorada*, mediante las propias revistas electrónicas. La ruta dorada consiste en que los artículos se publican en acceso abierto y los costos se trasladan a los autores o entidades a las cuales se adscriben los investigadores; estos costos, denominados cargos de procesamiento o APC (*article processing charges* por sus siglas en inglés), se definen como las tasas que cobran los editores a los autores por la publicación de artículos de revistas (Beasley, 2016). Por otro lado, la ruta verde implica que el autor de un manuscrito publicado en una revista de acceso cerrado deposite una copia gratuita en un repositorio digital en su versión preliminar (*preprint*) o versión corregida (*postprint*). Además de estas dos categorías de publicación en acceso abierto, existen las denominadas *híbrido*, *bronce*, y la ruta *cerrada*. Piwowar *et al.* (2018) definen las cinco rutas de publicación como sigue:

- Dorada (*gold*): publicado en una revista de acceso abierto indexada por el DOAJ.
- Verde (*green*): acceso de pago en la página del editor, pero con una copia gratuita en un repositorio.
- Híbrido (*hybrid*): gratis bajo una licencia abierta en una revista de acceso de pago.
- Bronce (*bronze*): acceso libre en la página del editor, pero sin una licencia claramente identificable.
- Cerrado (*closed*): todos los demás artículos, incluidos los que sólo se comparten en una red privada o de acceso mediante Sci-Hub.

No obstante, a pesar de la expansión del movimiento de acceso abierto y su filosofía de abrir el conocimiento científico para todos, son los propios autores quienes no asimilan esta práctica como una vía crucial para masificar el acceso al conocimiento, además de la posibilidad de medir el impacto y reconocimiento de sus publicaciones. Este último se traduce en las citas recibidas, un indicador por excelencia para medir el impacto académico de una publicación científica; no obstante, indicadores complementarios y recientes como la repercusión en plataformas sociales y medios web podrían dar indicios tempranos de los artículos que serán posteriormente citados. Para Bornmann (2014), uno de los cambios producidos al interior de la cienciometría es el acceso abierto y el movimiento del análisis de citación en la web hacia el análisis del uso de los medios sociales, fenómeno conocido actualmente como *altmetrics* o *altmétricas*, en español. Estas métricas alternativas suponen un complemento a la evaluación de la influencia de la investigación, a partir de nuevas audiencias que van más allá de la propia comunidad académica y científica puesto que las herramientas tradicionales para el análisis bibliométrico, como

el factor de impacto y el índice H, no capturan el impacto de productos de investigación conectados a la web (Lahikainen, 2016). Este enfoque de complementariedad al análisis de citas tradicionales parece ser más significativo y relevante en comparación con el uso aislado de los tradicionales indicadores bibliométricos (Melero, 2015; Alonso-Arévalo, 2016; Bornmann, 2014; Costas, Zahedi y Wouters, 2015; Haustein, 2016; Wilsdon *et al.*, 2015). Respecto a estas métricas se plantean interrogantes sobre las posibles limitaciones en el campo de la bibliometría y evaluación de la ciencia. Una de ellas es la falta de precisión respecto al objeto de medición: ¿impacto o influencia?, ¿social o científica? (Sud y Thelwall, 2014; Haustein, Bowman y Costas, 2015). Por otro lado, surgen los problemas de normalización e inconsistencia en la recuperación de los datos (Robinson-García *et al.*, 2014) y errores de los propios autores (Alhuay-Quispe y Bautista-Ynofuente, 2020), sumado a las diferentes coberturas de las plataformas que proporcionan los datos altmétricos.

La mayoría de los estudios bibliométricos en torno a la producción científica peruana se centra en las áreas de la salud como la medicina y la odontología, mientras que sólo algunos se enfocan en las ciencias sociales como la psicología (Carranza Esteban *et al.*, 2020; Livia, 2008) y otros analizan casos de instituciones específicas (Delgado Arenas, Delgado Arenas y Hermitaño Atencio, 2021; Romaní, 2020) o se comparan con otros países (Limaymanta *et al.*, 2020; Armijos Valdivieso, Avolio Alecchi y Arévalo-Avecillas, 2021). En cambio, las investigaciones de estudios métricos de la información de alcance altmétrico a partir de publicaciones científicas de procedencia peruana son escasas o de reciente interés (Vílchez-Román, Huamán-Delgado y Alhuay-Quispe, 2021).

Mugnaini *et al.* (2021) afirman que los estudios sobre la disponibilidad y uso de los DOI (siglas en inglés de *Digital Object Identifier*) pueden ayudar en el flujo de información científica en la era digital y garantizar una mayor coherencia al realizar análisis bibliométricos. En ese sentido, las herramientas de monitoreo de la actividad altmétrica de las publicaciones basan la recolección de datos a partir de varios identificadores persistentes. Estos códigos de identificación varían de acuerdo con el tipo de documento, tipo de plataforma web o áreas de publicación, tales como:

- a) Identificadores de repositorios: utilizados en plataformas del mismo nombre.
 - PubMed iD en ciencias de la salud.
 - arXiv iD en física, matemáticas y ciencias de la computación.
 - ADS iD en astrofísica.
 - SSRN iD en ciencias sociales.

- RePEC iD en economía.
- b) Identificadores por tipos de objetos:
 - Handle usados en repositorios digitales.
 - URN (*Uniform Resource Name*).
 - URL: cualquier URL como de los repositorios.
 - ISBN: usado en libros.
 - YouTube iD: URL de video en YouTube.
 - Clinical Trial iD: usado en repositorio de ensayos clínicos.
 - OCLC iD: usado para registros del catálogo OCLC.
 - DOI: utilizado para artículos, tesis, libros, archivos de datos, entre otros tipos de documentos.

Actualmente existen diferentes plataformas que rastrean datos alométricos (Altmetric, ImpactStory, PaperCritic, PLoS Impact, Plum Analytics, *Dimensions*, ScienceCard, CrowdoMeter, NewsFlow), siendo Altmetric (altmetric.com) la plataforma comercial y proveedor de datos alométricos más importante. Esta plataforma captura información del impacto de un documento de varias fuentes de medios sociales y calcula una puntuación ponderada denominada Altmetric Score (Robinson-García *et al.*, 2014). Actualmente, incluye como fuente de datos a varias plataformas sociales en línea (*Tabla 1*).

Dimensión	Alcance	Fuente	Situación
Policy documents	Documento de política, orientación o directrices de una organización gubernamental o no gubernamental.	Sitios web oficiales	Vigente
Mainstream media	Colección de fuentes de noticias curada manualmente y captadas en tiempo real.	Más de 5 000 medios	Vigente
Blogs	Seguimiento de una lista seleccionada de blogs académicos y no académicos realizado automáticamente mediante canales RSS.	Más de 15 000 blogs	Vigente
Online reference managers	Menciones en recursos de marcadores sociales y gestores bibliográficos.	Mendeley	Vigente
		CiteULike	Descontinuada
Post-publication peer-review	Seguimiento en plataformas web de reconocimiento de revisión de expertos.	Publons	Vigente
		PubPeer	Descontinuada

Social media	Menciones en posts, grupos o páginas públicas de varias fuentes de redes sociales y <i>microblogging</i> .	Twitter	Vigente
		Facebook	Vigente
		Reddit	Descontinuada
		Google+	
		Pinterest	
		LinkedIn	
Patent citations	Seguimiento de las citas de patentes de varias jurisdicciones de todo el mundo.	10 oficinas de patentes	Vigente
Citations	Citaciones recibidas en documentos incluidos en fuentes y bases de datos de citación.	<i>Dimensions</i>	Vigente via Altmetric Explorer
		Web of Science	
Other online sources	Seguimiento de atención de varias fuentes no incluidas en las categorías anteriores	Wikipedia	Vigente
		Q&A Forums	Vigente
		Faculty Opinions	Vigente
		YouTube	Vigente
		Open Syllabus	Vigente

Tabla 1. Categorías de repercusión y fuentes de cobertura en Altmetric

Fuente: Altmetric (2020)

En líneas generales, los objetivos del artículo se enmarcan en responder las siguientes interrogantes:

- ¿En qué tipos de acceso se publican los artículos de cinco áreas temáticas de las Ciencias Sociales?
- ¿Cuál es la distribución de fuentes altmétricas según el tipo de acceso publicado?
- ¿Existe asociación entre las menciones obtenidas de las fuentes altmétricas?
- ¿Existe diferencia en la repercusión mediática según el tipo de acceso?

METODOLOGÍA

Diseño

La investigación es de tipo descriptivo y correlacional, de carácter transversal y emplea la revisión documental a fin de identificar literatura previa sobre indicadores altmétricos.

Muestra

Se emplea una muestra de tipo intencional no probabilística conformada por los artículos con filiación a instituciones peruanas indizados en Scopus para el periodo 2011-2020 y clasificados en cinco áreas de las Ciencias Sociales: Artes y Humanidades, Negocios y Contabilidad, Ciencias de la Decisión, Economía y Finanzas, y Psicología. A partir de una muestra inicial de 3 092 artículos, se identifica una muestra final de 1 186 artículos correspondientes a las publicaciones con identificador DOI y al menos una atención altmétrica recibida.

La recolección de los datos se realizó en abril de 2021 y se empleó a Scopus como fuente de datos para recuperar artículos científicos. Se construyó una estrategia de búsqueda que permitió identificar publicaciones de tipo “artículo” y de tipo “revisión” con filiación peruana (campo AFFILCOUNTRY) para los años de estudio y delimitado a cinco áreas temáticas dentro de las Ciencias Sociales del sistema de clasificación ASJC (All Science Journal Classification Codes) que emplea Scopus (*Tabla 2*).

El año de inicio para nuestro corte de tiempo (2011) corresponde al año después de la publicación del Manifiesto Altmetrics; se toma el umbral de tiempo a dos periodos de cinco años en referencia a los fundamentos de los indicadores basados en este periodo: *H5-index*, *Impact factor 5-year*.

Los datos exportados de Scopus en formato de archivo delimitado por comas (csv) fueron depurados, y luego se seleccionaron los artículos con identificador DOI para la recuperación de los datos altmétricos mediante el recurso Altmetric Explorer de la plataforma comercial Altmetric. La evidencia del tipo de acceso abierto a los documentos proviene de la información registrada en *Dimensions* y extraída desde Altmetric Explorer.

Análisis de datos

Los análisis estadísticos se efectuaron mediante el software IBM SPSS 23. En el análisis correlacional bivariado se emplea el coeficiente Rho de Spearman y la prueba de hipótesis considera valores p inferiores a 0.05 como significativos. Los análisis incluyen las siguientes variables:

- Áreas: cinco áreas de ciencias sociales (*Tabla 2*).
- Documentos: número de artículos con DOI y actividad altmétrica.
- Menciones en Twitter.
- Menciones en Facebook.
- Repercusión en blogs.
- Repercusión en sitios de noticias.

- Menciones en Wikipedia.
- Lecturas en Mendeley.
- Tipo de acceso: abierto (*bronze, gold, green, hybrid*) y restringido.
- Ratio mediática: cociente entre menciones en una fuente y documentos del área.

Código	Categoría	Área ASJC
arts	Arts and Humanities	Archeology (arts and humanities)
		Arts and Humanities (miscellaneous)
		Arts and Humanities (all)
		Classics
		Conservation
		History
		History and Philosophy of Science
		Language and Linguistics
		Literature and Literary Theory
		Museology
		Music
		Philosophy
		Religious studies
		Visual Arts and Performing Arts
busi	Business, Management, and Accounting	Business, Management, and Accounting (all)
		Accounting
		Business and International Management
		Business, Management, and Accounting (miscellaneous)
		Industrial relations
		Management Informational Systems
		Management of Technology and Innovation
		Marketing
		Organizational Behavior and Human Resource Management
		Strategy and Management
		Tourism, Leisure, and Hospitality Management

deci	Decision Sciences	Decision Sciences (all)
		Decision Sciences (miscellaneous)
		Information Systems and Management
		Management Science and Operations Research
		Statistics, Probability and Uncertainty
econ	Economics, Econometrics and Finance	Economics, Econometrics, and Finance (all)
		Economics, Econometrics, and Finance (miscellaneous)
		Economics and Econometrics
		Finance
psyc	Psychology	Applied Psychology
		Clinical Psychology
		Developmental and Educational Psychology
		Experimental and Cognitive Psychology
		Neuropsychology and Physiological Psychology
		Psychology (all)
		Psychology (miscellaneous)
		Social Psychology

Tabla 2. Áreas temáticas de ciencias sociales ASJC

RESULTADOS

Presencia en fuentes altmétricas

La *Tabla 3* muestra que de la población inicial recuperada (3 092 artículos), la cantidad de artículos con DOI representa 80 % (2 466), mientras que sólo 38 % son artículos con identificador DOI y por lo menos una actividad altmétrica que corresponde al número final de publicaciones incluidas en el análisis (1 186).

Áreas	Docs	% Docs	DOI	% DOI	DOI + Altm	% DOI + Altm
arts	872	28.20	661	26.80	290	24.45
busi	788	25.49	585	23.72	230	19.39
deci	198	6.40	131	5.31	49	4.13
econ	639	20.67	538	21.82	289	24.37

psyc	595	19.24	551	22.34	328	27.66
Total	3 092	100.00	2 466	100.00	1 186	100.00

Docs: Total de documentos recuperados

DOI: Documentos recuperados con DOI

Altm: Documentos con DOI y repercusión alométrica

Tabla 3. Proporción muestral de documentos con atención alométrica

Los datos del análisis descriptivo (Tabla 4) muestran que Mendeley es la fuente que proporciona mayor número de menciones en fuentes alométricas con una media de 49.63 pero también es la que mayor varianza en los datos (4 799.41) posee. Otras fuentes con mayor número de datos son las menciones en Twitter y las citas en *Dimensions*; mientras que los blogs y Wikipedia son las fuentes con menor datos alométricos.

Estadístico	Blogs	News	Facebook	Twitter	Mendeley	Wikipedia	<i>Dimensions</i>
Media	0.11	0.22	0.22	5.64	49.63	0.06	11.84
Mediana	0.00	0.00	0.00	1.00	27.00	0.00	4.00
Varianza	0.34	2.36	0.41	354.60	4 799.41	0.15	810.46
Desv. estándar	0.58	1.54	0.64	18.83	69.28	0.39	28.47
Mínimo	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Máximo	8.00	33.00	7.00	355.00	960.00	8.00	589.00
Rango	8.00	33.00	7.00	355.00	960.00	8.00	589.00
Asimetría	8.71	13.13	4.40	10.29	4.44	11.89	10.23
Curtosis	93.97	222.56	27.32	144.34	36.23	199.04	165.06

Tabla 4. Análisis descriptivo de las variables

La Tabla 5 presenta los datos acumulados de las menciones para las cinco áreas de estudio según las siete fuentes alométricas. Los datos son favorables para las fuentes Mendeley y Twitter, para los campos de Economía (econ) y Psicología (psyc), que son dos de las tres áreas con mayor número de artículos analizados (*size depending*), mientras que las Ciencias de la Decisión (deci) presentan una menor puntuación alométrica.

Áreas	Blogs	News	Facebook	Twitter	Mendeley	Wikipedia	<i>Dimensions</i>
arts	20	40	75	977	7.292	30	1.969
busi	5	16	28	730	11.645	6	2.043
deci	10	13	10	402	2.600	2	688

econ	46	62	69	2.469	16.216	10	3.825
psyc	46	125	84	2.101	21.055	27	5.511

Tabla 5. Menciones altmétricas acumuladas por áreas

Áreas de publicación y tipo de acceso

El 54 % de los artículos de Negocios y Contabilidad (125 de 230) analizados fueron publicados en revistas de acceso restringido, mientras que los manuscritos de Psicología (203 de 328) y de Ciencias de la Decisión (30 de 49 artículos) fueron publicados en acceso abierto con 62 % y 61 % respectivamente del total de artículos revisados (Figura 1).

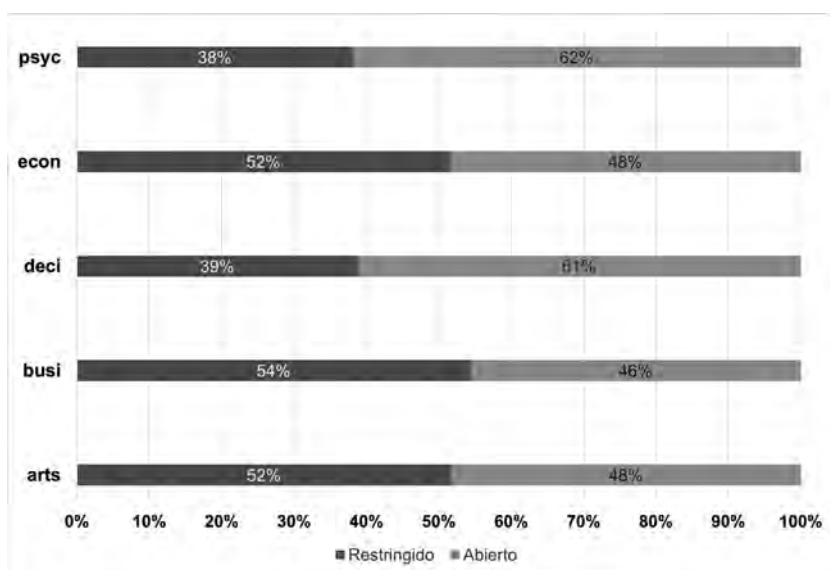


Figura 1. Proporción del tipo de acceso según área

Un análisis desagregado por tipo de acceso abierto y fuentes de datos altmétricos, empleando un cociente entre la cantidad de menciones sobre el número de artículos del área denominado “ratio mediática”, muestra que las menciones en Mendeley y las citas en *Dimensions* predominan en las rutas de acceso restringido y verde; por otro lado, la proporción mediática en Twitter aparece representada en casi todas las rutas de publicación (Figura 2).

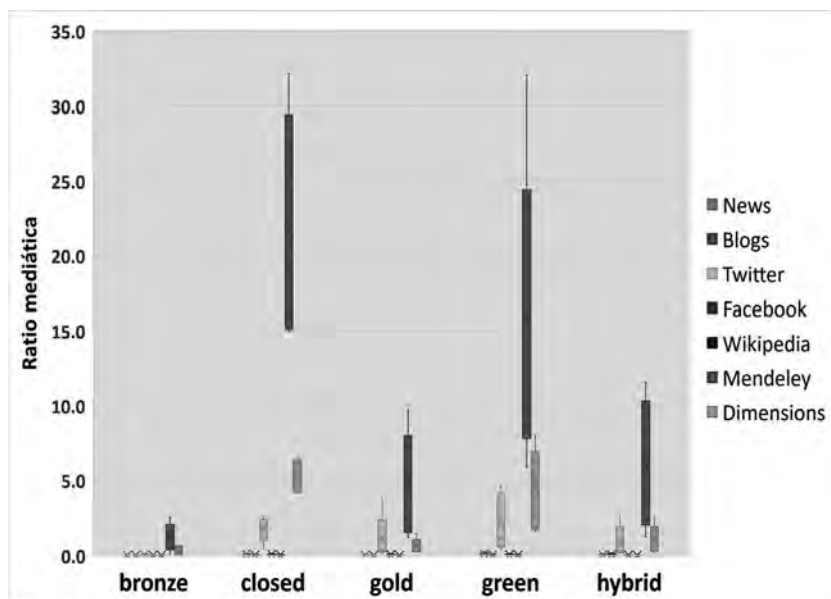


Figura 2. Ratio de menciones por documento según fuentes y tipo de acceso

Correlación de indicadores

Se obtuvo que existe una alta correlación entre el número de citas en *Dimensions* y Mendeley ($r = 0.797$), además de una asociación moderada ($r = 0.413$) entre las menciones en blogs y sitios de noticias; mientras que para los demás indicadores altmétricos la correlación resulta muy baja (Tabla 6).

Altmétricas	1	2	3	4	5	6	7
1. Blog	1.000	0.413	0.141	0.159	0.184	0.105	0.164
2. News		1.000	0.155	0.172	0.180	0.111	0.188
3. Facebook			1.000	0.127	0.137	0.047	0.102
4. Twitter				1.000	0.246	0.057	0.134
5. Mendeley					1.000	0.114	0.797
6. Wikipedia						1.000	0.147
7. Dimensions							1.000

Nota: La correlación es significativa en el nivel 0.01 (bilateral)

Tabla 6. Análisis de correlación entre indicadores altmétricos

La *Figura 3* presenta la distribución de los datos del análisis bivariado con mayor asociación a partir de la *Tabla 6*, se muestra un análisis complementario mediante el coeficiente de determinación lineal obteniéndose una alta correlación ($R^2 = 0.650$).

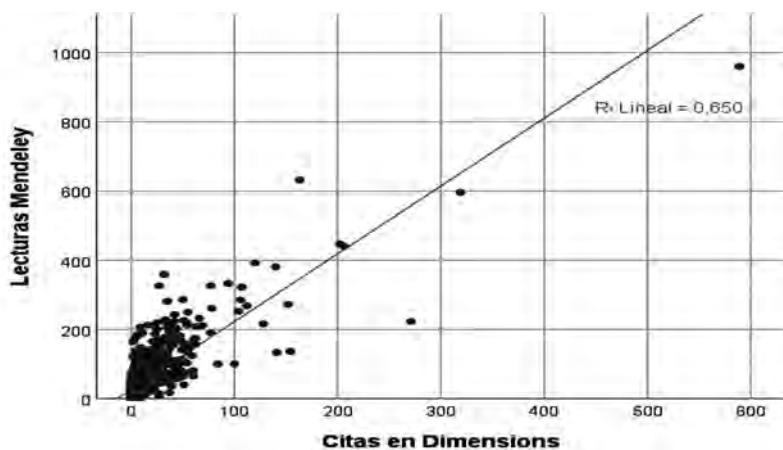


Figura 3. Correlación entre citas de *Dimensions* y lecturas de Mendeley

Finalmente, como respuesta a la última pregunta del estudio, se evidenció que existe diferencia estadísticamente significativa según el tipo de acceso (abierto o restringido) sólo para las menciones provenientes de Facebook ($p = .000$) y Twitter ($p = .005$), pero no para la repercusión altmétrica obtenida de las demás fuentes (*Tabla 7*).

Estadístico	Blog	News	Facebook	Twitter	Mendeley	Wikipedia	<i>Dimensions</i>
U de Mann-Whitney	176446.0	175911.0	189459.5	191383.0	169349.0	175751.0	163927.0
W de Wilcoxon	367099.0	366564.0	380112.5	382036.0	360002.0	366404.0	354580.0
Estadístico de contraste	176446.0	175911.0	189459.5	191383.0	169349.0	175751.0	163927.0
Sig. Asintótica*	0.615	0.772	0.000	0.005	0.318	0.802	0.053

* Bilateral. Nivel de significancia = .05

Tabla 7. Estadísticos de contraste por tipo de acceso

DISCUSIÓN

Estudios previos analizaron la correlación entre indicadores bibliométricos (número de citas recibidas) e indicadores altmétricos (capturas, menciones y uso), como Thelwall *et al.* (2013) y Costas, Zahedi y Wouters (2015), que analizan si las *altmetrics* se complementan y en qué medida con los indicadores tradicionales, incluso contemplan las diferencias disciplinares. De igual forma, algunos estudios tratan sobre el impacto de las métricas alternativas en áreas temáticas específicas, como Htoo y Na (2017), que revisan el impacto altmétrico de artículos incluidos en la colección SSCI (Science Citation Index) de Web of Science y quienes encontraron que la repercusión en medios altmétricos favorece a las publicaciones de tres áreas biomédicas que a las sociales (Library Science, Business-Finance, Law, Political Science). Por su parte, Mohammadi y Thelwall (2014) estudiaron artículos de las categorías WoS con temáticas en Ciencias Sociales y Humanidades, y evidenciaron que existe mayor correlación entre estadísticas de lecturas y citas en tres de las 10 áreas analizadas: Business & Economics ($p = 0.573$), Information Science & Library Science ($p = 0.535$), y Psychology ($p = 0.514$).

La mayoría de los estudios previos estudian la proporción de acceso abierto de los artículos extraídos de Web of Science, Scopus y otros recursos de literatura científica, pero son escasos los estudios a partir de publicaciones procedentes de un país o región. Piwowar *et al.* (2018) reportan que más de 50 % de las publicaciones en investigación biomédica y matemáticas son de libre acceso, mientras que cerca de 80 % de los artículos en química e ingeniería se encuentran en acceso cerrado. Martín-Martín *et al.* (2018) evidencian que 5 % de los artículos de ciencias sociales y del comportamiento se encuentran disponibles por la vía bronce de acceso abierto y 14% en la ruta verde. Singh, Piryani y Srichandan (2020) reportan que las publicaciones de India en las rutas verde y dorada suman 33 % en las áreas de Ciencias Sociales. Martín-Martín *et al.* (2018) reportan que estudios previos identifican la disponibilidad de acceso abierto de los artículos a partir de diferentes fuentes como Unpaywall, DOAJ, ROAR, OpenDOAR, ResearchGate, Google Académico o hasta el propio buscador de Google, por lo que una limitante de la investigación se enmarca en la disponibilidad de las rutas de acceso abierto proporcionadas por la propia base de datos Scopus.

En las Ciencias de la Información (y Bibliotecología), algunos autores demuestran que la preferencia por la publicación en acceso abierto permite que los artículos obtengan un mayor impacto científico (citas) y en las *social media* (lecturas) (Cintra, Furnival y Milanez, 2017). Otros concluyen que la gestión de contenidos (metadatos) y el uso de un sistema de gestión de revistas (OJS)

contribuyen a incrementar el uso o consumo de los artículos (vistas, descargas) y eventualmente, repercutirá en el incremento del impacto científico de la revista y los artículos publicados (Uribe Tirado, Vallejo Echavarría y Betancur Marín, 2016).

En América Latina, las iniciativas regionales de acceso abierto tienen larga trayectoria y son ampliamente reconocidas por los países de la región y también a nivel global. Por un lado, se encuentra el proyecto SciELO Network (www.scielo.org), que se clasifica como vía de publicación en ruta dorada, según la definición de Piwowar *et al.* (2018); por el otro lado, iniciativas como La Referencia (www.lareferencia.info), que consiste en una red regional de repositorios institucionales mayormente provenientes de universidades de varios países latinoamericanos, permiten la explotación de la ruta verde. En el caso del Perú, hasta el primer semestre de 2020 se reportan en DOAJ 92 revistas en acceso abierto y el recolector nacional ALICIA agrega 169 repositorios institucionales (Concytec, 2021). Sin embargo, pese a la considerable cantidad de repositorios peruanos, la proporción de artículos frente a los demás tipos de documentos depositados en los repositorios de universidades del sector privado apenas se acerca a 6 % y en las universidades públicas a 0.1 % (Alhuay-Quispe *et al.*, 2017). Por lo que, aunque en los países en los que la infraestructura de acceso abierto es relativamente buena, es necesario desarrollar una infraestructura de acceso abierto mediante la cooperación (Seo, 2018).

Finalmente, ya sea que el artículo se publique en acceso abierto o que el autor asuma el APC ante la revista para la apertura del manuscrito, y que el comité editorial de una revista gestione de manera óptima las publicaciones, la presencia de los propios autores en las diversas plataformas sociales y científicas, y más aún, la constante interacción conducirán a obtener mayor visibilidad e impacto de su productividad desde lo altmétrico a lo bibliométrico (Uribe-Tirado y Alhuay-Quispe, 2017). Asimismo, los beneficios del acceso abierto deben comprenderse junto con las dificultades inherentes para gestionar este nuevo enfoque de la publicación académica para beneficio de la sociedad (Ellers, Crowther y Harvey, 2017).

CONCLUSIONES

- Los autores peruanos de Negocios y Contabilidad publican mayormente en revistas de acceso restringido; en cambio, en el área de Psicología publican generalmente en revistas bajo alguna ruta de acceso abierto.

- Las publicaciones peruanas en Ciencias Sociales obtienen mayor visibilidad en las fuentes de Mendeley y *Dimensions* siempre y cuando los artículos se publiquen por la ruta verde de acceso abierto e incluso en acceso restringido.
- Las lecturas obtenidas en Mendeley guardan estrecha relación con las citas recibidas en *Dimensions*, mientras que las menciones en blogs y sitios de noticias muestran una ligera asociación.
- La repercusión mediática en artículos peruanos de Ciencias Sociales se diferencia por el tipo de acceso únicamente en las fuentes de Facebook y Twitter.

Agradecimientos

Nuestros agradecimientos al “Programa de acceso a los datos para investigadores” de Altmetric (<https://www.altmetric.com/research-access/>) a través del cual se logró obtener los datos altmétricos para el presente estudio.

REFERENCIAS

- Alhuay-Quispe, Joel y Lourdes Bautista-Ynofuente. 2020. “Imprecisión de los datos en estudios altmétricos”. *Revista Cubana de Información en Ciencias de la Salud* 31 (4): e1655.
<http://www.rcics.sld.cu/index.php/acimed/article/view/1655>
- Alhuay-Quispe, Joel, David Quispe-Riveros, Lourdes Bautista-Ynofuente y Josmel Pacheco-Mendoza. 2017. “Metadata Quality and Academic Visibility Associated with Document Type Coverage in Institutional Repositories of Peruvian Universities”. *Journal of Web Librarianship* 11 (3-4): 241-254.
<https://doi.org/10.1080/19322909.2017.1382427>
- Alonso-Arévalo, Julio. 2016. “Altmetrics pueden no seguir siendo alternativa por mucho tiempo”. *Revista ORL* 7 (3): 169-177.
<https://doi.org/10.14201/orl201673.14190>
- Altmetric. 2020. “Sources of Attention”.
<https://www.altmetric.com/about-our-data/our-sources/>
- Armijos Valdivieso, Pablo, Beatrice Avolio Alecchi y Danny Arévalo-Avecillas. 2021. “Factors That Influence the Individual Research Output of University Professors: The Case of Ecuador, Peru, and Colombia”. *Journal of Hispanic Higher Education*. OnlineFirst.
<https://doi.org/10.1177/15381927211008684>
- Beasley, Gerald. 2016. “Article Processing Charges: A New Route to Open Access?” *Information Services and Use* 36 (3-4): 163-170.
<https://doi.org/10.3233/ISU-160815>

- Bornmann, Lutz. 2014. "Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics". *Journal of Informetrics* 8 (4): 895-903.
<https://doi.org/10.1016/j.joi.2014.09.005>
- Carranza Esteban, Renzo F., Oscar Mamani-Benito, Josué R. Turpo Chaparro, Percy G. Ruiz Mamani y Ronald M. Hernández. 2020. "Student Scientific Output in Social Science Journals: A review of Peru, Ecuador and Paraguay". *International Journal of Early Childhood Special Education* 12 (1): 425-432.
<https://doi.org/10.9756/INT-JECSE/V12I1.201022>
- Cintra, Paulo Roberto, Ariadne Chloe Furnival y Douglas Henrique Milanez. 2017. "Vantagens de citação do acesso aberto em periódicos selecionados da Ciência da Informação: uma análise ampliada aos indicadores altmétricos". *Informação & Informação* 22 (1): 129-149.
<https://doi.org/10.5433/1981-8920.2017v22n1p129>
- Concytec (Consejo Nacional de ciencia, Tecnología e Innovación Tecnológica). 2021. "Listado de instituciones y repositorios participantes". Fecha de consulta: 9 de junio de 2021.
<https://alicia.concytec.gob.pe/instituciones/>
- Costas, Rodrigo, Zohreh Zahedi y Paul Wouters. 2015. "Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective". *Journal of the Association for Information Science and Technology* 66 (10): 2003-2019.
<https://doi.org/10.1002/asi.23309>
- Delgado Arenas, Raúl, Antonio L. Delgado Arenas y Bernardo C. Hermitaño Atencio. 2021. "Determinantes para publicación de artículos científicos en revistas indexadas: caso Universidad Nacional Agraria del Perú". *Revista General de Información y Documentación* 31 (1): 317-330.
<https://doi.org/10.5209/rgid.76972>
- Ellers, Jacintha, Thomas W. Crowther y Jeffrey A. Harvey. 2017. "Gold Open Access Publishing in Mega-Journals: Developing Countries Pay the Price of Western Premium Academic Output". *Journal of Scholarly Publishing* 49 (1): 89-102.
<https://doi.org/10.3138/jsp.49.1.89>
- Harnad, Stevan. 2005. "The Implementation of the Berlin Declaration on Open Access: Report on the Berlin 3 Meeting Held 28 February - 1 March 2005, Southampton, UK". *D-Lib Magazine* 11 (3).
<https://doi.org/10.1045/march2005-harnad>
- Haustein, Stefanie. 2016. "Grand challenges in altmetrics: Heterogeneity, data quality and dependencies". *Scientometrics* 108 (1): 413-423.
<https://doi.org/10.1007/s11192-016-1910-9>
- Haustein, S., T. D. Bowman y R. Costas. 2015. "When is an article actually published? an analysis of online availability, publication, and indexation dates", en *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*, 1170-1179.
https://www.issi-society.org/proceedings/issi_2015/1170.pdf
- Htoo, Tint Hla Hla y Jin-Cheon Na. 2017. "Disciplinary differences in altmetrics for social sciences". *Online Information Review* 41 (2): 235-251.
<https://doi.org/10.1108/OIR-12-2015-0386>

- Lahikainen, Johanna. 2016. "Altmetrics in Social Sciences and Humanities: Possibilities, Challenges, and Experiences". Trabajo presentado en IFLA World Library and Information Congress - WLIC 2016, Columbus, OH - Connections. Collaboration. Community, Ohio, USA, 16 de agosto de 2016.
<http://library.ifla.org/id/eprint/1356>
- Limaymanta, César H., Hilda Zulueta-Rafael, Cristina Restrepo-Arango y Patricio Alvarez-Muñoz. 2020. "Análisis bibliométrico y cienciométrico de la producción científica de Perú y Ecuador desde Web of Science (2009-2018)". *Información, Cultura y Sociedad* (43): 31-52.
<https://doi.org/10.34096/ics.i43.7926>
- Livia, José. 2008. "La producción científica y los estudios de post grado en Psicología en el Perú". *Interamerican Journal of Psychology* 42 (3): 431-445.
- Martín-Martín, Alberto, Rodrigo Costas, Thed Van Leeuwen y Emilio Delgado López-Cózar. 2018. "Evidence of Open Access of Scientific Publications in Google Scholar: A Large-Scale Analysis". *Journal of Informetrics* 12 (3): 819-841.
<https://doi.org/10.1016/j.joi.2018.06.012>
- Melero, Remedios. 2015. "Altmetrics – a complement to conventional metrics". *Biochemia medica* 25 (2): 152-160.
<https://doi.org/10.11613/BM.2015.016>
- Mohammadi, Ehsan y Mike Thelwall. 2014. "Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows". *Journal of the Association for Information Science and Technology* 65 (8): 1627-1638.
<https://doi.org/10.1002/asi.23071>
- Mugnaini, Rogério, Grischa Fraumann, Esteban F. Tuesta y Abel L. Packer. 2021. "Openness trends in Brazilian citation data: factors related to the use of DOIs". *Scientometrics* 126: 2523-2556.
<https://doi.org/10.1007/s11192-020-03663-7>
- Peña-Rey Lorenzo, Isabel. 2004. "Estudio bibliométrico de la producción científica sobre dioxinas a través de las bases de datos Pubmed e I.M.E. (19972003)". Tesis doctoral, Universidad de Murcia, España.
<http://hdl.handle.net/10201/167>
- Piowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Mathias, Bree Norlander, Ashley Farley, Jevin West y Stefanie Haustein. 2018. "The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles". *PeerJ* 6: e4375.
<https://doi.org/10.7717/peerj.4375>
- Robinson-García, Nicolas, Daniel Torres-Salinas, Zohreh Zahedi y Rodrigo Costas. 2014. "New data, new possibilities: Exploring the insides of altmetric.com". *El profesional de la información* 23 (4): 1386-6710.
<https://doi.org/10.3145/epi.2014.jul.03>
- Romaní, Franco. 2020. "Análisis bibliométrico de las publicaciones científicas originales del Instituto Nacional de Salud del Perú en el periodo 1998-2018". *Revista Peruana de Medicina Experimental y Salud Pública* 37 (3): 485-494.
<https://doi.org/10.17843/rpmesp.2020.373.5470>
- Seo, Tae-Sul. 2018. "Open Access Full-Text Databases in Asian Countries". *Science Editing* 5 (1): 26-31.
<https://doi.org/10.6087/kcse.114>

- Singh, Vivek Kumar, Rajesh Piryani y Satya Swarup Srichandan. 2020. "The Case of Significant Variations in gold-green and Black Open Access: Evidence from Indian Research Output". *Scientometrics* 124 (1): 515-531.
<https://doi.org/10.1007/s11192-020-03472-y>
- Sud, P. y M. Thelwall. 2014. "Evaluating altmetrics". *Scientometrics* 98, 1131-1143.
<https://doi.org/10.1007/s11192-013-1117-2>
- Thelwall, Mike, Stefanie Haustein, Vincent Larivière y Cassidy R. Sugimoto. 2013. "Do altmetrics work? Twitter and ten other social web services". *PloS one* 8 (5): e64841.
<https://doi.org/10.1371/journal.pone.0064841>
- Uribe-Tirado, Alejandro y Joel Alhuay-Quispe. 2017. "Estudio métrico de ALFIN en Iberoamérica: de la bibliometría a las altmetrics". *Revista española de Documentación Científica* 40 (3): e180.
<https://doi.org/10.3989/redc.2017.3.1414>
- Uribe Tirado, Alejandro, Juan Camilo Vallejo Echavarría y Darío Alexander Betancur Marín. 2016. "Somos visibles y tenemos impacto. Análisis desde datos de acceso abierto, altmetrics y otros de la Revista Interamericana de Bibliotecología". *Revista Interamericana de Bibliotecología* 39 (3): 243-275.
<https://doi.org/10.17533/udea.rib.v39n3a04>
- Vílchez-Román, Carlos, Farita Huamán-Delgado y Joel Alhuay-Quispe. 2021. "Social dimension activates the usage and academic impact of Open Access publications in Andean countries: a structural modeling-based approach". *Information Development* 37 (2): 209-220.
<https://doi.org/10.1177/0266666920901849>
- Wilsdon, James, Liz Allen, Eleanora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, Richard Jones, Roger Kain, Kerridge Simon, Mike Thelwall, Jane Tinkler, Ian Viney, Paul Wouters, Jude Hill y Ben Johnson. 2015. *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. HEFCE.
<https://kar.kent.ac.uk/81123/>

Para citar este texto:

- Alhuay-Quispe, Joel y Lourdes Bautista-Ynofuente. 2021. "Repercusión altmétrica y tipo de acceso en artículos peruanos de Ciencias Sociales". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 91-110.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58369>

SIGI, an integrated scientometric and curricular information system for teaching-research institutions

María Magdalena Sierra Flores*

Romel Calero Ramos**

María Victoria Guzmán Sánchez***

Juan Claudio Toledo-Roy*

Yolsy Gabriela Gamboa Calderón*

Artículo recibido:

20 de marzo de 2021

Artículo aceptado:

21 de junio de 2021

Artículo de investigación

ABSTRACT

We hereby present SIGI, an integral automated, web-based system for Research Data Management and generation of metric indicators. SIGI is an information platform and database system created to manage all data related to the academic activity of a teaching-research institution, and to generate annual reports and metric indicators at the individual and institutional levels. It integrates data related to traditional academic bibliographic products with other aspects of the academic activity (e.g., those related to teaching or

* Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, México
sierra@nucleares.unam.mx juan.toledo@nucleares.unam.mx
gamboa.yolsy@nucleares.unam.mx

** Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, México
romel.calero@c3.unam.mx

*** Instituto Finlay de Vacunas, Cuba
mvguzman@finlay.edu.cu

outreach). From this rich database, SIGI automatically generates metric indicators that are used in the creation of individual researcher profiles and in institutional evaluations and reports. We also show two examples of scientometrics analyses that can be carried out using the system: the internal collaboration network and the international collaboration map of a particular research institute. We conclude by discussing the example institute's overall experience with the system and how it could be adopted by other institutions, both nationally and internationally.

Keywords: Bibliometric Indicators; Research Data Management; Integrated Information System; Web Applications

SIGI, un sistema integral de información cuantitativa y curricular para instituciones de investigación-enseñanza

María Magdalena Sierra Flores, Romel Calero Ramos, María Victoria Guzmán Sánchez, Juan Claudio Toledo-Roy y Yolsy Gabriela Gamboa Calderón

RESUMEN

Presentamos SIGI, un sistema integral de gestión de información para datos de investigación y generación de indicadores métricos automático y con interfaz para la web. SIGI es un sistema de plataforma de información y de base de datos creado para gestionar todos los datos relacionados con la actividad de una institución de enseñanza-investigación, y para generar reportes anuales e indicadores métricos a nivel individual e institucional. Integra datos relacionados a productos bibliográficos académicos tradicionales con otros aspectos de la actividad académica (por ejemplo, relacionados con enseñanza y vinculación). De esta base de datos, SIGI automáticamente genera indicadores métricos utilizados en la creación de perfiles y evaluaciones individuales y en evaluaciones y reportes institucionales. También mostramos dos ejemplos de análisis cuantitativos que pueden ser hechos con este sistema: la red de colaboración interna y el mapa de colaboración internacional de un instituto específico. Concluimos con el ejemplo de la experiencia general del instituto

con el sistema y cómo éste podría ser adoptado por otras instituciones, tanto a nivel nacional como internacional.

Palabras clave: Indicadores Bibliométricos; Gestión de Datos de Investigación; Sistemas de Información Integrados; Aplicaciones Web

INTRODUCTION

The principles for the integration of Research Data Management (RDM) and bibliometry, and in particular the role of academic libraries in this process, have been explored in detail in the past. Torres-Salinas and Jiménez-Contreras (2012) proposed a model for the creation of bibliometric units within universities to function as central actors in managing research data and to impact institutional decision making regarding scientific policy. Cox et al. (2017) identified in a series of case studies that RDM is often accompanied by the generation of new services intended to support research and academic communication.

Many RDM systems have been developed and are currently in use by universities (Schöpfel, Prost, and Rebouillat, 2017), with Spain clearly being the breeding ground for the earlier initiatives. A pioneering example is that of Universitas XXI (1994, <https://www.universitasxxi.com/>), a modular system developed as a joint effort by a large consortium of Spanish universities including Alcalá, Carlos III de Madrid, Castilla-La Mancha, Salamanca and Valladolid. This platform performs varied administrative functions (including human resources and finances), with a modular focus on RDM, specifically management of researcher CVs. Another example is GREC (<https://www.uv.es/uvweb/servicio-investigacion/es/investigacion-uv/produccion-cientifica/introduccion-datos-grec-1285907192537.html>), developed by the Universitat de Barcelona and other research institutions of the Comunitat Valenciana, Catalunya y Balears and intended to centralize research information of the participating institutions. Lastly, GesBIB (<http://bibliotecas.csic.es/es/gesbib>) is an RDM system developed at the national level by the Scientific Information Resources Unit (URICI) of the Network of Libraries and Archives of Spain's Superior Council for Scientific Research (CSIC).

An example in Europe outside Spain is FRIDA, created in 2003 and consolidated as CRISTINA in 2010. Developed initially by the IT department of the University of Oslo and then as a joint project with other Norwegian

universities (Toft, 2003; Royal Ministry of Education and Research, 2004), it manages information on research results, a catalog of researchers, annual reports and research projects. Many examples also exist in Latin America, for example in Argentina the Information System for University Management (SIGUA) developed by the Superior Art Institute (ISA) integrates information of students, university workers, teaching resources and research reports of the institution (Guillot Jiménez, 2011). In Chile, the University of Talca develops the Research Management System (SGI, <http://sgi.utalca.cl>), which includes an academic directory, research statistics, and deploys a web site that serves to facilitate the interaction between academia, the private sector and society. In Colombia, the EAFIT University develops π EAFIT (<http://www.eafit.edu.co/investigacion/produccion-investigativa/Paginas/produccion-investigativa.aspx>) that collects information related to academic production, scientific activities, CVs of teaching staff and students, and also facilitates information that can be used to generate metric indicators. Another example in Colombia is CRAI-USTA (<https://crai.usta.edu.co/>), a digital platform intended to organize and offer resources related to teaching and research.

In Mexico, the state university of Veracruz maintains the Sistema Integral de Información Universitaria, SIIU (<https://www.uv.mx/siiu>), which is oriented towards academic and administrative management, automatizing processes, methods and procedures that are common in academic-administrative functions. It also produces some indicators targeted at orienting decision-making processes. At UNAM, the SIAE (<https://www.dgae-siae.unam.mx/>) system collects information on student registration and curricular activity in a single unified university-wide system, and all student theses are collected in a centralized repository. However, there is no university-wide research data management system in place, instead leaving individual schools or institutes to tackle this task as they see fit. Some institutes at UNAM have made attempts at creating RDM systems and repositories. The Physics Institute created SALVA (<https://salva.fisica.unam.mx/>), an information service dedicated to academic curricula, but that does not report additional bibliometric indicators. The Scientific Research Coordination of UNAM (<http://www.cic-ctic.unam.mx:31220/cisic/index.cfm>) implements a web-based service for the capture of metric indicators related to scientific production at the institutional level. These systems are usually limited to a few specific aspects of RDM that are most relevant to the institution's needs and are not meant to act as unified, integral systems for the institute or the university.

The objective of this paper is to present SIGI (Sistema Integral de Gestión de Información, or Integrated Information Management System), a research data management (RDM) system developed in-house by the academic library of the Instituto de Ciencias Nucleares (ICN) at the Universidad Nacional Autónoma de México (UNAM). Our intention is to describe the design philosophy and main features and products of the system, to showcase two examples of scientometric analyses conducted from data integrated by the system, and to discuss our experiences, as an academic library, in developing and utilizing an RDM system. While the ICN is a basic science research institute focused mainly on physics and chemistry, both the SIGI system and our experience are directly applicable to academic institutes in other fields. It is a good example of how academic libraries can leverage information technologies to create added value services for their community.

The principal feature of SIGI is that it is a system designed to integrate information from *all* aspects of the academic life of the institute, from publications of academic articles and books, participation and organization of academic events, to research grants, curricular teaching and thesis direction. It also acts as the mechanism for researchers to create and submit annual activity reports, handles student registration for the institute, and automatically generates user-customizable researcher profiles. Finally, but importantly, SIGI generates a variety of metric indicators of academic activity, while also offering other more traditional library services to the community. We are aware of few RDM systems as complete as SIGI.

Our paper is organized as follows. In “The SIGI system” we describe the SIGI system, specifying in some detail its different modules. “Results” describes some of the bibliometric information products produced by SIGI using our particular institute (the ICN-UNAM) as a case study. The metric indicators (related to publications, conferences, teaching activities, etc.) generated from this database are discussed in “Bibliometric indicators generated by SIGI”, additional, non-traditional indicators (mostly related to teaching) are discussed in “Indicators related to other academic activities”, a study of the collaboration networks of the institute’s researchers is given in “Analysis of academic collaboration”, and a study of the main research lines determined using publication keyword frequency analysis is described in “Analysis of keyword frequencies in academic publications”. Finally, the results are summarized in “Conclusions”.

THE SIGI SYSTEM

As a project conceived under the RDM philosophy, SIGI implements an integrated approach to information management and was designed to satisfy international guidelines and standards. The theoretical design of SIGI is rooted in several RDM standards, including the Frascati Manual (OECD, 2015) and the OSLO Manual (OECD and Eurostat, 2018). It satisfies the three fundamental pillars identified by Torres-Salinas and Jiménez-Contreras for a bibliometric unit (González Fernández-Villavicencio, 2017), but builds beyond them by including other functionalities not usually present in this kind of platforms.

The SIGI system innovates over similar projects in that it was designed from the start as an *integrated* system that centralizes all academic activity of an academic institution across different domains of academic life. SIGI seeks to bring all relevant data into a single database, where it can be normalized, curated, cross-referenced and analyzed. The entry site of the system can be accessed at <https://sigi.nucleares.unam.mx/>, with the researcher directory available at <https://sigi.nucleares.unam.mx/sguicn/people> and the metric indicators subsystem at <https://sigi.nucleares.unam.mx/sguicn/statistics>. Many other parts of the system are only accessible after login.

Because of its integrated design, SIGI offers significant added value over common scientific publication databases such as the Web of Science (Arciniegas Tinjacá, Gómez Gutiérrez, and Gregorio-Chaviano, 2018; Tarrats Pons, 2012) or the NASA-Smithsonian Astrophysics Data System (Sierra et al., 2009). As our system is primarily fed by the annual reports of the researchers, it includes not only information concerning academic journal and conference publications, but also teaching (undergraduate and graduate level courses, theses direction), attendance and organization of academic conferences, obtained research grants, outreach activities, and received awards, to name but a few items. SIGI was developed and is currently used by a particular research institute in Mexico, so its user interfaces are currently in Spanish. However, the underlying framework and code are already prepared for language localization, allowing the software to be deployed in other languages.

SIGI is composed of six independent component *modules* that provide separate user-facing interfaces connected to a single centralized software backend, which acts as an intermediary between them and maintains the database where all data is stored (Figure 1).



Figure 1. SIGI's general architecture and main modules

- *Annual Research Reports*: allows users (researchers and librarians) to capture, validate and generate the individual annual academic reports. Through this module researchers self-report their production and activities through a user-friendly web-based graphical user interface (*Figure 2*). The module allows importing bibliographic records from major academic databases (Web of Science, Scopus, and ADS).
- *ICN Metrics*: computes and displays metric indicators of the Institute's productivity, including refereed and non-refereed publications (*Figures 3 and 4*), courses taught, graduated thesis students, etc.
- *Students Module*: tracks student information at all levels, from short research internships to undergraduate and graduate thesis students as well as postdoctoral researchers.
- *Researcher Profiles*: automatically generates researcher profiles, which are publicly viewable in a web interface. Information on scientific production, graduated students, projects and prizes is displayed (*Figure 5*).
- *Library Services*: offers a unified interface for the management of library services.
- *Admin Module*: the main interface through which the library staff validates and normalizes data input into SIGI, manages data catalogues, etc.

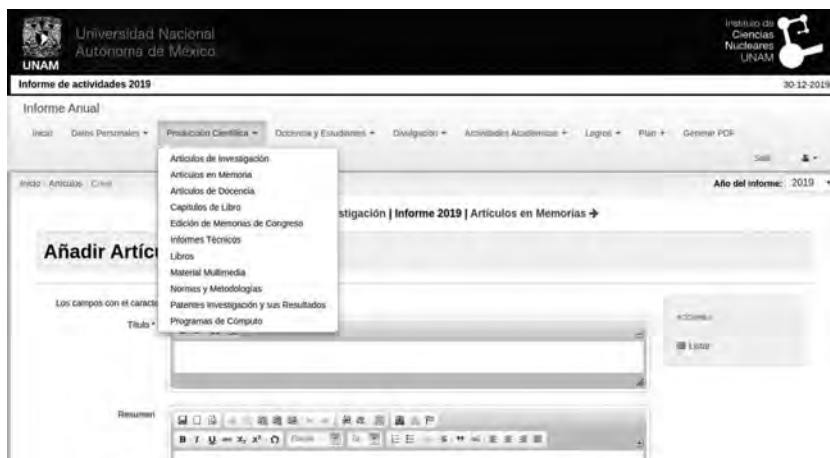


Figure 2. The Annual Research Reports main user interface

Bibliographic and academic information is entered into the SIGI system mainly through two means: by the researchers themselves, as the system is the means to fill in and generate their annual activity reports, and by library staff, who periodically review entered information and performs manual searches of new items. The system can import full bibliographic records from the Web of Science and Scopus databases, automatically translating imported fields into their equivalents in the local database (which is tailored to the information needs of the institute). Direct import capabilities from other indices such as Scielo, Redalyc and Latindex could also be added with ease.

The information collected through any of these modules is normalized and curated through several means, both (semi-)automatic and manual. For instance, research articles submitted by users to the system are automatically checked for possible duplicates (using fuzzy string matching), with the user being presented choices if a possible duplicate. Normalization is also done manually through intuitive web-based user interfaces, which allow library staff to search through data catalogues and easily edit information.

SIGI contrasts with commercial RDM solutions such as Elsevier's Pure (www.elsevier.com/solutions/pure), Clarivate's Converis (<https://clarivate.com/webofsciencegroup/solutions/converis/>) or Symplectic's Elements Platform (<https://www.symplectic.co.uk/theelementsplatform/>), which aim at offering one-size-fits-all data analysis services and products, at a substantial price tag. While this constitutes a possibly alternative, non-profit academic projects created and developed within universities (such as SIGI) are in a much better position to offer solutions tailored to the specific needs of the institutions with

a minimal resource investment. SIGI benefits from and builds upon the high adaptability and collaboration potential of open-source software.

Compared to similar open-source initiatives, such as VIVO (<https://duraspace.org/vivo/>), SIGI innovates by integrating a wider range of types of academic information (with the goal of building a more *complete* picture of academic life through data), and by offering added-value services to the institute, all from a centralized information platform. One example of the latter is the student information system, which is used not only to track academic progress and to generate associated indicators, but also to control student access at the institute's entrance booth.

RESULTS

Bibliometric indicators generated by SIGI

The SIGI database was designed with the generation of indicators in mind, be they of bibliometric type or related to human resources, teaching, outreach, etc. Because of this, SIGI offers indicators that cannot be computed from public bibliographic databases, including participation in academic events, directed theses, outreach events and publications for the general public, etc. Hence, SIGI can be used to characterize individual researchers, the department or the institution as a whole, not only by their traditional scientific production but by all aspects of academic life.

We started developing SIGI around 2005, directly using data from the ICN institute at the UNAM. The historical data of this institute since 1970 was retrospectively captured either from bibliographic electronic databases or manually from paper records, and now includes almost 8000 documents of all types, more than 2700 student records, and recognizes thousands of authors and hundreds of academic journals. From this ICN database, SIGI currently generates over a hundred indicators that allow to draw a very complete picture of the Institute's activity.

These indicators can be roughly grouped in 5 categories (Peralta González, Frías Guzmán, and Gregorio-Chaviano, 2015):

1. *Scientific production*: refereed and non-refereed research publications, patents and software. Indicators are computed at the individual researcher, department and institute levels.
2. *Visibility and impact*: the number and type of participation in academic conferences and outreach events, as well as published outreach documents and audiovisual items.

3. *Scientific activity, students and teaching*: statistics on courses taught, theses (graduated and in process), student projects and postdoctoral researchers.
4. *Scientific activity, research grants and financing*: research grants, student scholarships, and other financing sources obtained by researchers, as indicators of scientific activity and achievement.
5. *Scientific activity, individual achievements*: awards and prizes obtained by researchers and professional advancement categories (e.g., pay grade category and tenure status).

SIGI includes the capability to export data and indicators in popular data exchange formats such as CSV or XML, which allows the data to be loaded and analyzed using other tools.

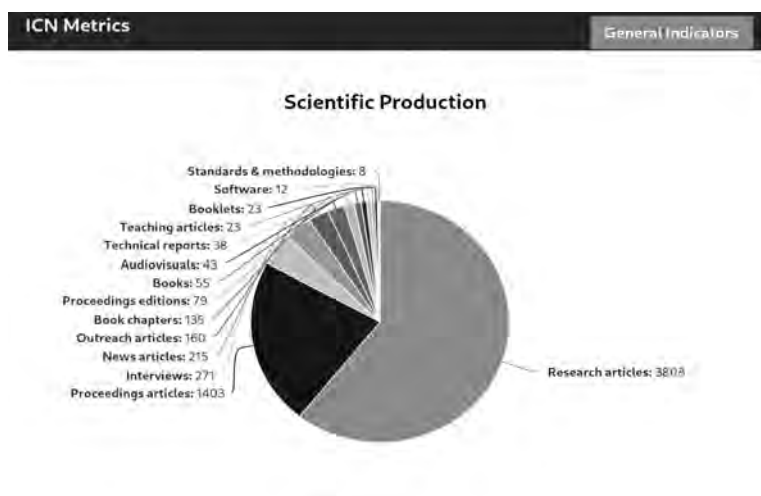


Figure 3. Scientific production of the Institute (all years on record) from the ICN Metrics module

In *Figure 3* we show the scientific production of the ICN for all years on record. It is clear that the Institute's main focus is the publication of academic refereed journal articles. *Figure 4* shows the academic journals in which the ICN publishes. We clearly see that physics (including astrophysics) and chemistry are the main research areas, and that the vast majority of publications are in international journals. *Figure 5* shows examples from the public profile generated for each ICN academic (<https://sigi.nucleares.unam.mx/sgiicn/people>), which is populated automatically with indicators and statistics obtained from the ICN Metrics module.

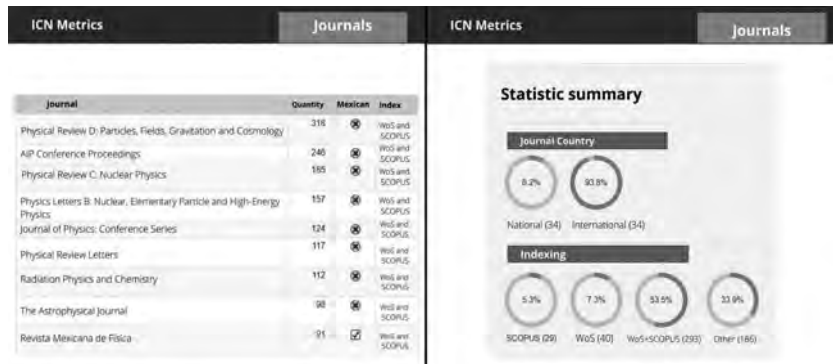


Figure 4. Statistics of academic journals in which ICN researchers publish (ICN Metrics)



Figure 5. Indicators displayed in a researcher public profile

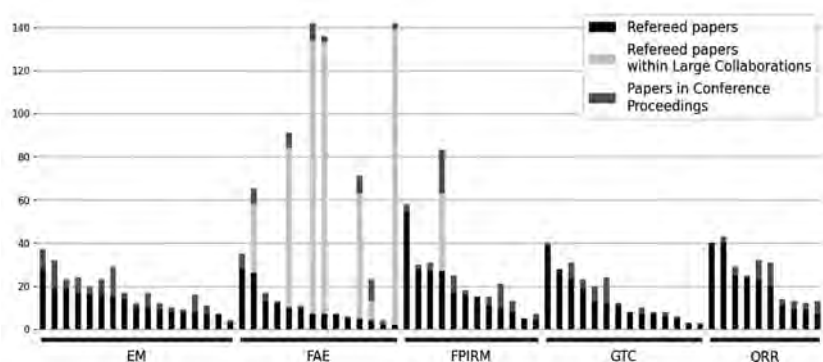


Figure 6. Scientific production (refereed journal publications and conference proceedings) of ICN broken down by individual researcher and by academic department in the 2015-2019 period. Articles with more than 50 authors are further identified as belonging to Large Collaborations

The scientific production can also be broken down by researchers and aggregated by academic department within the ICN. This is shown in *Figure 6* for the 2015-2019 period, yielding a useful comparison of the productivity of researchers and departments. Each vertical bar represents the production of individual researchers, who are grouped by academic department (EM: Structure of Matter, FAE: High Energy Physics, FPIRM: Plasma Physics, GTC: Gravitation and Field Theory, QRR: Radiation Chemistry).

Notably, we see that the production of the High Energy Physics (FAE) department is dominated by the participation in Large Collaborations (in dark black defined as articles having more than 50 authors), a rising mode of academic output in some fields of science. Excluding large collaborations, the distribution of individual productivity within each department is similar across departments.

Indicators related to other academic activities

The SIGI system includes information that is not found in public bibliographic databases such as the Web of Science (i.e., research publications) that are also important parts of academic activity. One such example are indicators related to participation in academic events. Also calculated are indicators related to academic mobility, such as research visits abroad by ICN researchers and visiting scholars from other institutions.

Another group of indicators not directly related to academic production computed by SIGI are those tied to student activity: university courses taught by institutional researchers (at both undergraduate and graduate levels), social

service students (a roughly 6-month academic internship that undergraduate students must complete during their last year), and, importantly, the direction of theses in the three academic levels recognized in Mexico (bachelor's, master's and doctorate). Tracking this kind of indicators is crucial if academic evaluation is to shift away from the dominance of research articles as the main evaluation metric used in many contexts worldwide. The task of tracking this information naturally falls on the information units and libraries of universities.

Analysis of academic collaboration

An interesting example of metric analysis that can be extracted from SIGI data is the collaboration network between the institutional researchers. Bibliographic data of all refereed research articles published by 104 researchers in the 2009-2019 period was exported from SIGI, including those of Large Collaborations. This yielded 1634 records. Using Python code we analyzed the authors field of the records, identifying co-authorships of the Institute's researchers. We excluded academics with no publications in the studied period (84 academics remained).

The resulting collaboration network was visualized using the open-source software Gephi (gephi.org) and is shown in *Figure 7*. Each circle (node) represents a researcher, its size proportional to the number of published articles in the studied period. Links between researchers indicate co-authorships, with the width of the line proportional to the number of collaborations. The tones of grey of the circles and links correspond to the different departments and information units of the ICN. The Fruchterman-Reingold algorithm was used to direct the node placement in the plot, resulting in a natural spatial grouping of clusters.

Several interesting patterns are observed in the map. While the collaborations mainly cluster according to the five academic departments (which means that intra-department collaborations are more frequent than inter-department ones), all departments are somewhat fragmented into smaller sub-clusters of researchers. For instance, one of the departments (FPIRM) is separated into a main cluster (indicated as A1 in the *Figure 7*), two smaller isolated clusters (A2 and A3) and one researcher with no intra-department collaborations (A4). Three of the other departments display two distinct clusters (B1 and B2, C1 and C2, D1 and D2). Curiously, the last department is split into one main cluster (E1) and a set of researchers with no collaborations whatsoever within the institute, at least in the study period (E2).

The strong clustering of departments is an indication of the strong coherency and affinity in the research areas of the Institute, while the sub-department fragmentation speaks of further specialization (for instance, the main cluster, A1, is composed of researchers in plasma astrophysics, while the researchers of the other clusters of this department work in fusion plasmas and laboratory plasmas) or of interpersonal relationships among researchers.

It is interesting to note that inter-department collaborations are often mediated by researchers that act as bridges by collaborating with researchers from more than one department. The two best examples are the two researchers indicated by the black arrows. The one on the left connects clusters A1 and B1 of two departments, while the other one connects two clusters B1 and B2 within the same department. These researchers act to transfer knowledge and techniques from one research area to another and constitute good examples of the rising trend of multidisciplinary research. The analysis also reveals that 10 out of 84 researchers do not collaborate (at least within the studied period) with the rest of the academic community of the Institute.

Figure 8 shows the same graph but with node sizes indicating the number of collaborations (not publications as in *Figure 7*). Only one researcher coincides in being both highly productive and highly collaborative (the central node in cluster B1). One could infer from this that, at least in this particular field of science, productivity and collaboration are not necessarily linked. This highlights the separate importance of highly collaborative researchers, suggesting that a full characterization of a researcher should look beyond common indicators like the number of publications.

In *Figure 9* we show computed network metrics for the large, connected component ($N = 52$, shown in the right panel) of the collaboration network of the ICN. The histograms are the degree distribution (left) and the distribution of node distances (shortest path lengths); also shown to the right are the other two smaller connected components. While this component is too small to reliably determine whether it is scale-free, the distribution does exhibit a downwards trend, with most researchers having few collaborations and a few researchers (the hubs that we identified above) having a large number of collaborations. The shortest path length distribution reveals a slightly right-tailed distribution with a small average distance of 4.26 and an intermediate average clustering coefficient of 0.412. This suggests that the network may be considered “small-world”, where the average distance between nodes is short despite the network not being very highly interconnected (as indicated by the intermediate clustering coefficient). This is confirmed by computing the “small-worldness coefficient” σ of Humphries and Gurney (2008), which for the largest component yields $\sigma = 3.41$, where a value of larger than one indicates a small-world network.

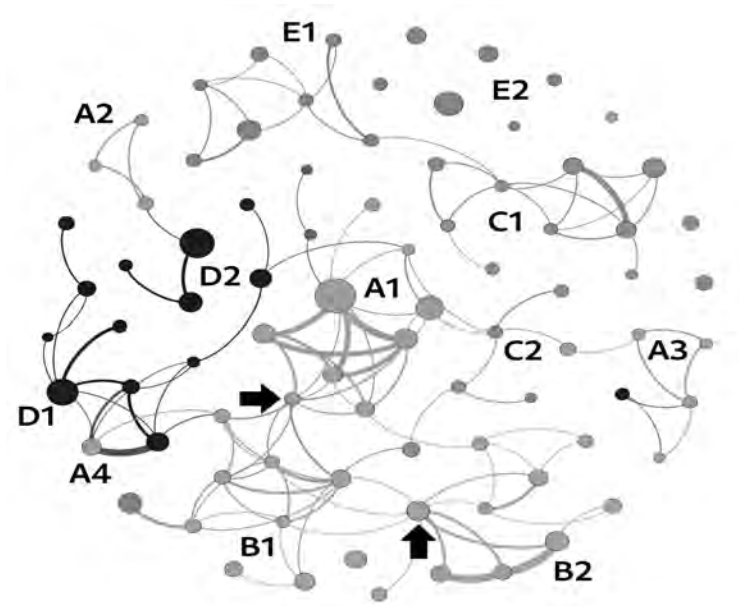


Figure 7. Scientific collaboration map of the Institute. Each circle is a researcher, its size proportional to the number of published articles in 2015-2019; links indicate the frequency of co-authorships between two researchers. The individual departments are indicated with different shades of grey

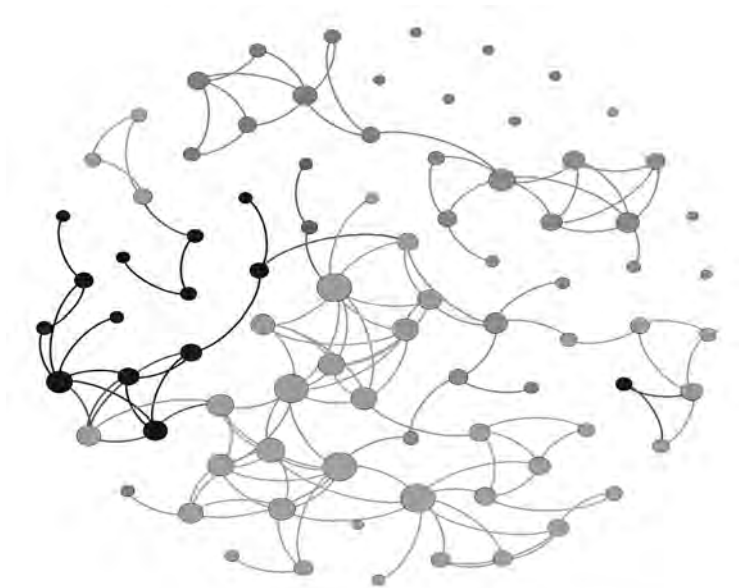


Figure 8. Same as Figure 7, except that node sizes indicate the number of collaborations

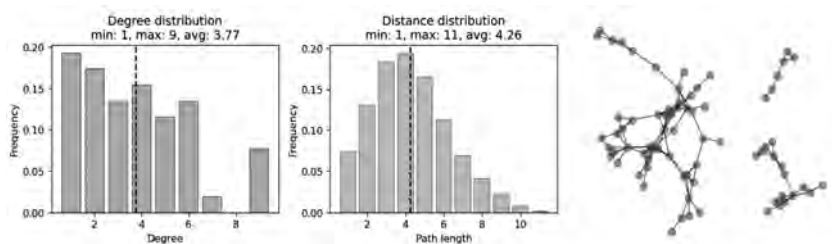


Figure 9. Network metrics of the ICN internal collaboration network: degree distribution (left) and path lengths distribution (center) for the largest connected component (52 nodes). This component, as well as the two other smaller connected components, are shown in the right panel

Using SIGI data we can also study the international collaboration network of the ICN. From the same data set spanning the 2009-2019 period we extracted the country information from the author affiliations. From this we mapped the geographical distribution of the ICN's international collaborations using the Python cartographic library (scitools.org.uk/cartopy).

The resulting map, with grey scale shading indicating the number of refereed articles with at least one non-Mexican co-author, is shown in Figure 10. The Institute maintains collaborations with researchers from 74 countries, with the bulk being with Germany and Italy in Europe and the United States in the American continent. We also identified a few collaborations with countries with small populations or with emerging scientific traditions, such as Bangladesh and Ghana.

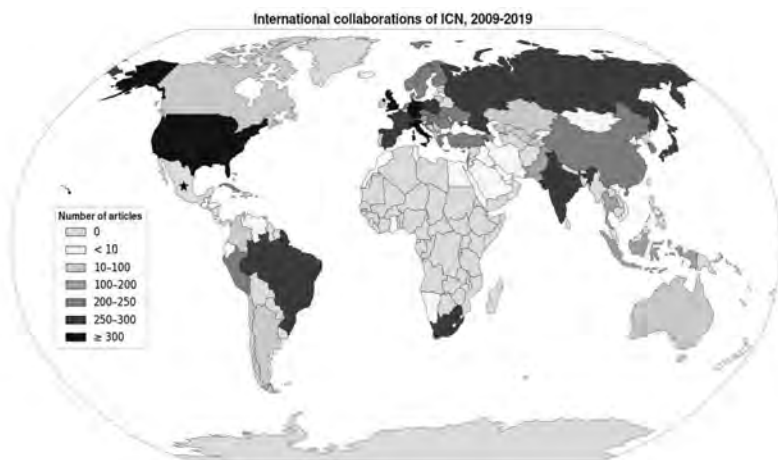


Figure 10. International scientific collaborations of ICN in the 2009-2019 period. The shaded categories indicate the number of refereed articles with at least one non-Mexican co-author

Analysis of keyword frequencies in academic publications

Another interesting analysis that can be extracted (and in the future automated) from the SIGI database is semantic analysis of the frequencies of keywords associated with academic publications. The keywords of 2465 refereed journal articles published by our example institution in the 2009-2019 period were extracted from our ICN database. We carried out an effort to manually normalize these keywords by identifying synonyms and abbreviations (such as “QFT” and “quantum field theory”) and grouping them into synonymic sets. Keywords of little semantic value such as “origin”, “generation” and “objects” were also removed from the analysis. The frequencies of the resulting keywords were then counted for articles published in two separate time periods, from 2009 to 2013 and from 2014 to 2019. Comparing the analyses over two periods lets us track the evolution of research interests over time while maximizing the statistics in each period.

Figure 11 shows two-word clouds of the 100 most common keywords found in all publications in the two time periods. It is made clear that the creation of theoretical models is the main academic focus of the institute and that this remains a constant through time. We also recognize the specific main fields of research of the institute by noticing keywords that remain frequent in both time periods: ISM jets & outflows (astrophysics), field theory, quantum mechanics and states, general relativity, gravitation (theoretical physics) and irradiation and thermoluminescence (chemistry).

Another interesting result is detecting keywords that either gained or lost prominence between these two time periods. One example is “carbon nanotubes”, which was the 7th most frequent keyword in the first period but fell to the 30th in the second, echoing the global decline in interest in this research topic after its peak in the first decade of the 21st century. Another example is “Mars”, which rose to become the most frequent keyword in the second period while not even appearing in the top 100 in the first period. In this case this reflects the simple fact that this keyword appears in NASA collaborations related to the Curiosity rover, which landed on Mars in 2012.



Figure 11. Word clouds showing the most frequent keywords in academic publications of the Institute for the 2009-2014 (left) and 2015-2019 (right) time periods

CONCLUSIONS

The SIGI system is a valuable tool for Research Data Management that allows the collection, storage, validation and analysis of all academic activity of a teaching-research institution. Based mainly on the data captured by the researchers as part of their annual reports, it integrates information from traditional scientific production with other aspects of academic activity into a user-friendly web-based application. From this rich database, SIGI automatically calculates a variety of metric indicators that are useful to fulfill evaluation requirements, as well as preparing data for further scientometric analysis.

In this article we described the ideas behind the genesis of the SIGI system, described the software architecture and the main modules and services, and showcased some of the generated scientometric indicators and analyses that can be performed. From this database SIGI builds an integral characterization of the academic trajectory of individual researchers, departments and the institution as a whole. This characterization goes beyond the main academic production metrics to include other activities that also constitute important parts of academic life.

Using a database generated for a particular research institute, the ICN-UNAM, we showed some of the main data products and indicators produced by SIGI. These include automatically generated public researcher profiles as well as various metric indicators and statistics that summarize the Institute's academic activity. We showed how from this information it is possible

to conduct further analyses, such as mapping the internal and external collaboration networks. We also showed other useful indicators not centered on direct academic production, such as participation in academic events and scholarly visits, as well as those associated with teaching and students (which is an important academic activity in all research universities), including courses taught and graduated students.

This paper gives an overview of the design philosophy, the implementation details and the actual results obtained with the new system. It is aimed at providing a valuable example for other academic libraries that wish to embark on the RDM initiative even if they do not have access to the resources that inter-institutional consortia or national efforts can leverage. In the future, we propose to support the implementation of the SIGI to study and manage the indicators of other institutions.

The potential impact of SIGI is clear from our institute's experience with the system, as not only the researchers but also the outreach and administrative departments of the institute have benefited from its information products and services. The system's adaptability and flexibility has allowed our small academic library to implement the RDM principles, with tangible benefits to our community.

We hope that this study can serve as an example of the creation and operation of RDM services by specialized units within academic libraries in Latin America. Our experience in developing this project, and the presented details on design, implementation and results, might be a useful reference for other small academic libraries seeking to assume a more central role in the RDM philosophy without a great expenditure of resources. Furthermore, the collected research data and generated metric indicators are by themselves interesting from a scientometric point of view, as shown by the two scientometric analysis examples that we presented.

REFERENCES

- Arciniegas Tinjacá, Eliana, Yury Marcela Gómez Gutiérrez, and Orlando Gregorio-Chaviano. 2018. "La biblioteca universitaria y su rol en los procesos de investigación: una mirada desde los servicios de información con enfoque bibliométrico en Colombia." *Biblios. Journal of Librarianship and Information Science* 72: 113-129.
<https://doi.org/10.5195/biblios.2018.439>
- Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. 2017. "Developments in research data management in academic libraries: Towards an understanding of research data service maturity." *Journal of the Association for Information Science and Technology* 68 (9): 2182-2200.
<https://doi.org/10.1002/asi.23781>

- Guillot Jiménez, Javier. 2011. "SIGUA: Sistema Informático de Gestión Universitaria." Presented at the 14th International Congress on Informatics for Education, Havana, Cuba, 2011.
<https://doi.org/10.13140/RG.2.2.17802.06084>
- González Fernández-Villavicencio, Nieves. 2017. "Unidades de Bibliometría y bibliotecas universitarias: hacia la transparencia." *ThinkEPI* 11 (1): 86-94.
<https://bit.ly/2PCSWcm>
- Humphries, Mark and Kevin Gurney. 2008. "Network 'small-world-ness': a quantitative method for determining canonical network equivalence." *PLoS One* 3 (4): e0002051.
<https://doi.org/10.1371/journal.pone.0002051>
- OECD (Organisation for Economic Cooperation and Development). 2015. *Manual de Frascati 2015: Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental*. Paris/FEYCT, Madrid: OECD Publishing.
<https://doi.org/10.1787/9789264310681-es>
- OECD and Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition*. (The Measurement of Scientific, Technological and Innovation Activities). Paris/Eurostat, Luxembourg: OECD Publishing.
<https://doi.org/10.1787/9789264304604-en>
- Peralta González, María Josefa, Maylín Frías Guzmán, and Orlando Gregorio-Chaviano. 2015. "Criterios, clasificaciones y tendencias de los indicadores bibliométricos en la evaluación de la ciencia." *Revista Cubana de Información en Ciencias de la Salud* 26 (3): 290-309.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-2113201500030000&lng=es
- Research Councils United Kingdom. 2015. *Guidance on best practice in the management of research data*. United Kingdom: Research Councils U.K.
<https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-Guidance-BestPracticeManagementResearchData.pdf>
- Royal Ministry of Education and Research. 2004. *CRISTin (Current Research Information System in Norway)*.
<https://www.cristin.no/teknisk/xsd/institusjonsdata/>
- Schöpfel, Joachim, Hélène Prost, and Violane Rebouillat. 2017. "Research data in current research information systems." *Procedia Computer Science* 106: 305-320.
<https://doi.org/10.1016/j.procs.2017.03.030>
- Sierra Flores, María Magdalena, María Victoria Guzmán, Alejandro Raga, and Ignacio Pérez. 2009. "The productivity of Mexican astronomers in the field of outflows from Young stars." *Scientometrics* 81: 765.
<https://doi.org.pbdi.unam.mx:2443/10.1007/s11192-008-2264-8>
- Tarrats Pons, Elisenda. 2012. "Sitkis: una herramienta bibliométrica para el desarrollo del estado de la cuestión." *Textos Universitaris de Biblioteconomia i Documentació* 28: 1-16.
<https://doi.org/10.1344/105.000001831>
- Toft, Martin. 2003. "Sparkar i gang Frida." *Nettavis for Universitetet i Oslo UNI-FORUM Okt.*
<https://www.uniforum.uio.no/nyheter/2003/10/sparkar-i-gang-frida.html>

- Torres-Salinas, Daniel and Evaristo Jiménez-Contreras. 2012. "Hacia las unidades de bibliometría en las universidades: modelo y funciones." *Revista Española de Documentación Científica* 35 (3): 469-480.
<https://doi.org/10.3989/redc.2012.3.959>
- UNAM (Universidad Nacional Autónoma de México). 2015. "Acuerdo por el que se establecen los Lineamientos Generales de la Política de Acceso Abierto de la Universidad Nacional Autónoma de México." *Gaceta UNAM*, 4721, Convocatorias, September 10, 29-30.
https://www.unamenlinea.unam.mx/files/TUL_AcuerdoLineamientosGeneralesPoliticaAccesoAbierto.pdf

Para citar este texto:

- Sierra Flores, María Magdalena, Romel Calero Ramos, María Victoria Guzmán Sánchez, Juan Claudio Toledo-Roy y Yolsy Gabriela Gamboa Calderón. 2021. "SIGI, an integrated scientometric and curricular information system for teaching-research institutions". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 111-131.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58431>

¿Son las investigadoras de Documentación más interdisciplinarias frente a sus colegas hombres? Un análisis aplicado al profesorado español

Manuel Blázquez Ochando*

Michela Montesi*

Isabel Villaseñor Rodríguez*

Artículo recibido:
29 de marzo de 2021

Artículo aceptado:
28 de junio de 2021

Artículo de investigación

RESUMEN

La actividad científica desempeñada por mujeres difiere bajo varios puntos de vista de la de los hombres, tanto en términos de producción como de impacto. En este trabajo, comparamos la actividad científica de 349 profesores y profesoras del área de Biblioteconomía y Documentación (ByD) afiliados con universidades españolas, con base en el análisis de su trayectoria de publicación en revistas vaciadas en Scopus. Se analizan las publicaciones de la muestra atendiendo a tres indicadores de interdisciplinariedad (variedad, equilibrio y disparidad), y midiendo la disparidad a través de métodos de Procesamiento del Lenguaje Natural

* Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid, España

manublaz@ucm.es

mmontesi@ucm.es

isavilla@ucm.es

(PNL) y Recuperación de Información (RI). A pesar de no encontrar diferencias importantes en las tres dimensiones, los datos obtenidos apuntan a la posible inadecuación de los índices de citas de reflejar las actividades académica desempeñadas por las mujeres en toda su variedad, así como a la necesidad de comparar la producción de hombres y mujeres con base en el análisis temático de los contenidos producidos.

Palabras clave: Interdisciplinariedad; Producción Científica; Sesgo de Género; Análisis Temático; Biblioteconomía y Documentación

Are Information Science female researchers more interdisciplinary than their male colleagues? An analysis of Spanish faculty

Manuel Blázquez Ochando, Michela Montesi and Isabel Villaseñor Rodríguez

ABSTRACT

The scientific activity of women differs in several ways from men's, and several studies have found differences in terms of production and impact. The present work compares the research activity of 349 LIS professors with a Spanish affiliation by studying their Scopus publications. The sample is analyzed from an interdisciplinary perspective in terms of variety, balance and disparity. Disparity was measured through methods of NLP and IR. Despite not finding important differences in the three dimensions for male and female professors, the data obtained point to the possible inadequacy of citation indexes with respect to the spectrum of academic activities carried out by women in all its variety, as well as the need to compare the production of men and women based on the thematic analysis of the contents produced.

Keywords: Interdisciplinarity; Scientific Production; Gender Bias; Thematic Analysis; Library and Information Science

INTRODUCCIÓN

Aunque de forma no consistente, la literatura especializada ha documentado diferencias en la labor científica desempeñada por las mujeres respecto a la de los hombres, tanto en términos de producción, especialización y colaboración, por el lado de la actividad, como en términos de citación y visibilidad, por el lado del impacto. Según la investigación previa, las mujeres investigadoras serían menos productivas (Larivière *et al.*, 2011; Larivière *et al.*, 2013; European Commission, 2018; Hernández-Martín *et al.*, 2019) y menos especializadas que los hombres (Leahey, 2006), mientras que sus patrones de colaboración serían más abiertos, cooperativos entre géneros e interdisciplinarios (Rhoten y Pfirman, 2007). Por el lado del impacto, la producción femenina, liderada por mujeres o realizada con una proporción mayor de mujeres tiende a citarse menos (Knobloch-Westerwick y Glynn, 2013; Chan y Torgler, 2020), incluso cuando se publica en revistas similares en términos de impacto (Beaudry y Larivière, 2016). Aunque los resultados de diferentes estudios no son coherentes en este sentido y algunos no detectan diferencias de impacto entre hombres y mujeres (Thelwall y Nevill, 2019), en general el trabajo de investigación femenino parece citarse con menor intensidad, por lo menos en muchas áreas. El sesgo de género se produce también en la comunicación científica en los medios sociales, donde la presencia de las mujeres sería en general menos visible que la de los hombres (Sugimoto *et al.*, 2017; Montesi, Rodríguez Villaseñor y Bittencourt Dos Santos, 2019). Las diferencias en términos de visibilidad e impacto, así como las documentadas en términos de producción y actividad, tienen repercusiones obvias en los procesos de evaluación, donde se enfatizan precisamente todas estas dimensiones. En el presente trabajo pretendemos ahondar en las peculiaridades de la actividad científica de las mujeres analizando las dinámicas de publicación de profesoras y profesores de Biblioteconomía y Documentación (en adelante ByD) afiliados con universidades españolas. Concretamente, nos centramos en los procesos de difusión del conocimiento y planteamos la cuestión de si las mujeres publican investigaciones más interdisciplinarias que los hombres. El trabajo de Jamali, Abbasi y Bornmann (2020) confirma la relación entre género e investigación interdisciplinaria (IID) en un conjunto de publicaciones del profesorado australiano en las áreas de Física, Química y Biología en el periodo 1980-2014, obtenidas de WoS. Sin embargo, la relación entre género e interdisciplinarietà no cuenta todavía con una base de evidencia sustancial.

Los paralelismos entre la IID y la investigación realizada por mujeres llevan a Rhoten y Pfirman (2007) a plantearse la cuestión de si existe un cruce

entre estos dos ámbitos. Según estas autoras, las mujeres se sienten atraídas naturalmente hacia la IID por una mayor inclinación a buscar conexiones entre ideas y contextos, sus menores vínculos institucionales con las normas del sistema de producción de conocimiento articulado en disciplinas, y por una epistemología basada en la racionalidad “afectiva”, la interconexión y el holismo. Tanto la IID como la investigación realizada por mujeres son a menudo objeto de discriminación en los procesos de revisión y evaluación. A través del caso de más de 18 000 solicitudes enviadas al *Australian Research Council's Discovery Programme*, Bromham, Dinnage y Hua (2016) demuestran esta desventaja para la IID pues encontraban que, a mayor grado de interdisciplinariedad de las propuestas, correspondía una menor probabilidad de financiación. Tanto la investigación llevada a cabo por mujeres como la IID pueden tener salida en revistas de menor impacto o recibir menos citas en casos como, por ejemplo, la investigación sobre las aguas subterráneas (Barthel y Seidl, 2017). Nielsen (2017) documenta diferencias a favor de los hombres en los procesos de evaluación de la comunidad científica danesa, que achaca a la escasa participación de las mujeres en los comités evaluadores, por un lado, y a la ventaja que el sistema asigna a la investigación colaborativa.

Por otro lado, existe un énfasis importante en promover desde la política tanto la IID como la investigación de mujeres (Rhoten y Pfirman, 2007). Se supone que ambos tipos de investigación pueden contribuir a la resolución de problemas del mundo real y así lo entienden muchas agencias financiadoras (Jacobs y Frickel, 2009). Repasando cuestiones pendientes con respecto a la IID, Jacobs y Frickel (2009: 52) constataban que la “interdisciplinarity is gendered” y que igual de importantes que las motivaciones sociales que llevan a promocionarla son las consecuencias de ésta, haciendo referencia en particular a los *women studies* y a otros dominios de IID marginalizados. En este sentido, cabe añadir que tanto la IID como la investigación de mujeres presentan niveles de productividad inferiores respecto al resto de investigaciones (Leahey, Beckman y Stanko, 2017) y, aunque el mayor calado social de estos dos tipos de investigación no se suele mencionar para explicar los niveles de producción inferiores, tampoco existen evidencias de lo contrario.

Finalmente, los propios patrones de colaboración de las mujeres podrían apuntar a una mayor inclinación hacia la diversidad respecto a los hombres, pues, en algunos estudios, tienden a ser más inclusivos en términos de género (Ghiasi, Harsh y Schiffauerova, 2018; Araújo *et al.*, 2017; Whittington, 2018), pudiendo resultar en una actitud más inclusiva también en términos disciplinarios. Asimismo, la colaboración de las mujeres sería más local, debido, en parte, a las necesidades de la conciliación familiar (Larivière *et al.*, 2011; Abramo, D'Angelo y Murgia, 2013). Cabe preguntarse si también son igual de

locales y cercanos los problemas de investigación que se plantean las mujeres investigadoras, así como las repercusiones de su investigación en la realidad más próxima; la investigación de Abramo, D'Angelo y Di Costa (2019) apoya, aunque moderadamente, la relación entre localidad de las redes de colaboración y la inclinación hacia una investigación más diversificada.

En el marco de estos paralelismos nos planteamos la cuestión de si las mujeres investigadoras son más interdisciplinarias que los hombres. En el ámbito de la Lingüística y la Sociología, Leahey (2006) encontraba que la mayor productividad de los hombres podía achacarse a su mayor especialización frente a la mayor diversificación de las mujeres. Por otro lado, Abramo, D'Angelo y Di Costa (2018), tras analizar la producción de investigadores italianos de Ciencias en el periodo 2004-2008, concluían que las diferencias entre hombres y mujeres en términos de producción no podían explicarse por una mayor o menor inclinación hacia la diversificación de las temáticas estudiadas. En Woolley *et al.* (2014) tampoco el género influye en la inclinación hacia la IID; sin embargo, otros estudios han comprobado la existencia de una conexión entre género e interdisciplinariedad (Van Rijnsoever y Hessels, 2011; Araújo *et al.*, 2017; Meng y Shapira, 2010) y recientemente Jamali, Abbasi y Bornmann (2020) concluían que las mujeres de universidades australianas presentaba una mayor inclinación a publicar investigación diversificada temáticamente en las áreas de Biología, Física y Química.

Medición de la interdisciplinariedad

Existen diferentes formas de analizar y medir la interdisciplinariedad de autores, publicaciones, revistas u otras entidades dentro del sistema de comunicación científica. Para determinar la pertenencia a cierto ámbito disciplinario, se suelen emplear los sistemas de clasificación en categorías de WoS y Scopus (que contemplan respectivamente cerca de 250 y 340 categorías respectivamente), u otros sistemas de clasificación como el *ECOOM Leuven-Budapest Subject Classification* que Zhang, Rousseau y Glänzel (2016) y Zhang *et al.* (2018) utilizan por incluir un número ni demasiado grande ni demasiado pequeño de grandes áreas (16) y sub-áreas (68), o el sistema de clasificación del profesorado italiano en 370 sectores científicos y 14 grandes áreas disciplinarias que aprovechan Abramo, D'Angelo y Zhang (2018). Solomon, Carley y Porter (2016) explican que las métricas de interdisciplinariedad pueden medir tanto la “integración” de conocimiento en el ámbito de las referencias citadas en artículos individuales –y, por extensión, en revistas, áreas de investigación o instituciones– como la “difusión” de conocimiento, en términos de citación de los artículos de un conjunto documental concreto

por parte de revistas pertenecientes a categorías diferentes. En ambos casos, la medición de la interdisciplinariedad suele atender a tres dimensiones diferentes de la integración y difusión del conocimiento: la variedad o el número de diferentes categorías en las cuales se asignan los elementos del sistema (referencias, citas o publicaciones, por ejemplo); el equilibrio o la uniformidad en la distribución de los elementos del sistema en las diferentes categorías, y la distancia o la disparidad entre los elementos analizados, se observa mayor disparidad entre las referencias, citas o publicaciones que se producen en categorías muy poco afines y distantes la una de la otra. Las tres dimensiones se pueden integrar en un único indicador, como el indicador *Rao-Stirling Diversity* (Porter y Rafols, 2009) o el *Integrated Diversity Index*, desarrollado en el trabajo de Zhang, Rousseau y Glänzel (2016) y de Zhang *et al.* (2018), o se pueden analizar por separado para determinar, por ejemplo, el efecto de cada una de ellas en los patrones de citación de los conjuntos documentales analizados (Wang, Thijs y Glänzel, 2015). Una vez calculados los diferentes indicadores, integrados o no, se pueden recabar estadísticas descriptivas para los conjuntos documentales analizados y ordenar los elementos examinados.

Debido a las varias posibles interpretaciones de la diversidad, los diferentes indicadores pueden devolver resultados poco coherentes (Huang *et al.*, 2021). Zhang *et al.* (2018) afirman que no existe de momento una metodología definitiva para medir la interdisciplinariedad, al igual que Adams, Loach y Szomszor (2016) y Leydesdorff y Rafols (2011). Como consecuencia de estas limitaciones y de los resultados incoherentes encontrados en estudios previos, Zhang *et al.* (2018) recomiendan revertir los enfoques de análisis interdisciplinario y, en lugar de contar con clasificaciones predefinidas, sugieren realizar análisis de abajo-arriba, incluyendo, entre varias posibilidades de análisis, la interpretación de los contenidos. En este sentido se orientan los estudios de Gomes y Dewes (2017) y Dong *et al.* (2018).

Objetivo

El objetivo del presente trabajo es comparar desde un punto de vista bibliométrico la actividad científica de 349 profesores y profesoras del área de ByD atendiendo a sus patrones de publicación. Si las políticas científicas de género miran a igualar hombres y mujeres, las críticas que reciben los criterios de evaluación normalmente adoptados (productividad e indicadores basados en citación como el factor de impacto) animan a buscar alternativas de evaluación que permitan al colectivo femenino poner en valor ciertas características de su actividad investigadora. Nuestra hipótesis de partida es que pueden existir diferencias en términos de interdisciplinariedad entre las

publicaciones de las profesoras y las correspondientes a profesores, y pretendemos comprobarlo a través del análisis de su producción científica recopilada en Scopus. Esta base de datos se escogió debido a su mejor cobertura de la producción en idiomas diferentes al inglés.

METODOLOGÍA

En el presente estudio, trabajamos con la muestra de 183 profesoras y 166 profesores afiliados con universidades españolas y analizada en un trabajo previo (Montesi, Rodríguez Villaseñor y Bittencourt Dos Santos, 2019). Los 349 individuos de la muestra se buscaron automática y manualmente en la base de datos Scopus para localizar las publicaciones que les correspondían desde 1980 hasta febrero de 2019. Se tuvieron en cuenta todos los tipos de documentos reflejados, en concreto libros y capítulos de libros, artículos en revistas científicas y revistas profesionales, y actas de congresos. Para determinar los indicadores de variedad y equilibrio, se utilizaron las 334 categorías que contempla Scopus, mientras que para la disparidad se atendió a las mismas categorías y además al título y resumen de cada documento. Cuando las revistas están clasificadas en más de una categoría se tuvieron en cuenta todas las categorías en las que aparecía. A partir de esta clasificación y de la descripción de contenido de los documentos, se obtuvieron tres indicadores para todos los individuos de la muestra: variedad, equilibrio y disparidad. La variedad mide la proporción de categorías no vacías para cada individuo en los dos conjuntos, es decir, la proporción de categorías en las cuales cada autor o autora tiene trabajos publicados sobre el total de categorías. En segundo lugar, el coeficiente de Gini se utilizó para medir el nivel de concentración de los trabajos de cada autor en las diferentes categorías temáticas. Normalmente, valores más altos del coeficiente de Gini denotan mayor desequilibrio y concentración de publicaciones en unas pocas categorías temáticas. Por eso, en los resultados, mostramos *1-Gini* para que el indicador tenga la misma dirección que los demás y los valores más próximos al 1 denotan mayor interdisciplinariedad. Finalmente, la disparidad de las publicaciones de cada autor se midió atendiendo al contenido de cada artículo. Esto se realizó evaluando el título y el resumen de cada publicación en relación con el resto de artículos de la categoría temática, mediante técnicas de PNL y RI, con base en las cuales se determinó un coeficiente de disparidad de la producción científica de cada individuo (ecuación 1).

$$D = 1 - \frac{\log_{10}(\sum_{ij} SimCos(d_i, d_j))}{N} \quad (1)$$

Donde:

SimCos = Similaridad del coseno

i = Categorías temáticas del autor o autora del que se estudia la disparidad en particular, sin repetición

j = Conjunto de categorías temáticas en las que ha publicado artículos cada individuo de la muestra, sin repetición

d_i = *n* documentos de la categoría *i*

d_j = *n* documentos de la categoría *j*

N = Número total de categorías temáticas en las que han sido clasificados los documentos de todos los autores y autoras de las dos muestras, sin repetición

La disparidad temática de un autor es igual a 1 menos la fracción del logaritmo en base 10 de la suma de los coeficientes de similaridad entre los documentos (calculado con base en su título y resumen) publicados por el autor en una categoría temática dada con respecto al resto de documentos de dicha categoría, publicados por el resto de autores de la muestra a estudiar, dividido por el número total de categorías en las que se han clasificado los documentos de la muestra, conformada por todos los autores estudiados. El resultado que se obtiene es la similaridad absoluta de la producción científica de cada individuo con respecto al resto de la muestra. Al restar 1 a ese coeficiente de similaridad se obtiene la diferencia, es decir, la disparidad.

Los coeficientes obtenidos por cada autor y autora se sometieron a las pruebas de significación pertinentes (prueba U de Mann-Whitney) con el objetivo de rechazar la hipótesis nula de no diferencia entre los dos conjuntos, a través de la herramienta IBM SPSS Statistics. Se escogieron pruebas estadísticas no paramétricas atendiendo a la distribución no normal del conjunto de datos manejados, que no cumplió con los requisitos de normalidad definidos en Stjernholm Madsen (2016: 55-60).

RESULTADOS

Según puede apreciarse en la *Tabla 1*, 55 % de las profesoras y 65 % de los profesores tienen por lo menos una publicación en Scopus. Aparte de la diferencia en términos de presencia, es posible apreciar otra en términos de producción, siendo la mediana de publicaciones por individuo inferior en el

caso de las mujeres (5 frente a 8.5). La diferencia entre los dos grupos en términos de productividad se acerca mucho a niveles de significación (U de Mann-Whitney = 4603, valor de $p = .051$).

	M	H
Número total de individuos en la muestra	183	166
Profesorado con presencia en Scopus	101	108
% Profesorado con presencia en Scopus	55 %	65 %
Número total de publicaciones	666	960
Mediana de publicaciones por persona	5	8.5
Promedio de publicaciones por persona	8.48	14.28

Tabla 1. Datos globales relativos a la muestra estudiada

Según puede apreciarse en los diagramas de las Figuras 1a y 1b, las publicaciones de profesores y profesoras se reparten de forma relativamente parecida en las cinco macrocategorías contempladas por Scopus (*Multidisciplinary*, *Health Sciences*, *Life Sciences*, *Physical Sciences*, *Social Sciences*) y en la categoría de ByD. Se aprecia una mayor inclinación de las mujeres por el ámbito tecnológico que se encuentra recogido en *Physical Sciences* (34 % de las publicaciones frente a 29 % de los hombres), incluso por encima de ámbitos de conocimiento más cercanos como las *Social Sciences* (23 % frente a 27 % de los hombres), algo que puede explicarse por una tendencia de las mujeres a trabajar en ámbitos más nuevos del conocimiento (Rhoten y Pfirman, 2007) y que también puede apuntar a una vocación más aplicada de éstas. Por lo demás, las diferencias en la actividad de publicación entre profesores y profesoras, a este nivel, resultan mínimas.



Figura 1a. Publicaciones por grandes áreas – Mujeres

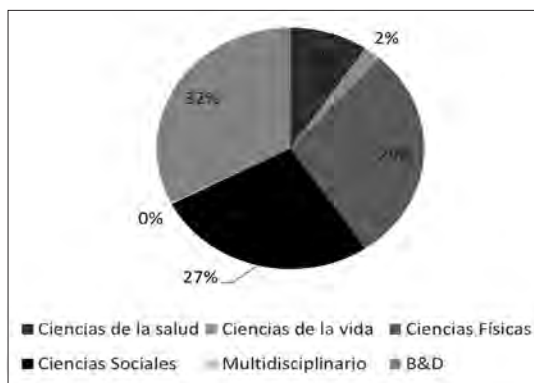


Figura 1b. Publicaciones por grandes áreas – Hombres

La *Tabla 2* muestra otros indicadores generales de actividad de los dos conjuntos. Las profesoras publican proporcionalmente menos en acceso abierto y proporcionalmente más bajo licencia de acceso. En promedio, las mujeres incluyen aproximadamente las mismas referencias en sus trabajos que los hombres; sin embargo, necesitan más referencias para obtener una cita respecto a los hombres. Es decir, que para obtener el mismo resultado que los hombres necesitan realizar un mayor trabajo de recopilación de fuentes y de fundamentación de su investigación.

Producción y actividad	M	H
Número total de publicaciones de acceso abierto	195	319
% de publicaciones en acceso abierto	29.3 %	33.2 %
Número total de publicaciones bajo licencia de acceso	471	641
% de publicaciones bajo licencia de acceso	70.7 %	66.8 %
Número total de referencias bibliográficas	17 285	24 126
Media de referencias bibliográficas por publicación	25.95	25.13
Media de referencias bibliográficas por citas recibidas	3.9	2.46

Tabla 2. Datos generales de producción y actividad en revistas de Scopus

La *Tabla 3* muestra los tres indicadores de interdisciplinariedad encontrados para los dos conjuntos de autores y autoras. En general, los hombres presentan una ligera ventaja en variedad, equilibrio y disparidad, aunque, como conjunto, presentan mayores diferencias internas y valores atípicos, atendiendo a la desviación estándar y al rango intercuartílico, con lo cual las diferencias entre hombres y mujeres parecen mínimas.

	Variedad		Equilibrio		Disparidad	
	M	H	M	H	M	H
Media	0.0182	0.0239	0.200	0.249	0.721	0.725
DE	0.0173	0.0241	0.184	0.234	0.172	0.197
Mediana	0.0120	0.0149	0.175	0.189	0.756	0.760
Rango int.	0.0180	0.0240	0.240	0.313	0.213	0.171

Tabla 3. Medidas de interdisciplinariedad

Las *Figuras 2a, 2b y 2c* muestran la distribución de los dos conjuntos en los tres indicadores, poniendo de manifiesto una mayor dispersión en el colectivo masculino respecto al femenino. En cuanto a la variedad, si ambos conjuntos se concentran en valores próximos al cero, delatando la tendencia a publicar en unas pocas categorías, esta inclinación es ligeramente más marcada en el conjunto femenino (*Figura 2a*).

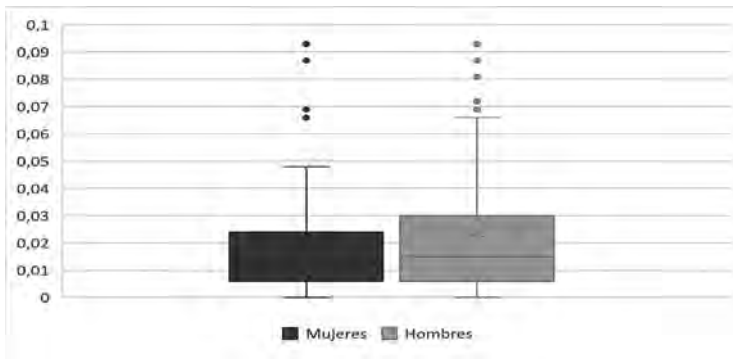


Figura 2a. Distribución de los coeficientes de variedad en los dos conjuntos

Lo mismo se aprecia en términos de equilibrio, y los dos conjuntos de profesores y profesoras no sólo publican en una proporción limitada de categorías, sino que además tienden a concentrar su producción en algunas, siendo esta tendencia más marcada en las mujeres (*Figura 2b*).

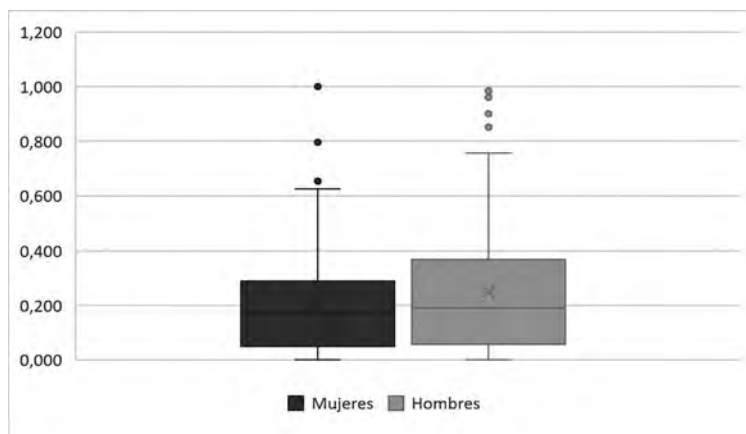


Figura 2b. Distribución de los coeficientes de equilibrio en los dos conjuntos

Finalmente, el análisis de la disparidad revela un conjunto de autoras y autores con líneas temáticas sustancialmente diferenciadas del resto, puesto que en este caso los coeficientes se sitúan más cerca del 1 que del 0, algo más en el conjunto masculino (*Figura 3*).

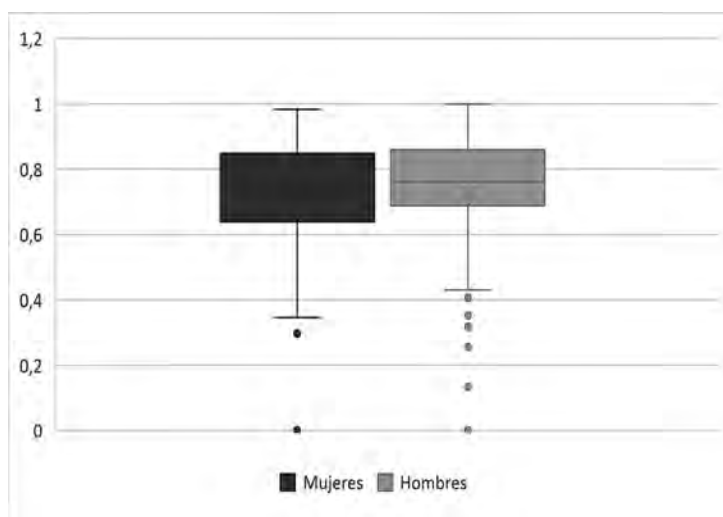


Figura 2c. Distribución de los coeficientes de disparidad en los dos conjuntos

Para rechazar la hipótesis nula de igualdad entre profesores y profesoras en las tres variables de variedad, equilibrio y disparidad, se realizó la prueba U de Mann Whitney, cuyos resultados se muestran en la *Tabla 4*.

	U de Mann-Whitney	Valor de <i>p</i>
Disparidad	5095	.477
Variedad	4954	.251
Equilibrio	4960	.258

Tabla 4. Resultados de la prueba U de Mann-Whitney

En ningún caso pudo rechazarse la hipótesis nula de igualdad, situándose los valores de *p* por encima del 0,05.

Por otro lado, los coeficientes de variedad y equilibrio en la muestra completa (hombres y mujeres) resultaron correlacionar con el número de documentos (*Tabla 5*), creciendo a la vez que éstos, explicando en parte la ligera ventaja en esos indicadores de los hombres que en la muestra son más productivos.

	Coefficiente Rho de Spearman	Valor de <i>p</i>
Disparidad	.034	.628
Variedad	.766	.000
Equilibrio	.747	.000

Tabla 5. Resultados de las pruebas de correlación

DISCUSIÓN Y CONCLUSIONES

Esta investigación pretendía documentar diferencias en la actividad investigadora de las profesoras frente a los profesores para detectar posibles diferencias cualitativas en la producción cuantitativamente más escasa del colectivo femenino. Con este fin, nos hemos planteado la pregunta de si la producción científica femenina puede considerarse más interdisciplinaria y diversificada temáticamente frente a la masculina, analizando la actividad de publicación del profesorado de ByD con filiación en España y reflejada en Scopus. Las mujeres presentan una ligera desventaja frente a los hombres en términos de presencia en la base de datos, número de publicaciones y proporción de trabajos publicados en abierto, mientras que parecen necesitar un mayor número de referencias para obtener una cita respecto a los hombres. Además, tienden a publicar un poco más que los hombres en revistas del área *Physical Sciences*. En cuanto a las diferencias relacionadas específicamente

con la interdisciplinariedad, los resultados muestran que la producción científica femenina no parece más interdisciplinaria que la masculina, por lo menos atendiendo a las categorías de Scopus y al análisis de la diversidad temática de títulos y resúmenes. Los indicadores de variedad y equilibrio tienden a crecer con el número de publicaciones, acentuando el sesgo de productividad a favor de los hombres que tienden a ser más productivos. Por otro lado, la similitud de los indicadores correspondientes a los dos colectivos puede interpretarse también como una característica propia de la disciplina que, en el estudio de Dong *et al.* (2018), se configura además como un ámbito cada vez más interdisciplinario.

También es importante tener en cuenta las limitaciones del presente estudio, empezando por el propio sistema de clasificación disciplinaria (Scopus), considerando que, según Huang *et al.* (2021), a veces se necesitan sistemas de clasificación muy granulares para poder apreciar la interdisciplinariedad. Otra limitación que cabe mencionar es la propia manera en la que se ha operacionalizado el concepto de disparidad en el presente estudio. Si por un lado se introdujeron técnicas propias del PNL y la RI para poder atender al contenido temático de la producción de cada autora y autor, en línea con las recomendaciones de la investigación más reciente, por otro lado el análisis se tradujo en la medición estadística de la disparidad de categorías, títulos y resúmenes más que en la valoración de los textos completos. En este sentido, en futuras investigaciones la medición de la disparidad debería buscar fórmulas que reflejen más fielmente las posibles diferencias en el contenido temático de publicaciones de hombres y mujeres, a ser posible trabajando con textos completos. Igualmente, sería recomendable ahondar en las temáticas concretas de investigación tratadas por las mujeres de forma cualitativa y no sólo cuantitativa. El origen de los temas de investigación de las mujeres está a menudo en las experiencias e interacciones de sus vidas de todos los días, chocando con la cultura académica más tradicional centrada en la literatura y convenciones disciplinarias (Gonzales, 2018), y es de esperar que las mujeres presenten perfiles investigadores temáticamente diferentes a los de los hombres.

Finalmente, otras razones que pueden haber impedido una diferenciación clara del trabajo de investigación femenino en el presente estudio incluyen las diferencias en el número de publicaciones analizadas para hombres y mujeres, los sesgos del propio sistema de clasificación utilizado, y el hecho de que hemos analizado una dimensión de la actividad investigadora, la publicación en revistas indexadas, que, por naturaleza, puede ser más conservadora y conforme con las normas de un sistema de comunicación científica que, en cierta literatura de corte feminista, se consideran como masculinas y basadas en valores, como la competición y la búsqueda del estatus, más favorecedores para

los hombres (Gunter, 2009). En este sentido, se puede añadir que reducir la actividad científica a la publicación en revistas recopiladas en bases de datos como Scopus y WoS puede resultar perjudicial para las mujeres que cada vez más van asumiendo tareas de gestión y docencia a expensas de la investigación o que se implican a menudo en el trabajo académico comprometido con la comunidad (Gonzales, 2018). Para el futuro, es importante demostrar el impacto de actividades académicas y de investigación alternativas a la publicación formal en revistas indexadas, especialmente de aquellas que permitan poner en valor la participación femenina en los procesos de generación y transmisión del conocimiento.

REFERENCIAS

- Abramo, Giovanni, Ciriaco D'Angelo y Flavia Di Costa. 2018. "The effects of gender, age and academic rank on research diversification". *Scientometrics* 114 (2): 373-387.
- Abramo, Giovanni, Ciriaco D'Angelo y Flavia Di Costa. 2019. "Authorship analysis of specialized vs diversified research output". *Journal of Informetrics* 13 (2): 564-573.
- Abramo, Giovanni, Ciriaco D'Angelo y Gianluca Murgia. 2013. "Gender differences in research collaboration". *Journal of Informetrics* 7 (4): 811-822.
- Abramo, Giovanni, Ciriaco D'Angelo y Lin Zhang. 2018. "A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list". *Journal of Informetrics* 12 (4): 1182-1193.
- Adams, Jonathan, Tamar Loach y Martin Szomszor. 2016. "Interdisciplinary research: Methodologies for identification and assessment. Digital research reports". Digital Science, London.
- Araújo, Eduardo B., Nuno A. M. Araújo, André A. Moreira, Hans J. Herrmann y Jose S. Andrade. 2017. "Gender differences in scientific collaborations: Women are more egalitarian than men". *PloS one* 12 (5): e0176791.
- Barthel, Roland y Roman Seidl. 2017. "Interdisciplinary collaboration between natural and social sciences—Status and trends exemplified in groundwater research". *PloS one* 12 (1): e0170754.
- Beaudry, Catherine y Vincent Larivière. 2016. "Which gender gap? Factors affecting researchers' scientific impact in science and medicine". *Research Policy* 45 (9): 1790-1817.
- Bromham, Lindell, Russell Dinnage y Xia Hua. 2016. "Interdisciplinary research has consistently lower funding success". *Nature* 534 (7609): 684.
- Chan, Ho Fai y Benno Torgler. 2020. "Gender differences in performance of top cited scientists by field and country". *Scientometrics* 125 (3): 2421-2447.
- Dong, Kun, Haiyun Xu, Rui Luo, Ling Wei y Shu Fang. 2018. "An integrated method for interdisciplinary topic identification and prediction: a case study on information science and library science". *Scientometrics* 115 (2): 849-868.

- European Commission. Directorate-General for Research and Innovation. 2018. *She Figures 2018*. Fecha de consulta: 27 de febrero de 2020. <https://op.europa.eu/en/publication-detail/-/publication/9540ffa1-4478-11e9-a8ed-01aa75ed71a1/language-en>
- Ghiasi, Gita, Matthew Harsh y Andrea Schiffauerova. 2018. "Inequality and collaboration patterns in Canadian nanotechnology: implications for pro-poor and gender-inclusive policy". *Scientometrics* 115 (2): 785-815.
- Gomes, Janaína y Homero Dewes. 2017. "Disciplinary dimensions and social relevance in the scientific communications on biofuels". *Scientometrics* 110 (3): 1173-1189.
- Gonzales, Leslie D. 2018. "Subverting and minding boundaries: The intellectual work of women". *The Journal of Higher Education* 89 (5): 677-701.
- Gunter, Ramona. 2009. "The Emergence of Gendered Participation Styles in Science-Related Discussions: Implications for Women's Place in Science". *Journal of Women and Minorities in Science and Engineering* 15 (1): 53-75.
- Hernández-Martín, Estela, Fernando Calle, Juan C. Dueñas, Miguel Holgado y Asunción Gómez-Pérez. 2019. "Participation of women in doctorate, research, innovation, and management activities at Universidad Politécnica de Madrid: analysis of the decade 2006–2016". *Scientometrics* 120 (3): 1059-1089.
- Huang, Ying, Wolfgang Glänzel, Bart Thijs, Alan L. Porter y Lin Zhang. 2021. "The comparison of various similarity measurement approaches on interdisciplinary indicators". *FEB Research Report MSI_2102*, 1-24.
- Jacobs, Jerry A. y Scott Frickel. 2009. "Interdisciplinarity: A critical assessment". *Annual review of Sociology* 35, 43-65.
- Jamali, Hamid R., Alireza Abbasi y Lutz Bornmann. 2020. "Research diversification and its relationship with publication counts and impact: A case study based on Australian professors". *Journal of Information Science* 46 (1): 131-144.
- Knobloch-Westerwick, Silvia y Carroll J. Glynn. 2013. "The Matilda effect—Role congruity effects on scholarly communication: A citation analysis of Communication Research and Journal of Communication articles". *Communication Research* 40 (1): 3-26.
- Larivière, Vicent, Chaoqn Ni, Yves Gingras, Blaise Cronin y Cassidy Sugimoto. 2013. "Bibliometrics: Global gender disparities in science". *Nature News* 504 (7479): 211.
- Larivière, Vicent, Etienne Vignola-Gagné, Christian Villeneuve, Pascal Gélinas y Yeves Gingras. 2011. "Sex differences in research funding, productivity and impact: an analysis of Québec university professors". *Scientometrics* 87 (3): 483-498.
- Leahey, Erin. 2006. "Gender differences in productivity: Research specialization as a missing link". *Gender & Society* 20 (6): 754-780.
- Leahey, Erin, Christine M. Beckman y Taryn L. Stanko. 2017. "Prominent but less productive: The impact of interdisciplinarity on scientists' research". *Administrative Science Quarterly* 62 (1): 105-139.
- Leydesdorff, Loet e Ismael Rafols. 2011. "Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations". *Journal of Informetrics* 5 (1): 87-100.
- Meng, Yu. 2018. "Gender distinctions in patenting: Does nanotechnology make a difference?" *Scientometrics* 114 (3): 971-992.
- Meng, Yu y Philip Shapira. 2010. "Women and patenting in nanotechnology: Scale, scope and equity", en *Nanotechnology and the challenges of equity, equality and development*, editado por Susan E. Cozzens y Jameson Wetmore, 23-46. Dordrecht: Springer.

- Montesi, Michela, Isabel Rodríguez Villaseñor y Fernando Bittencourt Dos Santos. 2019. "Presencia, actividad, visibilidad e interdisciplinariedad del profesorado universitario de Documentación en los medios sociales: una perspectiva de género". *Revista Española de Documentación Científica* 42 (4): 246.
- Nielsen, Mathias Wullum. 2017. "Gender consequences of a national performance-based funding model: new pieces in an old puzzle". *Studies in Higher Education* 42 (6): 1033-1055.
- Porter, Alan e Ismael Rafols. 2009. "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time". *Scientometrics* 81 (3): 719-745.
- Rhoten, Diana y Stephanie Pfirman. 2007. "Women in interdisciplinary science: Exploring preferences and consequences". *Research policy* 36 (1): 56-75.
- Solomon, Greg E. A., Stephen Carley y Alan L. Porter. 2016. "How multidisciplinary are the multidisciplinary journals Science and Nature?" *PloS one* 11 (4): e0152637.
- Stjernholm Madsen, B. 2016. *Statistics for non-statisticians*. 2a. ed. Berlín: Springer.
- Sugimoto, Cassidy R., Sam Work, Vincent Larivière y Stefanie Haustein. 2017. "Scholarly use of social media and altmetrics: A review of the literature". *Journal of the Association for Information Science and Technology* 68 (9): 2037-2062.
- Thelwall, Mike y Tamara Nevill. 2019. "No evidence of citation bias as a determinant of STEM gender disparities in US biochemistry, genetics and molecular biology research". *Scientometrics* 121 (3): 1793-1801.
- Van Rijnsvoever, Frank J. y Laurens K. Hessels. 2011. "Factors associated with disciplinary and interdisciplinary research collaboration". *Research policy* 40 (3): 463-472.
- Wang, Jian, Bart Thijs y Wolfgang Glänzel. 2015. "Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity". *PloS one* 10 (5): e0127298.
- Whittington, Kjersten Bunker. 2018. "A tie is a tie? Gender and network positioning in life science inventor collaboration". *Research Policy* 47 (2): 511-526.
- Woolley, Richard, Mabel Sánchez-Barrioluengo, Tim Turpin y Jane Marceau. 2014. "Research collaboration in the social sciences: What factors are associated with disciplinary and interdisciplinary collaboration?" *Science and Public Policy* 42 (4): 567-582.
- Zhang, Lin, Ronald Rousseau y Wolfgang Glänzel. 2016. "Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account". *Journal of the Association for Information Science and Technology* 67 (5): 1257-1265.
- Zhang, Lin, Beibei Sun, Zaida Chinchilla-Rodríguez, Lixin Chen y Ying Huang. 2018. "Interdisciplinarity and collaboration: on the relationship between disciplinary diversity in departmental affiliations and reference lists". *Scientometrics* 117 (1): 271-291.

Para citar este texto:

Blázquez Ochando, Manuel, Michela Montesi e Isabel Villaseñor Rodríguez. 2021. "¿Son las investigadoras de Documentación más interdisciplinarias frente a sus colegas hombres? Un análisis aplicado al profesorado español". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 133-149.

<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58433>

Interconexões temáticas em Ciência da Informação no ResearchGate: o caso de professores da Pós-graduação no Brasil

Bruna Lessa*
Hildenise Ferreira Novo*

Artículo recibido:
17 de abril de 2021

Artículo aceptado:
31 de agosto de 2021

Artículo de investigación

RESUMO

Analisa a interconexão temática nos títulos de trabalhos acadêmicos no domínio da Ciência da Informação publicados no *ResearchGate*. Como metodologia se utilizou a abordagem de análise de redes sociais, baseado em um sistema de representação do conhecimento - rede semântica, para potencializar os modos de visualização e representação da informação científica nesses ambientes, fundamentados na Teoria da Classificação Facetada (S.R. Ranganathan) e na Teoria do Conceito (Ingetraut Dahlberg) e associadas à base epistemológica de Gilles Deleuze e Félix Guattari. Foram analisados 324 perfis dos docentes dos 18 Programas de

* Departamento de Documentação e Informação, Instituto de Ciência da Informação, Universidade Federal da Bahia, Brasil
brunalessa@ufba.br
hnovo@ufba.br

Pós-graduação no domínio da Ciência da Informação, no Brasil para extrair títulos de suas pesquisas. Como resultado, a rede semântica criada possui caracterização topológica de rede de mundo pequeno, indicando convergência na interconexão entre os conceitos que compõem a rede, os quais estão representados e organizados em torno de um componente central. Conclui-se que os experimentos de validação da proposta sugerem um *continuum* na relação entre conceitos, evidenciando a interdisciplinaridade e evolução das pesquisas científicas. Uma estrutura geral de aplicação da análise de conceitos emergentes no *ResearchGate* é criada de modo a planejar o caminho teórico-metodológico apresentado para criação dessas redes.

Palavras-chave: Representação da Informação; Redes Semânticas; Comunicação Científica

Interconexiones temáticas en Ciencia de la Información en ResearchGate: el caso de profesores de posgrado en Brasil

Bruna Lessa e Hildenise Ferreira Novo

RESUMEN

Se analiza la interconexión temática en los títulos de trabajos académicos en el campo de las Ciencias de la Información publicados en ResearchGate. La metodología utilizada fue el enfoque de análisis de redes sociales, basado en un sistema de representación del conocimiento-red semántica, para mejorar las formas de visualización y representación de la información científica en estos entornos, a partir de la Teoría de la Clasificación Facetada (S. R. Ranganathan) y la Teoría de Concepto (Ingtraut Dahlberg) y asociado con las bases epistemológicas de Gilles Deleuze y Félix Guattari. Analizamos 324 perfiles de docentes de los 18 programas de posgrado en el campo de las ciencias de la información en Brasil para extraer los títulos de sus investigaciones. Como resultado, la red semántica creada tiene una caracterización topológica de red de pequeño mundo, lo que indica la convergencia en la interconexión entre los conceptos que conforman la red, que están representados y organizados en torno a un componente central. Se concluye que los experimentos de validación de la propuesta sugieren un continuo en la relación entre conceptos, destacando la

interdisciplinariedad y evolución de la investigación científica. Se crea un marco general para la aplicación del análisis de conceptos emergentes en Research Gate con el fin de planificar el camino teórico-metodológico presentado para la creación de estas redes.

Palabras clave: Representación de la Información; Redes Semánticas; Comunicación Científica

Thematic interconnections in Information Science at ResearchGate: the case of postgraduate professors in Brazil

Bruna Lessa and Hildenise Ferreira Novo

ABSTRACT

It analyzes the thematic interconnectedness in the titles of academic papers in the field of Information Science published in ResearchGate. The methodology used was the social network analysis approach, based on a knowledge representation system - semantic network, to enhance the ways of visualization and representation of scientific information in these environments, based on the Faceted Classification Theory (S.R. Ranganathan) and in Concept Theory (Ingtrout Dahlberg) and associated with the epistemological basis of Gilles Deleuze and Félix Guattari. We analyzed 324 profiles of teachers from the 18 Graduate Programs in the field of Information Science in Brazil to extract titles from their research. As a result, the semantic network created has topological characterization of small world network, indicating convergence in the interconnection between the concepts that make up the network, which are represented and organized around a central component. It is concluded that the validation experiments of the proposal suggest a continuum in the relationship between concepts, highlighting the interdisciplinarity and evolution of scientific research. A general framework for the application of the analysis of emerging concepts in ResearchGate is created in order to plan the theoretical-methodological path presented for the creation of these networks.

Keywords: Representation of Information; Semantic Networks; Scientific Communication

INTRODUÇÃO

Sistemas de Organização do Conhecimento (SOC) são desenvolvidos para estabelecer padrões que visam a adequação terminológica e a sistematização de conceitos para o estabelecimento relações semânticas, evitando ambiguidade tanto no âmbito da recuperação da informação, por usuários, quanto na indexação, ao orientar a escolha dos termos utilizados neste processo.

Desta maneira, seja para armazenamento e recuperação da informação em Unidades de Informação (biblioteca, arquivo, museu), criar terminologias para representação do conhecimento, ou ainda, descrever informações disponíveis na web e suas relações semânticas, os SOC objetivam, sobretudo, a organização da informação e do conhecimento, representando-os, de modo a facilitar a recuperação da informação.

A *web* semântica, ou *web* 3.0, vislumbra a geração de novos serviços para recuperação da informação. Assim, formas de busca implicaram em sua criação, pois o contexto informacional contemporâneo requer o estabelecimento das relações semânticas entre conceitos presentes em cada recurso informativo da *web*. Desse modo, as Tecnologias de Informação e Comunicação (TIC) alinham-se para a inovação na organização e representação dos conteúdos disponíveis na internet.

O meio acadêmico utiliza de recursos da *web* para a circulação e recuperação de informações científicas. Diversas possibilidades, bases de dados, repositórios institucionais, bibliotecas digitais, e mais recentemente, *sites* de redes sociais acadêmicos (SRSA), vêm ganhando espaço como fonte de informação informal, uma nova instância de comunicação científica entre a comunidade acadêmica. Potencializando a troca e compartilhamento de informações científicas de maneira rápida e dinâmica, favorecendo a agilidade na transmissão da informação.

Nesta perspectiva, tem-se um novo condicionante a incorporar este contexto que é a evolução e/ou obsolescência de conceitos característicos de um domínio do conhecimento e sua organização e representação no ambiente virtual. A multirreferencialidade¹, existente nas abordagens teóricas e práticas na elaboração da pesquisa científica, requer representações que possibilitem perceber a interconexão entre os conceitos que convergem das áreas do saber, o que de certo modo, desconfigura a ideia estrutural de um conhecimento isolado, hierarquizado, dotado do “poder” que é oriundo de conceitos estritamente genuínos de um domínio específico.

1 A concepção de multirreferencialidade presente neste trabalho parte da ideia de pluralidade nas abordagens interdisciplinares sobre objetos de estudos fronteiriços em diferentes domínios do conhecimento, apresentando diferentes perspectivas, seguindo o conceito delineado em Fróes Burnham e Fagundes (2002).

Contudo, salienta-se que os problemas relacionados à organização da informação na *web*, que interferem na recuperação da informação, e que por sua vez, configura-se como um dos motivos que inibem o uso de *sites* de redes sociais para pesquisa científica, é pouco explorado na literatura da Ciência da Informação (CI). Nota-se, que no campo de estudos voltados para esta problemática, que envolve a representação e organização da informação na *web*, tem-se utilizado sistemas de organização do conhecimento (SOC), tais como taxonomias, ontologias, mapas conceituais, redes semânticas, *folksonomias* mas, no entanto, não se verificou até o momento de elaboração desta pesquisa, estudos que abordem esta temática em *sites* de redes sociais no âmbito de informações científicas.

Diante da transformação dos modos de como se faz, busca, organiza e dissemina a informação, em especial, a informação científica, este trabalho, apresenta a estrutura geral de aplicação da análise de conceitos emergentes no *site* de rede social acadêmico *ResearchGate*, representando a modelagem realizada para o processo de coleta, extração, análise e representação da informação científica em Ciência da Informação no Brasil, utilizando-se da abordagem de análise de redes sociais para criação de uma rede semântica a fim evidenciar as relações entre conceitos e interconexões temáticas com outros domínios do conhecimento.

A COMUNICAÇÃO CIENTÍFICA E O PROCESSO DE ORGANIZAÇÃO, DISSEMINAÇÃO E USO DO CONHECIMENTO CIENTÍFICO

O termo “comunicação científica” foi cunhado pela primeira vez em 1939 por John Desmond Bernal, em seu livro *The social function of science*. Em 1976, Wilfred D. Garvey e S. D. Gottfredson, com ideias similares as de Bernal, propuseram um sistema integrado de periódicos para submissão de *preprints*. Na perspectiva de Garvey e Gottfredson (1976), cada área do conhecimento teria um único artigo que embasasse suas teorias e práticas, e vários *preprints* para complementar a publicação tradicional nos periódicos acadêmicos, permitindo a comunicação científica de forma mais rápida. Contudo, tais propostas não tiveram êxito devido à ausência de tecnologia disponível na época.

Para uma nova estruturação da comunicação científica, na década de 1990, Steven Harnad trouxe uma polêmica solução: *skywriter* – os pesquisadores difundiam suas publicações científicas na *web* com acesso aberto. A partir disso, novas instâncias de comunicação científica poderiam ser criadas em redes *on-line*, promovendo o livre acesso à informação científica, dinamizando

o fluxo da informação e a produção científica. Com o caminho traçado para a reconstrução de um modelo para o sistema de comunicação científica, também se faz necessário pensar sobre o comportamento dos usuários e produtores da informação frente aos canais formais e informais de informação.

Neste contexto de inovações, surgem os *sites* de redes sociais acadêmicas, os quais se inserem na terceira evolução das redes sociais da internet – Redes 3.0 –, caracterizadas pela “[...] integração com múltiplas redes, plataformas e funcionalidades através do uso de aplicativos e de mídias móveis.” (Santalla e Lemos, 2010: 59). Além dos *sites* de redes sociais, tais como *Facebook*, *LinkedIn* e *Twitter*, utilizados cada vez mais como recurso para disseminar a comunicação em pesquisa científica e, conseqüentemente, uma fonte informal de informação científica, sites especializados para a comunidade científica compartilhar suas pesquisas e desenvolver projetos compartilhados, por exemplo, *ResearchGate*, *Academia.edu* e *Mendeley*, já se configuram como fontes de pesquisa científica e espaço de interação e diálogo entre cientistas para colaboração aberta em pesquisa, mecanismo para classificação de pesquisadores e suas universidades de origem, construção de redes/grupos profissionais e de pesquisa, fornecimento, obtenção e divulgação de resultados de pesquisa, gerenciamento de documentos e citações, entre outros.

Evidencia-se, portanto, a transição dos modos e instrumentos para organização e recuperação da informação em espaços sociais semânticos, o que sustenta a elaboração desta pesquisa, uma vez que o trabalho coletivo em rede potencializa a adoção e classificação de vocabulários no contexto digital. O que acontece quando surge um novo conceito dentro de um domínio do conhecimento? Ao tratar sobre o futuro da investigação em redes sociais, Lévy (2010) adverte sobre o processo evolutivo da internet, voltado para o compartilhamento semântico, recomendando algumas etapas a considerar para a gestão do conhecimento nesse espaço virtual, dentre elas a categorização dos conteúdos em *tags*, classes ou ontologias, a síntese do conteúdo organizado, o compartilhamento informacional em mídias sociais.

Pensar sobre representação da informação e do conhecimento em ambientes colaborativos *on-line*, tais como os *sites* de redes sociais, torna-se imprescindível no momento em que avanços tecnológicos, no âmbito do armazenamento de dados têm sido experimentados, o que demanda mais uma etapa a ser compreendida e analisada pelos cientistas da informação no que tange à materialização ou, ainda, desmaterialização da informação e do conhecimento representados nas conexões estabelecidas em rede.

A REPRESENTAÇÃO DA INFORMAÇÃO CIENTÍFICA NO SÉCULO XXI: O PRINCÍPIO DA CONEXÃO E AS RELAÇÕES FUNCIONAIS ENTRE CONCEITOS

A necessidade de criar sistemas de classificação que permitissem combinar várias relações para representar exatamente o assunto dos documentos, levou Ranganathan à constatação de que as classificações bibliográficas não poderiam seguir somente os predicáveis representadas na árvore de Porfírio, tais como as que possuem estruturas enumerativas – Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU), por exemplo, – elas deveriam originar-se de subdivisões em categorias, mais próximas das teorias de Aristóteles, sugerindo sua representação pela Árvore Baniana (*Banyan Tree*), quando os conceitos se relacionam a outros diversos conceitos, semelhantemente ao processo de ramificação.

Sob esta lógica, Ranganathan (1967: 372) apresentou a *Espiral do Desenvolvimento de Assuntos* para associar a atividade de organização do conhecimento ao crescimento e expansão da produção do conhecimento e, consequentemente, o desenvolvimento de novos assuntos, permitindo assim a dinâmica das constantes modificações no universo do conhecimento.

Esta analogia com a Árvore Baniana feita por Ranganathan, coaduna-se com o conceito de rizoma de Gilles Deleuze e Félix Guattari. Em *Mil platôs: capitalismo e esquizofrenia*, Deleuze e Guattari (2011) constroem (a partir de um pensamento orientado pela Filosofia Continental, largamente praticada na Alemanha e França) a noção da palavra de ordem como um conceito filosófico sobre a linguagem, explicando-a por meio do conceito de rizoma. Para os autores, não há início e nem fim para os rizomas, eles se conectam, estão em constante interconexão seguindo uma linha de evolução. Nessa linha evolutiva, os rizomas são estratificados, territorializados, organizados, cartografados, significados, alterando também sua natureza ao se conectar a outros rizomas, assumindo sua multiplicidade, sua heterogeneidade, desenvolvendo, assim, princípios que apontam para as características de um rizomorfo, dentre eles o princípio da conexão, no qual um rizoma pode ser conectado a qualquer outro.

Tais princípios norteiam o entendimento de que diferentemente da ideia de “árvore”, que possui uma estrutura hierárquica, estrutural e filiativa, o rizoma em Deleuze e Guattari representa a região de platôs (zonas de intensidades contínuas), é a aliança, é cruzamento, o “inter-ser”, o “entre”. Nesse contexto, definem o conceito de agenciamento como “[...] precisamente este crescimento das dimensões numa multiplicidade que muda necessariamente de natureza à medida que ela aumenta suas conexões.” (Deleuze e Guattari, 2011:17).

Conforme Deleuze e Guattari (2011: 67; vol. 2) a expressão mais real de um rizoma é, portanto, a multiplicidade, pois está sempre em movimento, variando sua distância em relação aos outros, em virtude das interpretações que são realizadas pelos sujeitos, ou ainda “interpretância”. Há então um ciclo, uma espiral que representa o regime de signos, conteúdos e expressões específicas emitidas por meio dos agenciamentos, e assegura sua expansão “[...] para que novos círculos brotem ou para que os antigos sejam realimentados.” (*Figura 1*).

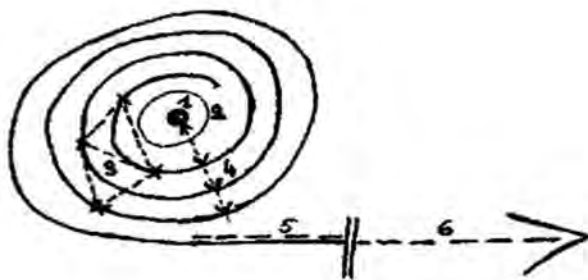


Figura 1. Espiral das transformações analógicas nos regimes de signo (Deleuze e Guattari, 2011: 59)

Esta relação entre a concepção de Ranganathan sobre os conceitos e de Deleuze e Guattari acerca dos rizomas, potencializa a observação sobre as possíveis interconexões entre os conceitos de um determinado domínio do conhecimento a partir de suas multiplicidades, justificando-se, portanto, a evolução da organização do conhecimento, sobretudo das teorias da classificação, de uma estrutura arbórea, com início, meio e fim, para uma concepção rizomática, que parte do meio para todos os possíveis inícios e fins, em constante dispersão.

Nesse contexto, adaptamos então o que Deleuze e Guattari (2011) chamam de “plano de consistência” à tentativa de organizar os conceitos segundo suas relações/conexões em suas próprias multiplicidades, observando também a Teoria do Conceito de Dahlberg (1978), quando é possível fazer a leitura de que se um determinado conceito possui a mesma característica de um outro, é estabelecida relações entre conceitos. Esta relação conceitual é chamada por Dahlberg de relação funcional, quando as relações entre os conceitos são verificadas com base na valência semântica dos verbos, que considera questões complementares, tais como contexto, propriedades, formas, etc.

O método desenvolvido por Dahlberg para a formação de conceitos, baseado na tríade conceitual, pode ser compreendida também a partir do

regime de signos de Deleuze e Guattari (2011), evidenciando a relevância da tradição semiótica pragmática e sua compreensão de conceitos como signos, especialmente quando Dahlberg inclui nas relações entre conceitos, as relações funcionais.

O encontro convergente entre tais autores para o desenvolvimento desta pesquisa está, portanto, nas relações conceituais que ocorrem em um mesmo domínio do conhecimento, e das relações existentes entre este domínio e tantos outros a partir de suas características, constituindo-se assim, categorias, que são o próprio conceito em sua mais ampla extensão, e daí sua forma verbal.

Com base nas teorias e bases epistemológicas aqui apresentadas, é que se pensou a estruturação das relações conceituais para o *ResearchGate*, site de rede social acadêmico voltado para o conhecimento de pesquisa científica. Criado em 2008, tem como objetivo o compartilhamento de artigos científicos e a socialização entre pesquisadores e cientistas. O *ResearchGate* permite fazer *upload* e *download* de trabalhos científicos, e possui interface acessível e atrativa, capaz de fornecer um resumo do perfil profissional e de pesquisa, o que enriquece a interação entre investigadores de todo o mundo.

Esse *site* tem como objetivo promover a conexão entre cientistas com a finalidade de potencializar a colaboração e o compartilhamento de pesquisas científicas, acelerando o processo de aquisição do conhecimento. Um dos serviços disponibilizados, permite que qualquer utilizador coloque questões de pesquisa, e qualquer outro possa respondê-las. A indexação é realizada a partir dos nomes dos autores, vinculando-os aos seus artigos. Possui ainda, uma área para anúncios de emprego científico, produzindo alertas de sugestões baseadas no perfil do usuário, o qual só pode fazer parte da rede se tiver vínculo com alguma instituição de pesquisa, sendo obrigatória a verificação dos dados.

CAMINHOS METODOLÓGICOS

O caminho metodológico traçado nesta pesquisa se adequa ao campo da pesquisa exploratória no âmbito do ambiente de redes sociais acadêmicas. O planejamento teórico se baseia no fazer científico da organização e representação da informação, domínio da Ciência da Informação. Desse modo, a pesquisa estrutura o método de análise em duas dimensões: teórica e aplicada.

A dimensão aplicada, de acordo com Gil (2008), tem por objetivo o conhecimento produzido em aplicações práticas, a exemplo de redes semânticas, que ao serem exploradas, revelam sentidos das pesquisas por meio dos conceitos que emergem nessas redes. Quanto aos objetivos a pesquisa se

configura como exploratória e descritiva, a fim de explicitar ideias e descrever determinada população ou fenômeno.

A abordagem foi hipotético-dedutiva, onde se partiu da percepção de uma lacuna em pesquisas sobre a representação da informação científica em ambientes *on-line*, no que concerne analisar o espaço de conhecimento traduzido através de conceitos extraídos dos títulos dos estudos publicados em novas instâncias de comunicação científica, tais como as redes sociais acadêmicas *on-line*.

Delineamento do corpus da pesquisa

Para criação da rede semântica de títulos de trabalhos acadêmicos do domínio da CI no *ResearchGate*, foram extraídos dados dos perfis de 324 docentes dos 18 Programas de Pós-Graduação (PPG) no domínio da CI com perfis ativos (entende-se como perfil ativo o docente que publicou mais de um trabalho acadêmico no *site*), sendo necessário a criação de um algoritmo que permitisse a coleta, com o auxílio da ferramenta *Selenium Web Driver*, utilizando a linguagem de programação *Java*, além das ferramentas *ChromeDriver*, responsável por toda navegação que será feita no *site*, e *WebDriverWait*, responsável por processar o tempo necessário, definido no código, para que os elementos da página fossem carregados.

Após esta etapa, os dados extraídos foram convertidos para uma planilha de arquivo *.xlsx* e, onde foi armazenado todas as informações, já definidas anteriormente, de cada perfil identificado. A análise passou também por um tratamento manual, quando foi realizada a mineração de texto, em quatro etapas, nos títulos dos trabalhos acadêmicos localizados, adotando-se o método de eliminar palavras sem significado intrínseco e transformar as palavras restantes para sua forma canônica (Pereira *et al.*, 2011) (*Figura 2*).

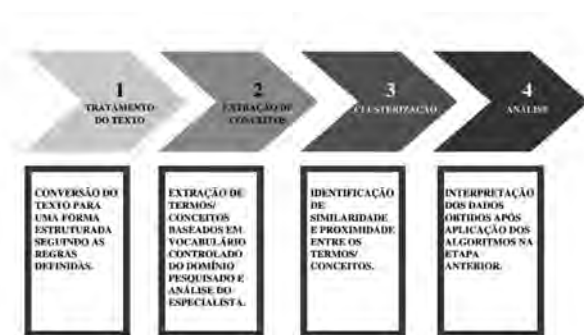


Figura 2. Etapas da mineração de texto aplicada nos títulos dos trabalhos acadêmicos extraídos do *ResearchGate*

Na etapa de extração de conceitos, adotou-se como vocabulário controlado o Tesouro Brasileiro de Ciência da Informação (TBCI)², bem como a análise feita pela pesquisadora baseada na literatura científica da CI e tendências em pesquisa neste domínio. Por exemplo, para o título “A adoção das árvores hiperbólicas como instrumentos de representação de conhecimento nos sistemas de informação multimodais”, após remoção dos elementos definidos nas regras, foram selecionados os conceitos que representam o domínio em análise, com valor semântico, combinando-os em um único componente: “*arvorehiperbolica*” “*instrumento*” “*representacaodeconhecimento*” “*sistemasdeinformacaomultimodais*”.

Para as etapas de clusterização e análise, as palavras foram submetidas ao tratamento computacional para classificar, modificar e eliminar palavras, caso necessário, evitando assim, ambiguidades e distinguindo palavras lexicais de palavras gramaticais, bem como a construção automática de taxonomia entre os conceitos, a partir das relações e similaridades semânticas definidas por meio de índices da Teoria de Redes, através do Netpal.exe³, programa para criar redes de palavras, que usa o Pacote UNITEX⁴ e o programa Ambisin.exe⁵, e para a visualização da rede, utilizou-se os *softwares Gephi* e *Pajek*.

Compõe o *corpus* da pesquisa 500 títulos de trabalhos acadêmicos publicados pelos docentes dos PPG no domínio da CI no *ResearchGate*, entre os anos 2013 e 2019, permitindo a representação e visualização da informação científica em *sites* de redes sociais a partir de rede semântica, proposta metodológica deste estudo.

Utiliza-se a abordagem de análise de redes sociais (ARS) para criação da rede semântica de títulos. Após exportação do corpus, utilizou-se o *software Gephi* para visualização da rede e interpretação de suas propriedades (*Tabela 1*).

- 2 Publicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), em 2014. Disponível em: http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-da-informacao-1/copy_of_TESAURCOMPLETOFINALCOMCAPA24102014.pdf
- 3 Programa que constrói redes de cliques a partir da estrutura criada pelo Pacote UNITEX. Desenvolvido pelo Prof. José Garcia Vivas Miranda e seu grupo de pesquisa (Caldeira, 2005).
- 4 Conjunto de programas baseados em léxicos para o processamento de corpos textuais. Disponível em: <http://unitexgramlab.org/pt>
- 5 Programa que determina o tipo de análise que se pretende realizar com as redes de palavras desenvolvido por Silvia Maria Gomes Caldeira (Caldeira, 2005).

Índices	Definição	Valores
Grau médio ($\langle k \rangle$)	Representa o valor médio em que cada conceito/termo se conecta a outro	6,19
Densidade (Δ)	Define a quantidade entre um conceito/termo e outro (maior menor caminho), expressando o quão perto o grafo está de se tornar completo.	0,006
Diâmetro (D)	Representa o caminho mais longo de todos os caminhos mais curtos calculados entre dois conceitos/termos.	11
Caminho mínimo médio (L)	Distância média do caminho entre os conceitos/termos.	3,789
Coeficiente de aglomeração (C)	Probabilidade de conexão entre conceitos/termos quando agrupados próximos de determinado conceito/termo.	0,863
Modularidade	Identifica presença de comunidades entre os conceitos/termos.	0,71

Tabela 1. Índices das propriedades da rede semântica dos títulos dos trabalhos acadêmicos extraídos do *ResearchGate* (2013-2019)

Assim, foram obtidos 1.287 conceitos, sendo a rede composta por 1.287 nós e 3.512 arestas.

Rede Semântica de títulos de trabalhos acadêmicos no domínio da Ciência da Informação publicados no ResearchGate

Nas redes semânticas os conceitos são modelados como nós em uma rede de relacionamentos variáveis. Na representação da informação e do conhecimento, as redes semânticas são bidimensionais, ou seja, a relação entre os objetos e/ou conceitos analisados podem ter mais de um tipo de relacionamento, estrutura básica para OC em redes. São mais ricas que os tesauros na definição de categorias ou relações semânticas e, “[...] representa o conhecimento na forma de nós (conceitos, atributos) ligados por arcos (relacionamentos).” (Sales, Campos e Gomes, 2008: 64). Elas surgem como uma adaptação ao tesouro conceitual, objetivando uma melhor visualização da relação entre os conceitos, possibilitando navegar através de grafos de conexão e, assim, representar o conhecimento.

Após análise dos índices das propriedades da rede semântica dos títulos dos trabalhos acadêmicos extraídos do *ResearchGate* (2013-2019) (Tabela 1), foi possível a identificação do (L) e (C) e comparação com rede aleatória (RA) do componente gigante da rede real, com os mesmos valores de n (número de vértices) e k (grau), da rede semântica de títulos de trabalhos acadêmicos, foi possível verificar comportamento característico de rede de mundo pequeno ou *small world*, pois há um valor alto de C ($0,92 \leq C \leq 0,97$) e, o valor de L

semelhante a rede semântica de títulos de trabalhos acadêmicos correspondente ($1,65 \leq L \leq 2,68$), além das propriedades específicas de redes de mundo pequeno, tais como: não direcionada, não ponderada, simples, esparsa e conectada, delineadas por Watts e Strogatz (1998).

A caracterização topológica desta rede semântica indica uma convergência na interconexão entre os conceitos que a compõem, os quais estão representados e organizados em torno de um componente central - “*cienciadainformacao*” – com 168 nós conectados, ou seja, o nó com maior grau da rede. A partir desse resultado, fez-se uma análise do nó “*cienciadainformacao*”, aplicando o algoritmo Rede Ego do *software Gephi*, que indicou grau médio de aproximadamente dois nós (1,963). Isso significa que o caminho mínimo médio entre cada nó dessa rede é de dois conceitos, o que resulta na compreensão de outro índice, o diâmetro (D), o maior – mais longo – de todos os caminhos mais curtos calculados entre os conceitos (2), inferindo o quanto cada conceito está afastado na rede.

Esta análise demonstra que o estudo sobre Ciência da Informação é o que mais emerge nas publicações analisadas, indicando a constante busca pela consolidação do domínio enquanto Ciência, bem como seu caráter interdisciplinar ao relacionar-se com outros temas que não são necessariamente do seu campo teórico e epistemológico, a exemplo do conceito “*educacao*” (Figura 3), fato que sugere o porquê da necessidade do uso do conceito “*cienciadainformacao*” na maioria dos títulos de trabalhos acadêmicos analisados.



Figura 3. Sub-rede do conceito “*educacao*”

O conceito “*educacao*” possui conexão mais próxima ao conceito central da rede “*cienciadainformacao*” em relação aos demais, assumindo duas funções

nesta sub-rede: “[...] gerar o significado básico da representação” e “[...] determinar a organização global de todos os elementos” (Sá, 1996: 22). Ao buscar o conceito “*educacao*” no TBCI, verificou-se que os outros conceitos relacionados ao conceito “*educacao*” na Figura 4 não fazem parte das relações hierárquicas – termo específico (TR) e das relações associativas – termo associado (TR) do TBCI (*direitos humanos, competência em informação, ficção juvenil brasileira, arte, letramento informacional, mercado*).

Tais exemplos indicam um *continuum* na relação entre conceitos no domínio da CI, quando a interdisciplinaridade e evolução das pesquisas científicas tem se encarregado de assegurar aos conceitos criados por uma comunidade científica, a expansão, inclusão de outras características e criação de novas formas verbais, onde os conteúdos dissolvem neles mesmos suas características próprias, tendo-se neste caso, sobretudo, uma relação funcional entre os conceitos, expressada nos títulos de trabalhos acadêmicos, os quais traduzem os discursos produzidos pelos pesquisadores ao sintetizar seus estudos.

Dessa maneira, apresenta-se a estrutura geral de aplicação da análise de conceitos emergentes no *ResearchGate* (Figura 4), o qual possibilitou representar as interconexões temáticas em Ciência da Informação, partindo-se do entendimento de que a construção do conhecimento é viabilizada pelas interações entre elementos numa rede. Assim, no âmbito dos estudos sobre a representação da informação, poderá ser aplicado como um modelo para análises da mesma natureza em outros *sites* de redes sociais acadêmicos e, consequentemente, identificar conceitos emergentes em um domínio do conhecimento e suas relações interdisciplinares.

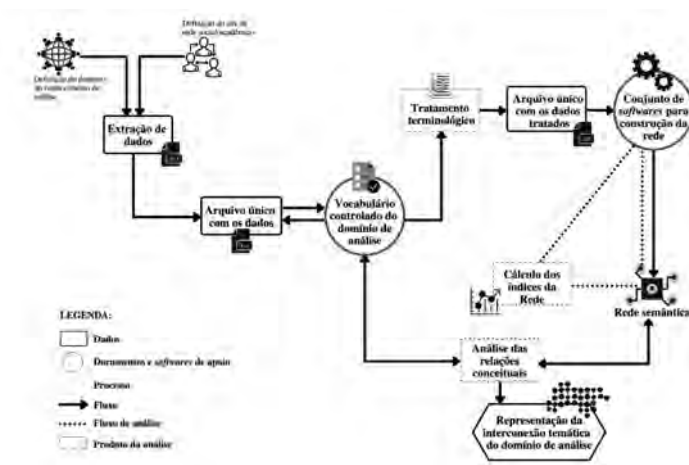


Figura 4. Estrutura geral de aplicação da análise de conceitos emergentes no *ResearchGate*

Este modelo tem como objetivo tornar visível os conceitos emergentes de pesquisas interdisciplinares e multirreferenciais que se relacionam com a CI, sobretudo, no ambiente dos SRSA, potencializando a representação da informação científica nestas novas instâncias de comunicação da ciência.

CONSIDERAÇÕES FINAIS

As redes semânticas foram utilizadas como um sistema para representação da interconexão temática no *ResearchGate* para o domínio da Ciência da Informação devido a possibilidade de uma organização em macroestrutura semântica e visualização das relações entre conceitos, considerando o contexto em que se expressam e de como são arranjados e sintetizados em um título de trabalho científico, por exemplo.

Salienta-se que, não se deve confundir a “transparência” das redes sociais acadêmicas com a falta de semântica nos produtos que ali são compartilhados. A narrativa presente nos assuntos em cada título de trabalho acadêmico carrega significados diversos que dependem do contexto do leitor, do pesquisador, ou ainda, do uso dos conceitos.

Como a representação se dá em tempo e espaço, convém ressaltar que o universo do conhecimento é um *continuum*, desse modo, novas formas de representação da informação e conhecimento se manifestarão acompanhando o desenvolvimento das TIC. Portanto, novos conceitos surgem para acompanhar a atualização e evolução das pesquisas desenvolvidas nos diferentes domínios de conhecimento.

Nesse sentido, utilizou-se neste estudo a fundamentação teórica baseada na TCF e TC, bem como as contribuições filosóficas trazidas por Deleuze e Guattari (2011) sobre o que é conceito e a noção de rizoma, associadas à abordagem de ARS, considerando o uso do TBCI para perceber a eminência de novos conceitos e suas relações com outros já consolidados. Teve como objetivo contribuir para a representação da informação em ambientes complexos e dinâmicos, tais como os SRSA, considerados aqui como novas instâncias de comunicação científica.

Desse modo, acredita-se que, no âmbito da análise conceitual de um domínio do conhecimento nesses ambientes, deve-se partir de uma concepção rizomática, para a qual não há início e nem fim para a relação entre conceitos. Acredita-se, portanto, que esta pesquisa contribui, como um modelo que poderá ser aplicado em outros *sites* de redes sociais acadêmicas para a representação da informação científica de um domínio do conhecimento, além de identificar conceitos emergentes e em evolução.

REFERÊNCIAS

- Bernal, J. D. 1939. *The social function of science*. London: George Routledge & Sons.
- Caldeira, S. M. G. 2005. “Caracterização da Rede de Signos Lingüísticos: um modelo baseado no aparelho psíquico de Freud”. Dissertação de mestrado interdisciplinar em modelagem computacional, Fundação Visconde De Cairu- Salvador-Bahia-Brasil.
- Dahlberg, I. 1978. “Teoria Do Conceito”. *Ciência Da Informação* 2 (7): 101-107.
<http://revista.ibict.br/ciinf/article/view/115>
- Deleuze, G. e F. Guattari. 2011. *Mil platôs: capitalismo e esquizofrenia*. 2a. ed. São Paulo: Editora 24.
- Frões Burnham, T. e N. C. Fagundes. 2002. “Transdisciplinaridade, Multirreferencialidade e Currículo”. *Revista da FAGED* (Salvador, Bahia) (5): 39-55.
<https://periodicos.ufba.br/index.php/entreideias/article/view/2837/2013>
- Garvey, W. D. e S. D. Gottfredson. 1976. “Changing the system: Innovations in the interactive social system of scientific communication”. *Information Processing and Management* (12):165-176.
[https://doi.org/10.1016/0306-4573\(76\)90003-0](https://doi.org/10.1016/0306-4573(76)90003-0)
- Gil, A. C. 2008. *Métodos e técnicas de Pesquisa social*. São Paulo: Ed. Atlas.
- Harnad, S. 1990. “Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry”. *Psychological Science* (1): 342-343.
<http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>
- Lévy, P. 2010. “O futuro da investigação em redes sociais”. Comunicação apresentada em Conferência internacional sobre redes sociais, Curitiba, Brasil, março de 2010.
<http://escoladeredes.net/video/pierre-levy-o-futuro-da-1>
- Pereira, H. B. B., I. S. Fadigas, V. Senna e M. A. Moret. 2011. “Semantic networks based on titles of scientific papers”. *Physica A: Statistical Mechanics And Its Applications* 6 (390): 1192-1197.
<https://doi.org/10.1016/j.physa.2010.12.001>
- Ranganathan, S. R. 1967. *Prolegomena to Library Classification*. Bombay: Asia Publ. House.
- Sá, C. P. de. 1996. “Representações Sociais: teoria e pesquisa do Núcleo Central”. *Temas em Psicologia* 4 (3):19-33.
http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X1996000300002&lng=pt&tlng=pt
- Sales, L. F., M. L. A. Campos e H. E. Gomes. 2008. “Ontologias De Domínio: um estudo das relações conceituais”. *Perspectivas Em Ciência Da Informação* 2 (13): 62-76.
<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/219>
- Santaella, L. e R. Lemos. 2010. *Redes Sociais Digitais: a cognição conectiva do Twitter*. São Paulo: Paulus.
- Watts, D. J. e S. H. Strogatz. 1998. “Collective dynamics of small world networks”. *Nature* 4 (393): 440-441.
<https://www.nature.com/articles/30918>
- Wüster, E. 1998. *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Barcelona: IULA, Universitat Pompeu Fabra.

Para citar este texto:

Lessa, Bruna, e Hildenise Ferreira Novo. 2021. “Interconexões temáticas em Ciência da Informação no *ResearchGate*: o caso de professores da Pós-graduação no Brasil”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 151-167.

<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58449>

DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58449>

Evaluación UX para sitios web orientados a bibliotecas digitales con usuarios ciegos

Teresita Álvarez-Robles*

Yadira Alejandra Orozco Osuna**

Francisco Javier Álvarez Rodríguez***

Artículo recibido:

23 de abril de 2021

Artículo aceptado:

16 de agosto de 2021

Artículo de investigación

RESUMEN

El objetivo de este trabajo es evaluar y mejorar la experiencia del usuario con discapacidad visual al hacer uso de sitios web del área de las bibliotecas digitales. Para lograr el objetivo se realizaron pruebas con métodos de evaluación de la experiencia del usuario. Las pruebas se llevaron a cabo en los sitios web de dos bibliotecas públicas de la ciudad de Guadalajara, en México, las cuales a simple vista no toman en cuenta las pautas de accesibilidad y usabilidad suficientes para que sean utilizados por personas ciegas. Los instrumentos que se

* Departamento Académico de Sistemas Computacionales, Universidad Autónoma de Baja California Sur, México
alvarez88.t@gmail.com

** Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, México
yadira.orozco4156@alumnos.udg.mx

*** Departamento de Ciencias Básicas, Universidad Autónoma de Aguascalientes, México
fjalvar@correo.uaa.mx

utilizaron para las pruebas fueron entrevistas, evaluación de heurísticas, y el método “pensando en voz alta” modificado para usuarios ciegos (Álvarez-Robles, 2019). Con base en los resultados obtenidos en dichas pruebas, es posible notar que los usuarios presentaron dificultades para hacer uso del catálogo web. Se identificaron los principales problemas y se realizó una propuesta de guía de diseño para el desarrollo de sitios web enfocados en las bibliotecas digitales. Como conclusión se puede observar la importancia de realizar pruebas e incluir a los usuarios ciegos desde fases tempranas del desarrollo y con esto poder evitar problemas o errores futuros en el sitio web.

Palabras clave: Experiencia del Usuario; Accesibilidad Web; Personas Ciegas; Usabilidad

UX Evaluation on digital libraries-oriented websites with blind people

Teresita Álvarez-Robles, Yadira Alejandra Orozco Osuna and Francisco Javier Álvarez Rodríguez

ABSTRACT

The aim of this work is to evaluate and improve the capacity of the blind user when making use of websites focused on the digital libraries area. To achieve this objective, tests with user experience evaluation methods were carried out. The tests were done on two public libraries websites in the city of Guadalajara, Mexico, which at first glance don't consider sufficient accessibility and usability guidelines to be used by blind people. The instruments used for the tests were: interviews, heuristic evaluation and the thinking aloud method modified for blind users (Álvarez-Robles, 2019).

Based on the results obtained in the said tests, it was possible to note that users had difficulties in using the Web catalog. The main problems were identified and a design guide proposal was made for the development of Web sites focused on digital libraries. As a conclusion, the testing importance and the inclusion of blind people became clear from the early stages of the evaluation and thus for being able to avoid future problems or errors on the website.

Keywords: User Experience; Websites Accessibility; Blind People; Usability

INTRODUCCIÓN

El uso de las Tecnologías de la Información y la Comunicación (TIC) resulta algo prácticamente indispensable para la población hoy en día e inevitablemente no se encuentra al alcance de todos por diferentes factores, es por esto que la accesibilidad juega un papel importante en las nuevas formas de adaptarse e interactuar con la tecnología y marca seriamente la diferencia para las personas ciegas.

Existen una serie de pautas de accesibilidad desarrolladas por parte del *World Wide Web Consortium* (W3C) con el objetivo de proporcionar un único estándar para la accesibilidad del contenido en la web (W3C, 2005).

Por otra parte, también se debe tener en cuenta el concepto de usabilidad, que se refiere al alcance en el que un producto puede ser utilizado por usuarios específicos para alcanzar metas específicas con efectividad, eficiencia y satisfacción en un contexto específico de uso (Martínez de la Teja, 2006).

Tomando en cuenta los conceptos mencionados, actualmente muchas bibliotecas cuentan con un catálogo web de libros disponibles con el objetivo de evitar acudir físicamente a la biblioteca, esto permite que las personas ciegas puedan hacer una consulta sin necesidad de trasladarse; sin embargo, se sabe que en su mayoría las bibliotecas no cuentan con un sitio y aquellas que sí no realizan las pruebas necesarias con los usuarios objetivo (*El Informador*, 2010a).

En la Red Nacional de Bibliotecas Públicas (conformada por más de 7 000 bibliotecas) se contabilizan 100 bibliotecas en las que no sólo se ofrece el servicio de sala braille, sino que además cuentan con lecturas en voz alta, grabación de texto en audio, impresión de textos en braille y amplificadores de imagen, así como orientación y apoyo para usuarios con discapacidad visual (*El Informador*, 2010b).

En México, cerca de 6 % de la población padece algún tipo de discapacidad. De este porcentaje, se estima que poco más de 58 % presenta algún tipo de discapacidad visual (Carrillo, 2018). En el caso específico de Jalisco, existen 23 bibliotecas que cuentan con recursos para personas con discapacidad visual, por lo tanto, es el estado que se encuentra en el primer lugar al hacer uso de este tipo de infraestructura (*El Informador*, 2010b).

Por esta razón, todos los sistemas que operan las bibliotecas públicas, principalmente las que cuentan con salas braille, deberían cumplir con las pautas de accesibilidad y usabilidad ofreciendo así una experiencia de usuario óptima a los usuarios ciegos.

Evaluación de la usabilidad

La usabilidad se asocia a tres elementos: eficacia, eficiencia y satisfacción (Reyes Vera, Berdugo Torres y Machuca Villegas, 2016). Los métodos para evaluarla, a su vez, se clasifican en tres grupos, como se muestra en la *Tabla 1*.

Inspección	Indagación	Test
- Evaluación heurística. - Recorridos cognitivos.	- Observación de campo. - Grupos de discusión (<i>focus group</i>).	- Pensar en voz alta (<i>thinking aloud</i>).

Tabla 1. Ejemplos de métodos para evaluación de usabilidad
Fuente: Granollers, Lorés y Cañas (2011)

Experiencia del usuario

Para D'Hertefelt (2000), la experiencia del usuario o *User eXperience* (UX por sus siglas en inglés) representa un cambio emergente del propio concepto de usabilidad, donde el objetivo no se limita a mejorar el rendimiento del usuario en la interacción, sino que se intenta resolver el problema estratégico de la utilidad del producto y el problema psicológico del placer y diversión de uso. En otras palabras, podemos definir la UX como la sensación, sentimiento, respuesta emocional, valoración y satisfacción del usuario respecto a un producto (Álvarez-Robles, 2019).

Factores que componen la experiencia del usuario

Son pocos los modelos que descomponen la UX en las diferentes variables que la condicionan y modelan. El modelo de Peter Morville, por ejemplo, plantea que la UX está compuesta por seis factores: usabilidad, accesibilidad, utilidad, credibilidad, deseabilidad, y encontrabilidad, con lo cual el producto obtiene valor (Morville, 2004) (*Figura 1*).

Particularmente, cada factor se encarga de que el producto, sistema o servicio sea de calidad para el usuario final obteniendo con esto una buena experiencia de usuario (Morville, 2004).

Métodos de evaluación de la UX

Los métodos de evaluación de la UX buscan encontrar cómo el usuario percibe lo que está experimentando al relacionarse con un sitio web, se toman

en cuenta sus sentimientos, opiniones, dificultades, entre otros elementos; por esta razón existen diferentes formas de realizar dicha evaluación (Granolers, Lorés y Cañas, 2011).

Con base en <https://www.allaboutux.org/> son 86 los métodos de evaluación de la UX que existen hoy en día; para este proyecto se tomaron en cuenta las condiciones de los usuarios, donde los métodos se adaptan al usuario ciego (Álvarez-Robles, 2019).



Figura 1. Modelo de Peter Morville
Fuente: Morville (2004)

Desarrollo

Se hizo una investigación de las bibliotecas públicas en la ciudad de Guadalajara, Jalisco, México. De 12 bibliotecas encontradas, se seleccionaron aquellas que contaran con algún tipo de recurso para personas ciegas y manejaran un catálogo web. Con base en lo anterior, se tomó la decisión de trabajar con los catálogos web de la Biblioteca Central Profesor García Ruiz (BCPGR) y de la Biblioteca Iberoamericana Octavio Paz (BIOP).

Perfil del usuario

Para las pruebas se contactó a la Organización de Invidentes Unidos de Jalisco A. C., y se obtuvo la participación de cinco usuarios ciegos (Tabla 2).

Usuario	Género	Edad	Tipo de ceguera
1	Masculino	24	Ciego Total. Congénito
2	Femenino	26	Ciego Total. Congénito
3	Femenino	40	Ciego Total. Congénito
4	Masculino	56	Ciego Total. Adquirido
5	Masculino	60	Ciego Total. Congénito

Tabla 2. Descripción de los usuarios

Todos los usuarios que participaron en las pruebas están familiarizados con los lectores de pantalla, principalmente el sistema JAWS. Cabe mencionar que no habían hecho uso de un catálogo web previamente pues prefieren intercambiar libros por medio de USB.

Selección de casos de estudio

Caso 1: BCPGR. La biblioteca cuenta con un acervo aproximado de 50 000 volúmenes en circulación, que están a disponibilidad de los usuarios en la web a través del sistema *Open Public Access Catalog* (OPAC) (*El Informador*, 2018).

Este sitio cuenta con un apartado de Accesibilidad donde puede modificarse el formato de la página a sus versiones de fuentes grande, alto contraste o sólo texto (*Figuras 2 y 3*).



Figura 2. Página principal del sitio web de la Biblioteca Central
Fuente: Red Estatal de Bibliotecas Públicas de Jalisco (2015)



Figura 3. Versión sólo texto del sitio web de la Biblioteca Central
Fuente: Red Estatal de Bibliotecas Públicas de Jalisco (2015)

Caso 2: BIOP. El OPAC de la Universidad de Guadalajara (UdeG) contiene los registros bibliográficos de toda la Red de Bibliotecas de la Universidad de Guadalajara; sin embargo, en esta plataforma también se incluye la Biblioteca Iberoamericana y otras tres bibliotecas públicas (*Figuras 4 y 5*).



Figura 4. Página principal del catálogo general donde aparece la Biblioteca Iberoamericana
Fuente: Universidad de Guadalajara (2009)

Figura 5. Página principal del sitio web de la Biblioteca Iberoamericana
Fuente: Universidad de Guadalajara (2009)

Una evaluación heurística tiene como objetivo medir la calidad de la interfaz de cualquier sistema interactivo en relación con su facilidad para ser aprendido y usado por un determinado grupo de usuarios en un determinado contexto de uso (Arancibia Céspedes y González Canales, 2017). Se tomó como guía el listado de principios heurísticos de Jakob Nielsen, los cuales son generales y se pueden aplicar a cualquier sitio web (Nielsen, 1994):

- H1. Visibilidad del sistema
- H2. Coincidencia entre el sistema y el mundo real
- H3. Control y libertad del usuario
- H4. Consistencia y estándares
- H5. Prevención de errores
- H6. Minimizar la carga de memoria
- H7. Flexibilidad y eficacia del uso
- H8. Diseño estético y minimalista
- H9. Ayuda al usuario para reconocer, diagnosticar y recuperarse de errores
- H10. Ayuda y documentación

Para la evaluación de la UX durante las pruebas a los sitios web se usó el método “pensando en voz alta”, también conocido como “*Thinking aloud*”, modificado para usuarios ciegos (Álvarez-Robles, 2019). En general es un método que requiere que los participantes verbalicen lo que están haciendo y piensen mientras completan una tarea, revelando aspectos que puedan gustar, confundir o frustrar (Ericsson y Simon, 1980; Hughes, 2012).

Evaluación de Heurística- Aplicación Caso 1: BCPGR (Biblioteca 1). Para el caso de la Biblioteca 1 se identificaron los errores usando la versión de sólo texto disponible en el apartado de accesibilidad. Los errores se centran en la adecuación del lector de pantalla y la movilidad ofrecida por medio del teclado, además de problemas relacionados al uso del formulario (Tabla 3).

ID	Definición del problema	Heurística incumplida
P1	Falta notificación al marcar un libro	H1
P2	Difícil acceso desde el teclado	H4
P3	No hay instrucciones de llenado de formulario	H1
P4	Apartado no cuenta con el botón volver (resultados)	H5
...	...	...

Tabla 3. Problemas heurísticos identificados en el sitio web de la Biblioteca Central Profesor García Ruiz

Evaluación de Heurística- Aplicación Caso 2: BIOP (Biblioteca 2). La Biblioteca 2 no cuenta con un apartado de accesibilidad, por lo que sólo se evaluó su adecuación con el lector de pantalla y la posibilidad de desplazamiento por medio del teclado. Algunos de los problemas encontrados con base en la evaluación heurística se muestran en la Tabla 4.

ID	Definición del problema	Heurística incumplida
P1	No accesible desde el teclado al menú principal	H3
P2	Notificaciones de error que no se leen hasta que las apuntas	H1 H5
P3	No se sabe si la escritura está habilitada	H3 H5
P4	Formulario expandible que no se lee de forma automática	H5 H9
...	...	...

Tabla 4. Problemas heurísticos identificados en el sitio web de la Biblioteca Iberoamericana Octavio Paz

Con base en los problemas identificados se diseñaron una serie de tareas para evaluar si los errores son detectados por los usuarios ciegos, y de ser así, contemplar sus comentarios para la propuesta de la guía de diseño para sitios web accesibles para usuarios ciegos.

Tareas y resultados

Las tareas establecidas para el Caso 1 y el Caso 2 se encuentran en los *Anexos A y B* respectivamente, en ellos se enlistan las tareas establecidas para ser realizadas por los usuarios ciegos al usar cada uno de los sitios web.

Caso 1. Sitio web BCPGR (Biblioteca 1)

Tarea 1. El criterio de éxito era encontrar el botón de Accesibilidad en el sitio y seleccionar la opción de sólo texto.

Datos cualitativos por usuario

- Usuario 1. Tiene dificultades con el uso del teclado para moverse por la web; al final no le fue posible cumplir con la tarea. Comenta: “No tengo idea de donde podría encontrar el botón de Accesibilidad”
- Usuario 2. Logró completar la tarea después de familiarizarse con el sitio. Comenta: “Debería ser más fácil llegar al apartado”.
- Usuarios 3, 4 y 5. Les tomó bastante tiempo llegar al apartado, y si lo pasaban tenían que escuchar todo nuevamente. Comentaron: “Muy complicado de encontrar”.

Datos cuantitativos generales

Como se logra observar en la *Figura 6*, 20 % de los usuarios logró realizar la tarea, mientras que 60 % de los usuarios cumplieron con la tarea con bastante dificultad y 20 % restante no llegó al objetivo.

Si bien la mayoría de los usuarios lograron cumplir con la tarea aun con dificultades, se concluye que el apartado de Accesibilidad no tiene la mejor ubicación pues las personas que precisan de esta opción se desesperan y comentan que es difícil de encontrar.



Figura 6. Nivel de cumplimiento de la Tarea 1 (Caso 1)

Tarea 2. Tiene como criterios de éxito consultar los datos en la biblioteca deseada, buscar un libro por autor y marcar el resultado.

Datos cualitativos por usuario

- Todos los usuarios. Tuvieron dificultades para lograr el objetivo y sólo 40 % completó la tarea con ayuda del facilitador para saber si ya estaba marcada la búsqueda, debido a que el programa no se los notifica. Comentaron: “Debería de contar con notificaciones o sonidos que permitan saber si puedes escribir en el formulario o si ya marcaste un libro”.

Datos cuantitativos generales

La Figura 7 muestra que 40 % de los participantes logró realizar la tarea con éxito, mientras que 60 % de ellos tuvieron inconvenientes para llegar al objetivo y no pudieron cumplir.

Por lo tanto, se concluye que el formulario es deficiente, las instrucciones de llenado no son claras y la página no “avisa” al usuario que puede escribir en el formulario; además, no se sabe cuándo se puede marcar un resultado o si ya lo hicieron.



Figura 7. Nivel de cumplimiento de la Tarea 2 (Caso 1)

Tarea 3. El criterio de éxito fue buscar un libro por materia y dejar un comentario.

Datos cualitativos generales

- Ningún usuario pudo cumplir con esta tarea, debido a que la mayoría no sabía llenar el formulario. Los usuarios comentaron que “El formulario debería limpiarse por sí solo y además debería ser más sencillo”.

Datos cuantitativos generales

Como se puede observar en la *Figura 8*, 100 % de los usuarios no logró realizar la tarea, ya que cada uno de ellos encontró algún obstáculo dentro de la interfaz que no permitió que cumplieran con la misma.



Figura 8. Nivel de cumplimiento de la Tarea 3 (Caso 1)

Se concluye nuevamente que el formulario es deficiente aun para la modalidad de sólo texto y que sería mejor si fuera más minimalista.

Tarea 4. El criterio de éxito fue buscar un libro en versión sonora y volver al formulario de búsqueda.

Datos cualitativos generales

- Ninguno pudo cumplir con la tarea y si ocurría un error en la búsqueda no eran notificados, lo cual es importante para mantener el control del usuario sobre el sitio web. Comentaron: “El formulario es complejo y cuando se comete un error no te dice por qué, además hace falta un botón para volver”.

Datos cuantitativos generales

En la *Figura 9* se observa que ningún usuario logró realizar esta tarea, todos tuvieron dificultades que no pudieron superar y no llegaron al objetivo.



Figura 9. Nivel de cumplimiento de la Tarea 4 (Caso 1)

Se concluye que el formulario debería ser fácil de llenar y que cuando se presente un error el lector de pantalla debería poder leer al usuario el porqué de dicho error, preguntando si desean volver a un estado anterior del sistema.

Resumen de cumplimiento de tareas Caso 1. En la *Tabla 5* se presenta un resumen con base en el cumplimiento o no de las tareas propuestas para el primer caso de estudio con la Biblioteca 1.

Usuarios / Tareas	Tarea 1	Tarea 2	Tarea 3	Tarea 4
Usuario 1	X	X	X	X
Usuario 2	0	0	X	X
Usuario 3	0	0	X	X
Usuario 4	0	X	X	X
Usuario 5	0	X	X	X

Cumplimiento de la tarea - 0

No cumplimiento de la tarea - X

Tabla 5. Tabla de cumplimiento para las tareas propuestas

En general y con base en los resultados, se observa que el sitio web no cumple con un diseño accesible y usable para ser utilizado por personas ciegas.

Caso 2. Sitio web de la Biblioteca Iberoamericana (Biblioteca 2)

Tarea 1. El criterio de éxito fue buscar un libro por autor, establecer un rango de años en el formulario, identificar el tercer resultado y marcarlo.

Datos cualitativos por usuario

- Todos los usuarios tuvieron dificultades para llenar el formulario pues no decía a qué campo pertenecía el espacio que seleccionaban. Comentaron: “No sé en qué apartado me encuentro, no sé cómo limitar la búsqueda por año”, “Hay algunas opciones que no entiendo porque no están en español, sin embargo, pude completar la tarea”, “El formulario de búsqueda no me permite regresar a las opciones del inicio por eso me tomará más tiempo”.

Datos cuantitativos generales

En la *Figura 10* se observa que 40 % de los usuarios logró realizar la tarea, mientras que 60 % restante tuvo dificultades y no llegaron al objetivo.



Figura 10. Nivel de cumplimiento de la Tarea 1 (Caso 2)

Por lo tanto, se concluye que existen problemas de diseño en cuanto a la usabilidad, sobre todo en la parte del desplazamiento por medio del teclado.

Tarea 2. Los criterios de éxito eran buscar el tema, seleccionar un formato específico y abrir el resultado.

Datos cualitativos por usuario

- Todos los usuarios. Se presentaron inconvenientes con el formulario, debido a que algunos usuarios no pudieron regresar para hacer otra búsqueda. Comentaron: “No sé qué es lo que está pasando en la pantalla porque no me arroja los resultados de la búsqueda que realicé”, “Para ir al formulario nuevamente no hay un botón para volver lo cual me facilitaría realizar la consulta”.

Datos cuantitativos generales

En la *Figura 11* se muestra que sólo 20 % de los usuarios logró realizar con éxito esta tarea, mientras que 80 % restante tuvo inconvenientes que le impidieron cumplir con el objetivo.



Figura 11. Nivel de cumplimiento de la Tarea 2 (Caso 2)

Se concluye que el formulario tiene deficiencias para la mayoría de los usuarios, las cuales han sido identificadas y tomadas en cuenta para la propuesta de la guía.

Tarea 3. El criterio de éxito era ir a las búsquedas anteriores, seleccionar una y borrarla de la búsqueda.

Datos cualitativos por usuario

- Aunque a los usuarios les costó trabajo entender la tabla de resultados, la mayoría pudo seleccionar una opción y borrarla tal y como se describe en la tarea. Comentaron: “El menú principal tiene muchas opciones, pero en realidad es muy útil”, “La tabla de resultados es un poco complicada de interpretar, pero es cuestión de tomarte el tiempo de explorarla”.

Datos cuantitativos generales

En la Figura 12 se observa que 80 % de los usuarios logró realizar con éxito esta tarea, mientras que sólo 20 % no llegó al objetivo.



Figura 12. Nivel de cumplimiento de la Tarea 3 (Caso 2)

En este caso se observa que las actividades dentro del sitio web que no requieren del llenado de un formulario resultan más sencillas de completar, por lo que es importante tomar medidas y realizar adecuaciones de usabilidad y diseño al sitio para obtener una mejor respuesta por parte de los usuarios objetivo.

Resumen de cumplimiento de tareas Caso 2. En la *Tabla 6* se presenta el resumen del cumplimiento o no de las tareas propuestas para la prueba del segundo caso de estudio.

Usuarios / Tareas	Tarea 1	Tarea 2	Tarea 3
Usuario 1	X	X	X
Usuario 2	0	0	0
Usuario 3	0	X	0
Usuario 4	X	X	0
Usuario 5	X	X	0

Cumplimiento de la tarea - 0
No cumplimiento de la tarea - X

Tabla 6. Tabla de cumplimiento para las tareas propuestas (Caso 2)

En general, los usuarios presentaron dificultades para cumplir con las tareas, sobre todo con el formulario de búsqueda, debido a que no se adapta a las pautas de accesibilidad y compatibilidad con lectores de pantalla; sin embargo, el diseño es sencillo pues no cuenta con mucho texto que pueda desorientar a los usuarios. Otra ventaja es que el menú siempre se encuentra al principio y no hay necesidad de buscarlo.

A manera de conclusión, se logra observar que en ambos casos de estudio los sitios web comparten problemas de diseño que pueden ser fáciles de resolver, también se percibe que el error más común se da al llenar un formulario y la compatibilidad con el lector de pantalla y el sitio web. Con base en lo anterior y la retroalimentación por parte de los usuarios se realiza una guía de diseño para el desarrollo y mejora de sitios web enfocados en el área de las bibliotecas digitales que pretenden ser accesibles para usuarios con discapacidad visual.

PROPUESTA DE GUÍA PARA SITIOS WEB EN BIBLIOTECAS CENTRADOS EN USUARIOS CIEGOS

A continuación se presentan los puntos de la guía de diseño, los cuales se pueden aplicar a sitios web enfocados en bibliotecas digitales con apartados de accesibilidad:

1. En todos los casos se debe verificar que exista compatibilidad y retroalimentación entre lo que está haciendo el usuario y su lector de pantalla.
2. Se debe revisar que exista un buen desplazamiento por el sitio web haciendo uso del teclado.
3. Con énfasis en el primer punto, tener presente que debe existir un acoplamiento al lector de pantalla respecto a las descripciones alternativas, asegurando que toda la información posible sea fácil de leer por el propio lector de pantalla.
4. Se debe revisar la existencia de textos alternos garantizando así que el usuario sabrá lo que está sucediendo en la pantalla.
5. Se debe agregar siempre un botón de atrás/volver en todos los apartados, que sea fácil de encontrar.
6. Si el sitio cuenta con formularios, deben ser simples, todo en él debe poder ser identificado y leído por el lector de pantalla y debe preferentemente limpiarse automáticamente cada vez que se finalice una búsqueda.

7. Si el sitio web ya cuenta con algunas pautas de accesibilidad, el apartado de “Accesibilidad” donde se puede acceder a ellas, es decir, activarlas o usarlas, debe ser fácil de localizar.

Con estos sencillos pasos se puede marcar la diferencia entre cumplir o no con el objetivo del sitio web y las características del usuario objetivo.

Guía de diseño aplicada a los casos de estudio

En los siguientes apartados se muestran los cambios realizados a cada uno de los sitios web con base en la propuesta de la guía de diseño para sitios especializados en bibliotecas digitales.

Caso 1: Biblioteca 1 (BCPGR)

En el primer caso de estudio se identificaron cinco problemas relevantes, los cuales no pudieron ser resueltos por el usuario. En la Tabla 7 se muestran algunos de los cambios relevantes (debido a su extensión), tomando en cuenta la guía de diseño.

Problema identificado	Resolución con base en la guía/Imagen	Observaciones
Falta notificación al marcar un libro.		Agregar un texto alternativo al <i>check</i> que retroalimente al usuario, dependiendo de si está activo o no.
Falta botón “atrás” (volver al formulario de búsqueda).		Agregar un botón “volver / atrás”, posiblemente en la parte inferior para que tenga oportunidad de evaluar todas las opciones.

Tabla 7. Solución de problemas haciendo uso de la guía de diseño

Se espera que con estos pequeños ajustes mejore considerablemente la experiencia del usuario ciego en la consulta de libros en este sitio web.

Caso 2: Biblioteca 2 (BIOP)

Por la extensión del documento se visualizan algunos cambios una vez aplicada la guía propuesta. En este caso se identificaron cinco problemas relevantes, los cuales no pudieron ser resueltos por el usuario. En la *Tabla 8* se muestran los cambios relevantes, tomando en cuenta la guía de diseño.

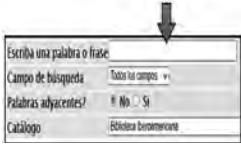
Problema identificado	Resolución con base en la guía/Imagen	Observaciones
Falta notificación (verbal o sonora) que indique si es posible escribir en el formulario.		Notificar si está activada o no la función de escritura.
Tabla de resultados compleja para verbalizar.		Debe ser fácil verbalizar la información. Además, en este caso se cuenta con espacios que usualmente no se usan.

Tabla 8. Solución de problemas haciendo uso de la guía de diseño

CONCLUSIONES

La inclusión en el ámbito tecnológico resulta ser un aspecto importante para el desarrollo académico, laboral y cultural de personas con discapacidad visual. Brindar accesibilidad web a sitios centrados en bibliotecas digitales no es muy complicado. Con esta guía de diseño básica es fácil brindar una mejor experiencia al usuario; teniendo en cuenta los factores que repercuten en la misma, le permite al usuario sentirse más orientado al saber qué es lo que está sucediendo en la pantalla por medio de notificaciones, sonidos o instrucciones que se mencionan en la propuesta de este artículo.

Implementar la guía de diseño desde una fase temprana del desarrollo del sitio web daría paso a una mejor adaptación por parte de las pautas y las funciones principales del sitio. Con la propuesta de diseño basada en los co-

mentarios y pruebas emitidos por los usuarios ciegos, es posible contemplar a grandes rasgos una forma simple de evaluar cualquier otro sitio web relacionado con las bibliotecas digitales, específicamente en catálogos web, y de esta manera complementar las pautas de accesibilidad.

Los resultados de las pruebas con usuarios ciegos revelaron que, en este caso, para ambos catálogos web el problema heurístico que más se destacó por su deficiencia fue el de control y libertad del usuario, esto puede proyectarse sobre todo en el llenado de los formularios de búsqueda.

La propuesta de diseño es simple pues incluye mejor compatibilidad con el lector de pantalla, desplazamiento óptimo por medio del teclado, notificaciones verbales o sonidos que indiquen lo que sucede en la pantalla, opciones para volver y mejoras en el formulario de búsqueda.

A futuro se toma la posibilidad de hacer evaluaciones similares con diferentes sitios web con el mismo propósito y fomentar la participación de las personas con discapacidad visual en bibliotecas que cuentan con recursos para ellos.

Agradecimientos

Se agradece el apoyo a la Asociación de Invidentes Unidos de Jalisco A. C. de la ciudad de Guadalajara, Jalisco, México: al coordinador y sus integrantes por participar en las pruebas de este trabajo de investigación. También se agradece al Programa Interinstitucional para el Fortalecimiento de la Investigación y el Posgrado del Pacífico que hizo posible esta investigación.

REFERENCIAS

- Álvarez-Robles, T. 2019. “Metodología para la evaluación de la experiencia del usuario de sistemas de software interactivos para usuarios ciegos”. Tesis doctoral, Universidad Veracruzana, Facultad de Estadística e Informática, México.
- Arancibia Céspedes, D. y S. González Canales 2017. “Heurísticas de Usabilidad y Experiencia del usuario en redes sociales”. Informe final de proyecto, Pontificia Universidad Católica de Valparaíso, Chile.
http://opac.pucv.cl/pucv_txt/txt-4500/UCC4799_01.pdf
- Carrillo, M. 2018. “En México millones de personas con discapacidad visual”, *El Dictamen*, 14 de octubre de 2018.
<https://www.eldictamen.mx/nacional/en-mexico-millones-de-personas-con-discapacidad-visual/>
- D’Hertefelt, S. 2000. “Emerging and future usability challenges: designing user experiences and user communities”, *Interaction Architect*.
- El Informador*. 2010a. “Guadalajara, una ciudad que no ve a los invidentes”, 11 de abril de 2010.

- <https://www.informador.mx/Jalisco/Guadalajara-una-ciudad-que-no-ve-a-los-invidentes-20100411-0219.html>
- El Informador*. 2010b. "En Jalisco el 10% de bibliotecas tiene un área para invidentes", 28 de julio de 2010.
<https://www.informador.mx/Cultura/En-Jalisco-el-10-de-bibliotecas-tiene-un-area-para-invidentes-20100728-0234.html>
- El Informador*. 2018. "Un espacio renovado para la lectura", 7 de junio de 2018.
<https://www.informador.mx/cultura/Un-espacio-renovado-para-la-lectura-20180606-0166.html>
- Ericsson, K. y H. Simon. 1980. "Verbal reports as data." *Psychological review* 87 (3): 215.
<https://doi.org/10.1037/0033-295X.87.3.215>
- Granollers, T., J. Lorés y J. Cañas. 2011. *Diseño de sistemas interactivos centrados en el usuario*. Barcelona: Editorial UOC.
- Hughes, M. 2012. *Uxmatters*. Marzo. Fecha de consulta: agosto de 2020.
<https://www.uxmatters.com/mt/archives/2012/03/talking-out-loud-is-not-the-same-as-thinking-aloud.php>
- Martínez de la Teja, G. 2006. *Usabilidad y Accesibilidad en Web*.
<http://www.semac.org.mx/archivos/6-11.pdf>
- Morville, P. 2004. "User Experience Design". *Semantic Studios*, 21 de junio de 2004.
https://semanticstudios.com/user_experience_design/
- Nielsen, J. 1994. *10 Usability Heuristics for User Interface Design*. NN/g Nielsen Norman Group, 24 de abril de 1994, última actualización 15 de noviembre de 2020.
<https://www.nngroup.com/articles/ten-usability-heuristics/>
- Red Estatal de Bibliotecas Públicas de Jalisco. 2015.
<http://consultareb.jalisco.gob.mx/cgi-bin/abnopac/O7032/ID79ec-65b3?ACC=101>
- Reyes Vera, J. M., M. I. Berdugo Torres y L. Machuca Villegas. 2016. "Evaluación de usabilidad de un sistema de administración de cursos basado en la plataforma Lingweb." *Ingeniare. Revista chilena de ingeniería* 24 (3): 435-444.
<https://doi.org/10.4067/S0718-33052016000300008>
- Universidad de Guadalajara. 2009. *Biblioteca Iberoamericana*. 2009.
http://148.202.105.23:8991/F/-/?func=find-b-0&local_base=IBE01
- W3C. 2005. *Web Content Accessibility Guidelines (WCAG) Overview*. Julio de 2005. Fecha de consulta: 22 de junio de 2018.
<https://www.w3.org/WAI/standards-guidelines/wcag/>

Para citar este texto:

Álvarez-Robles, Teresita, Yadira Alejandra Orozco Osuna y Francisco Javier Álvarez Rodríguez. 2021. "Evaluación UX para sitios web orientados a bibliotecas digitales con usuarios ciegos". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 169-194.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58451>

Anexo A. Formato lista de tareas Biblioteca Central

Lista de tareas (1)

Considere el siguiente escenario:

Usted como estudiante está realizando una investigación sobre Juan Rulfo y la cultura mexicana. La biblioteca más cercana es la Biblioteca Central Profesor García Ruíz. Por lo que antes de ir a ella decide consultar la existencia de libros que van a ser de su interés y en algún formato sonoro para su comodidad, usando el catálogo web de la biblioteca (<http://consultareb.jalisco.gob.mx/cgi-bin/abnopac/O7063/ID25bd74c0/NT2?ACC=101>).

Se les pide realizar las siguientes tareas, *comentando en voz alta durante todo el experimento*.

- Tarea 1

Buscar el apartado de accesibilidad y seleccionar la versión de solo texto.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P10	Apartado de accesibilidad se encuentra al final de la página	

- Tarea 2

Buscar en el catálogo de GDL / Biblioteca Central por autor un libro de “Juan Rulfo” publicado entre 1950 -1970 y marcar un resultado.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P1	Falta notificación al marcar un libro	
P3	No hay instrucciones de llenado de formulario	
P5	Botón BUSCAR despliega el catálogo completo (150 elementos)	
P7	Instrucciones para buscar aparecen al final	
P8	Siempre aparece el apartado de usuario al principio	
P14	Botón con abreviatura (Inf. Biblioteca)	
P15	Botón Inf. Biblioteca no parece llegar a ningún apartado especial	

- Tarea 3

Buscar por materia “Cultura México” usando la opción de índice automático. Abrir uno de los resultados. Abrir alguno de los libros que aparecen en pantalla y dejar un comentario.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P6	Palabras abreviadas (Ind.Aut.)	
P9	Texto desplegable que no se lee automáticamente (comentarios)	
P12	El formulario no se limpia por sí solo	

- Tarea 4

Buscar por título el libro de “Pedro Páramo”, seleccionar el *Tipo de Acervo* como *Sonoros*. Identificar las opciones que el programa ofrece ante problema (La consulta no recupera resultados) y volver al formulario para realizar una nueva búsqueda.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P2	Difícil acceso desde el teclado	
P4	Apartado no cuenta con el botón volver (resultados)	
P11	Formulario desplegable no se lee usando el puntero (acervo)	
P13	Oración que no se lee de corrido	

Anexo B. Formato lista de tareas Biblioteca Iberoamericana

Lista de tareas (2)

Considere el siguiente escenario:

Usted como estudiante está realizando una investigación sobre Juan Rulfo y la cultura mexicana. La biblioteca más cercana es la Biblioteca Iberoamericana Octavio Paz. Por lo que antes de ir a ella decide consultar la existencia de libros que van a ser de su interés y en algún formato de su comodidad, usando el catálogo web de la biblioteca (http://148.202.105.23:8991/F/-/?func=find-b-0&local_base=IBE01).

Se les pide realizar las siguientes tareas, *comentando en voz alta durante todo el experimento*.

- Tarea 1

Buscar un libro de “Juan Rulfo”, limitar la búsqueda entre el año 1950 -1970 y marcar el resultado número 3.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P1	No accesible desde el teclado al menú principal	
P3	No se sabe si la escritura está habilitada	
P4	Formulario expandible que no se lee de forma automática	
P6	Botones ilegibles para el lector de pantalla (ir al #)	
P7	Tablas de difícil acceso desde el teclado	
P11	Marcar resultados es difícil	
P12	Sugerencia para introducir datos al formulario no adaptado para lectores de pantalla (fecha)	
P14	Opciones en inglés (opciones de ordenamiento)	

- Tarea 2

Buscar a cerca de la “cultura mexicana” en formato de *archivos de computadora* y abrir un resultado.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P2	Notificaciones de error que no se leen hasta que las apuntas (formulario)	
P5	No existe un botón para volver	
P8	Se lee el menú cada que cambias de opción	
P9	Las imágenes que aparecen no son accesibles	
P10	Abreviatura en la palabra Biblioteca (menú principal)	
P15	Formulario de búsqueda no se limpia por sí solo	

- Tarea 3

Ir al apartado de búsquedas anteriores, seleccionar una y borrarla.

ID del problema	Descripción del problema	¿Fue de gran impacto para el usuario?
P13	Botones no legibles para el lector de pantalla (búsquedas anteriores)	

Olhando a dimensão epistemológica da organização do conhecimento brasileira com três lentes da análise de domínio: epistemológica, estruturação científica e terminológica

Maíra Fernandes Alencar*
Brígida Maria Nogueira Cervantes*

Artículo recibido:

5 de mayo de 2021

Artículo aceptado:

2 de agosto de 2021

Artículo de investigación

RESUMO

Considerando a importância de uma comunidade científica internacional para o cultivo de reflexões e pesquisas, selecionamos para ser mapeada a ISKO-Brasil. O objetivo do estudo é apresentar um mapeamento dos trabalhos que foram apresentados na dimensão epistemológica. A metodologia utilizada foi a análise de conteúdo para fins de categorização e análise de domínio com três enfoques de estudos: 1) Epistemológica - para identificar as temáticas das pesquisas; 2) Estrutura da comunicação científica - cujo foco está em descrever ano, autores e instituições; e por fim, 3) Estudos terminológicos, ao extrair dos títulos das publicações termos representativos, para categorizar e cotejar com o Tesouro Brasileiro

* Centro de Educação, Comunicação e Artes (CECA), Universidade Estadual de Londrina, Brasil alencarmfa@gmail.com brigidacervantes@gmail.com

de Ciência da Informação e o *Diccionario de organización del conocimiento: Clasificación, Terminología, Indización*. Os resultados revelam que as temáticas são predominantes de base conceitual, seguidas de diálogos interdisciplinares; os trabalhos estão a cargo de 26 (vinte e seis) instituições com maior incidência da Universidade Estadual Paulista “Júlio de Mesquita Filho” Campus – Marília; e no que tange aos estudos terminológicos, com base na categorização, os termos que prevalecem nos títulos são provenientes das áreas de Tratamento Temático da Informação e de Sistemas de Organização do Conhecimento, com um nível médio de compatibilização nos instrumentos terminológicos.

Palavras-chave: ISKO-BRASIL; Organização do Conhecimento; Análise de Domínio; Dimensão Epistemológica

Mirando la dimensión epistemológica de la organización del conocimiento brasileño con tres lentes de análisis de dominio: epistemológico, de estructuración científica y terminológico

Maira Fernandes Alencar e Brígida Maria Nogueira Cervantes

RESUMEN

Considerando la importancia de una comunidad científica internacional para el cultivo de las reflexiones e investigaciones, seleccionamos para ser mapeada la ISKO-Brasil. El objetivo del estudio es presentar un mapa de los trabajos que se presentaron en la dimensión epistemológica. La metodología utilizada fue el análisis de contenido para la categorización y el análisis de dominio con tres estudios de enfoque: 1) Epistemológico - para identificar los temas de investigación; 2) Estructura de la comunicación científica - cuyo enfoque es describir el año, los autores y las instituciones, y finalmente 3) Estudios terminológicos, mediante la extracción de términos representativos de los títulos de las publicaciones, para categorizar y comparar con el Tesoro Brasileño de Ciencias de la Información y el *Diccionario de organización del conocimiento: Clasificación, Indización, Terminología*. Los resultados revelan que los temas son predominantemente de base conceptual, seguidos de diálogos interdisciplinarios; los trabajos están a cargo de 26 (veintiséis) instituciones con mayor incidencia de la Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus Marília;

y en cuanto a los estudios terminológicos, con base en la categorización, los términos que prevalecen en los títulos provienen de las áreas de Tratamiento Temático de la Información y Sistemas de Organización del Conocimiento, con un nivel medio de compatibilidad en los instrumentos terminológicos.

Palabras clave: ISKO-BRASIL; Organización del Conocimiento; Análisis de Dominio; Dimensión Epistemológica

Looking at the epistemological dimension of Brazilian knowledge organization with three lenses of domain analysis: epistemological, scientific structuring and terminology

Maíra Fernandes Alencar and Brígida Maria Nogueira Cervantes

ABSTRACT

The objective of the study is to present a mapping of the works that were presented in the epistemological dimension. The methodology used was content analysis for purposes of categorization and domain analysis with three study approaches: 1) Epistemological - to identify the themes of research; 2) Structure of scientific communication - whose focus is on describing year, authors and institutions; and finally, 3) Terminological studies, by extracting from the titles of publications representative terms, to categorize and collate with the Brazilian Thesaurus of Information Science and the *Diccionario de organización del conocimiento: Clasificación, Indización, Terminología*. The results reveal that the themes are predominant on a conceptual basis, followed by interdisciplinary dialogues; the work is carried out by 26 (twenty-six) institutions with the highest incidence at the Universidade Estadual Paulista "Júlio de Mesquita Filho" *Campus* - Marília; and with regard to terminological studies, based on categorization, the terms that prevail in the titles come from the areas of Thematic Treatment of Information and Knowledge Organization Systems, with a medium level of compatibility in terminological instruments.

Keywords: ISKO-BRASIL; Knowledge Organization; Domain Analysis; Epistemological Dimension

INTRODUÇÃO

Os estudos sobre o saber registrado e socializados em rodas de conversas, contações de histórias, filmes, livros, documentos de arquivos, periódicos científicos, fotografias (entre tantos outros), desde as concepções de suas naturezas e interdisciplinaridades até as teorias e metodologias em como torná-los acessíveis, sustentam diferentes perspectivas no âmbito da Organização do Conhecimento (OC), com distintos níveis de interesses, intenções e ideologias.

Essas perspectivas se concretizam nas diversidades de 1) temáticas (sistemas e processos, com assuntos variáveis dos fenômenos que giram em torno do que é o conhecimento, a informação e seus desdobramentos em teorias e práticas profissionais); e 2) enfoques (sociais, políticos, críticos, econômicos, culturais, entre outros) nas pesquisas desenvolvidas.

Uma sociedade científica que contribui para fomentar os estudos dessas perspectivas é a ISKO Brasil (dentre os demais objetivos institucionais da sociedade) que se configura como um espaço teórico de diálogos e discussões em torno das questões da OC. Assim, ressaltamos que ter um espaço investigativo científico como este capítulo nacional da ISKO Brasil é fundamental para o desenvolvimento da ciência no Brasil.

Nesse sentido, Dodebei e Guimarães (2013: 15) situam que a área de Organização e Representação do Conhecimento no Brasil “enquanto um campo de estudos já consolidado nos domínios da Ciência da Informação e, em outros fronteiriços, notamos uma forte aproximação com seus fundamentos teórico-conceituais e metodológicos”.

Na ISKO do Brasil sua estrutura de trabalhos divide-se em três categorias, denominadas “dimensões”, sendo estas:

- dimensão epistemológica;
- dimensão aplicada; e
- dimensão político - social.

Ao analisarmos essas diferentes dimensões, é notável uma pluralidade de assuntos, o que corrobora para as discussões em torno da OC em que as pesquisas podem ser contínuas, e também tecer caminhos de reflexões, críticas, análises e construções que ampliem as reflexões sobre as questões do conhecimento, da informação e suas interfaces com os diferentes domínios.

Vale destacar que a palavra olhar no título, como metáfora ilustra aquilo que vemos, sempre mediada por lentes específicas, nesse caso, estas, são a análise de domínio.

O que iremos “olhar”, partiu da pesquisa de Santos, Hernandez e Vital (2019) (que mapearam as temáticas estudadas e a comunidade discursiva por meio da produção científica da *dimensão política e social* do capítulo da ISKO-Brasil). Nesse sentido, essa pesquisa parte da seguinte questão: como se caracteriza a dimensão epistemológica da ISKO-Brasil?

Nosso objetivo é apresentar uma análise e mapeamento no que tange a sua produção científica.

UMA TEIA DENOMINADA “EPISTEMOLOGIA” NA ORGANIZAÇÃO DO CONHECIMENTO

Para nos aproximarmos do tema de epistemologia, selecionamos uma poesia de Orides Fontela (1996), poeta brasileira contemporânea¹, que recebe o título de *Teia*:

A teia, não
mágica
mas arma, armadilha
a teia, não
morta
mas sensitiva, vivente
a teia, não
arte
mas trabalho, tensa
a teia, não
virgem
mas intensamente
prenhe:
no
centro
a aranha espera.

Em síntese, e sem reduzir as possibilidades de análises desse poema, nos interessa o que está estruturando adjetivos e os objetivos sobre a teia. O que é a teia? Quem constrói a teia? Para que e a quem a teia serve? Como ela é

1 Nascida em São João da Boa Vista (SP), Orides começou a escrever aos 7 anos e publicou seus primeiros poemas no jornal *O Município*, periódico de sua terra natal. Ao todo, a autora teve sete livros publicados, uma obra densa, que recebeu, entre outros reconhecimentos, o Prêmio Jabuti de Poesia em 1983, pelo livro *Alba* (Castro, 2015).

construída? Essas perguntas se feitas sobre a epistemologia, nos parece primordial, para situarmos a amplitude da temática e seu grau de complexidade.

Nesse sentido, iremos expor brevemente sobre seu conceito em dois diferentes dicionários de filosofia, e posteriormente, situá-la no âmbito específico da OC.

Epistemologia² no âmbito da filosofia é designado como teoria do conhecimento, “em inglês o termo *Epistemology* foi introduzido por J.F. Ferrier (Institutes of Metaphysics, 1854) [...]” (Abbagnano, 1998: 183).

Interessa-nos também a definição de Japiassú e Marcondes do *Dicionário Básico de Filosofia* (2006: 67) que atribui epistemologia (do gr. *episteme*: ciência, e *logos*: teoria) como “Disciplina que toma as ciências como objeto de investigação [...]”. Para os autores enquanto disciplina ela tenta reagrupar 3 aspectos: a) a crítica do conhecimento científico; b) a filosofia das ciências; c) a história das ciências.

Em outro momento adiante, os autores especificam que a epistemologia é uma disciplina proteiforme:

No pensamento anglo-saxão, epistemologia é sinônimo de teoria do conhecimento (ou gnoseologia), sendo mais conhecida pelo nome de “*philosophy of science*”. E neste sentido que se fala de epistemologia a propósito dos trabalhos de Piaget versando sobre os processos de aquisição dos conhecimentos na criança. O fato é que um tratado de epistemologia pode receber títulos tão diversos como: “A lógica da pesquisa científica”, “Os fundamentos da física”, “Ciência e sociedade”, “Teoria do conhecimento científico”, “Metodologia científica”, “Ciência da ciência”, “Sociologia das ciências” etc. Por essa simples enumeração, podemos ver que a epistemologia é uma disciplina proteiforme que, segundo as necessidades, se faz “lógica”, “filosofia do conhecimento”, “sociologia”, “psicologia”, “história” etc. (Japiassú e Marcondes, 2006: 67).

Podemos observar com base nessa diversidade de títulos o quanto tratar desse tema é complexo, haja visto que as variantes de possibilidade que podem ter sua concepção e denominação, razão pela qual não aprofundaremos detalhadamente para não sair do escopo do trabalho. O que nos interessa é situar que a OC ao ser um campo científico sistematizado, também se insere nessa discussão.

Por fim, para passarmos para a OC, identificaremos antes, a concepção de Japiassú e Marcondes (2006: 67-68), em relação ao modo de como eles identificam o problema central da epistemologia:

2 “Muitos questionamentos existem em torno do termo epistemologia. Esses questionamentos vão desde seus domínios, enquanto disciplina científica (Blanché, 1983; Japiassu, 1986), até seu aparecimento e etimologia (Carrilho; Sáaagua, 1991)” (Francelin, 2005: 103).

Seu problema central, e que define seu estatuto geral, consiste em estabelecer se o conhecimento poderá ser reduzido a um puro registro, pelo sujeito, dos dados já anteriormente organizados independentemente dele no mundo exterior, ou se o sujeito poderá intervir ativamente no conhecimento dos objetos. Em outras palavras, ela se interessa pelo problema do crescimento dos conhecimentos científicos. Por isso, podemos defini-la como a disciplina que toma por objeto não mais a ciência verdadeira de que deveríamos estabelecer as condições de possibilidade ou os títulos de legitimidade, mas as ciências em via de se fazerem, em seu processo de gênese, de formação e de estruturação progressiva.

Ao tomar como objeto a ciência em seu processo de se fazer, entendemos que a epistemologia em analogia a poesia de Orídes é de tal modo como uma teia, não apenas pelo devir de se ocupar de algo em construção, mas também por vias dos adjetivos que as qualificam, considerando que teia = epistemologia, e aranha = disciplina do conhecimento científico. Partindo dessa concepção que teia e aranha são diretamente relacionadas entre si, em uma relação de causa e efeito toda aranha faz suas teias, assim, todas as disciplinas científicas têm suas posições epistemológicas, ditas de formas explícitas ou não.

Continuando com esse argumento da teia como epistemologia, vale mais um elemento a ser considerado, a teia é uma construção solitária. A aranha tem um propósito: manter-se viva por meio de alimentos. Então ela escolhe um lugar ideal para construir uma armadilha. Tece fio a fio. E no final o que vemos? Uma construção geométrica perfeita, resistente. Fruto de um ser que teve muita destreza para construir e, muita calma para atingir o seu objetivo: devorar insetos ou animais que porventura fiquem presos em qual ponta da teia. Sim! A aranha construiu um método (aproveitando das sensibilidade que há nas suas patas) para sentir qualquer movimento em qualquer ponto da teia. A teia só foi possível existir, pois a aranha tem “compreensão” do mundo à sua volta.

As disciplinas do conhecimento científico assim como as “aranhas” constroem métodos, formas de compreender o mundo, e com isso tecem epistemologias, caminhos para compreender o como sabemos o que sabemos.

Situado dessa forma, não há magias, nem estagnação, o que há são trabalhos, gestações, todas construções dotadas de interesses e voltadas para usos que podem ser direcionados para a própria ciência, para a sociedade e/ou mercado, para uma cultura de violência ou de paz, entre tanto outros, éticos ou não. E no centro, quem espera são pessoas que criam, escolhem e/ou usam determinada sistematização de saberes para descrever e orientar formas de pensamentos, condutas de ações para vida, tipos de crenças, enfim, todos elementos de um processo feito com e para demonstrar quais são e como se obtém conhecimento, e que conhecimento é válido para quê, para quem e suas causas.

Após termos explanado essa aproximação da epistemologia enquanto uma teia, concordamos com Japiassú e Marcondes (2006: 67-68. Grifo nosso) a respeito do objeto da epistemologia enquanto uma ciência que se constrói, “*em via de se fazerem, em seu processo de gênese, de formação e de estruturação progressiva*” como argumentamos acima.

Como já dissemos, o fato de explanar que é uma construção significa dizer também que além de intenções, tem métodos, e nesse ponto, no *Dicionário de Filosofia* de Abbagnano (1998: 140. Grifo nosso) ao final da descrição do conceito de Teoria da Ciência, consta que “para a disciplina que considera as formas ou os procedimentos do conhecimento científico, têm sido usadas mais frequentemente às palavras *epistemologia* e *metodologia*”.

Visto essas concepções, passamos para a OC, e começamos por esclarecer que essa área como em todo campo de conhecimento científico tem na epistemologia uma temática importante, visto que é a base de toda pesquisa científica, além de ser uma das facetas da Análise de Domínio proposta por Hjørland (2002).

Além disso, na ISKO-Brasil há um eixo de estudos denominado “Dimensão Epistemológica”, essa dimensão compreende o foco desse estudo. Em razão disso, consideramos que o primeiro aspecto para ser analisado, é explicarmos o que é uma dimensão.

Em um trabalho publicado na ISKO Brasil, Smiraglia (2013: 18) parte dessa questão e descreve a epistemologia enquanto uma dimensão do domínio da OC:

A epistemologia é uma ferramenta essencial da organização do conhecimento e uma dimensão é uma expressão da extensão de um espaço. Portanto, na organização do conhecimento, a epistemologia representa uma dimensão, porque é como podemos medir ou expressar o espaço em nosso domínio, que vai do empírico ao racional, as duas principais posições epistemológicas.

O autor supracitado faz tal colocação, por entender como dimensão o enfoque conceitual oriundo da matemática e da física, para isso, ele usa como exemplo a figura de uma esfera sendo dividida por três dimensões (altura, largura e profundidade) assim, “cada dimensão viaja ao longo de uma daquelas setas que dividem o espaço” (Smiraglia, 2013: 18). Essa metáfora compara a OC como uma esfera, ou seja, um domínio específico em que há diferentes dimensões que divide o espaço intelectual da OC.

Essa dimensão epistemológica que divide o espaço, e fornece *trajetórias* para transitar, é vista também como uma possibilidade de estudos no domínio, como é o caso da tese intitulada “Epistemologia da Organização Do

Conhecimento: Um estudo Metateórico” de Araújo (2019: 133) que busca responder sobre a concepção de epistemologia no domínio da OC, e identifica, dentre outros aspectos da pesquisa, que: “[...] a concepção de epistemologia no domínio da OC é o estudo crítico dos princípios, hipóteses e produção do conhecimento no domínio”. Observamos que essa concepção está alinhada com a primeira característica que Japiassú e Marcondes (2006) citam.

No caso específico da OC a epistemologia está relacionada com a “produção do conhecimento científico pelo domínio da OC, bem como com a aplicação desse conhecimento para o desenvolvimento dos SOC e para dar suporte aos processos de OC” (Araújo, 2019: 133).

Epistemologia é essencial para desenhar e implementar os SOC, guia o processo de OC e destaca o seu nível conceitual, também torna claras as posições epistêmicas que influenciam a OC e ajuda a aplicar a análise de domínio já que provê ideias relacionadas às suposições das teorias e é uma das onze abordagens da análise de domínio. (Araújo, 2019: 132)

Nesse sentido, na pesquisa de Araújo (2019) são identificadas outras posições epistêmicas a partir do conceito de epistemologia:

As posições epistêmicas (atributos) que formam esta família de atributos são: empirismo, racionalismo, historicismo, pragmatismo, realismo construtivista, construtivismo social, holismo, pragmatismo realista, reducionismo, idealismo, relativismo, positivismo lógico, construtivismo, realismo hipotético, realismo, positivismo, pós-modernismo e teorias críticas (pósestruturalismo, desconstrução, epistemologia feminista, pós-colonialismo, teoria crítica, racial e teoria queer). (Araújo, 2019: 132-133)

Hjørland (2002) estuda o campo epistemológico na OC e frequentemente situa como epistemologia na OC quatro correntes: empirismo, racionalismo, historicismo e pragmatismo. Ele utiliza como exemplo para cada uma dessas posições o que pode ser os critérios de relevância, simplificados nessas quatro correntes epistemológicas, com duas categorias: o que é relevante e não relevante.

Quanto a essas categorias que estão representada em uma figura, vale ressaltar que em outro trabalho, o autor defende que tal figura “[...] é obviamente uma simplificação do complexo campo da epistemologia. Fornece, no entanto, uma base para compreender algumas das posições epistemológicas primárias e a sua inegável influência sobre o trabalho da biblioteca e da ciência da informação” (Hjørland e Hartel, 2003: 239).

A respeito do pragmatismo, na tese de Araújo (2019: 133), foi identificado que essa corrente é a posição epistêmica mais influente no domínio da OC;

e que o pragmatismo e historicismo são posições epistêmicas relacionadas, além de uma forte presença das teorias críticas.

METODOLOGIA

A metodologia utilizada foi a Análise de Domínio (AD) de Hjørland (2002) e a Análise de Conteúdo de Bardin (1977), com a Pré-Análise e Exploração do Material nos anais da ISKO “Estudos Avançados em Organização do Conhecimento” volumes 1-5, disponíveis no *website* da ISKO-Brasil:

- 2012 (Brasília). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade*
- 2013 (Rio de Janeiro). *Complexidade e organização do conhecimento, desafios do nosso século*
- 2015 (Marília). *Organização do Conhecimento e Diversidade Cultural*
- 2017 (Recife). *Memória, tecnologia e cultura na organização do conhecimento*
- 2019 (Belém). *Organização do conhecimento responsável: promovendo sociedades democráticas e inclusivas*

A trajetória da pesquisa foi realizada de acordo com as etapas em:

1. Estudos de Estrutura na Comunicação Científica, para identificar quantidade de trabalhos por ano; autores e autoras; e instituições
2. Estudos Epistemológicos e Críticos, para identificar as temáticas dos trabalhos.
 - 2.1 Categorização das temáticas (na fase de Tratamento dos Resultados, Inferências e Interpretações) com base em Dodebei e Guimarães (2013: 17) ao discorrer sobre o que compõe a dimensão epistemológica – bases conceituais, históricas, metodológicas, produção científica e diálogos interdisciplinares.
3. Estudos Terminológicos com o objetivo de verificar se os termos dos títulos estão presentes em dois instrumentos terminológicos - Tesauro Brasileiro de Ciência da Informação, pelo fato da predominância dos autores e autoras serem docentes de programas de pós-graduação em Ciência da Informação e o *Diccionario de organización del conocimiento: Clasificación, Indización, Terminología* por ser específico da temática do evento que está sendo analisado. Aqui não fez uso das palavras-chave, pois no formato dos trabalhos, estas só constam a partir do ano de 2017.

Tennis (2012: 6) distingue dois tipos de AD, a descritiva e instrumental, destacando que “a primeira é usada, e útil, somente em pesquisas básicas, e a última é usada para criar sistemas de organização do conhecimento”.

Com base em Santos, Hernandez e Vital (2019) sistematizamos para esse domínio em estudo, a proposta de Tennis (2012) sobre a importância dos dois eixos para a AD - áreas de modulação e graus de especialização, com a definição do domínio, escopo e alcance e propósito - situados em âmbito descritivo.

RESULTADOS E DISCUSSÕES

Com foco na AD descritiva, apresentamos a definição do domínio, escopo e alcance e propósito, e em seguida os Estudos de Estrutura na Comunicação Científica; Estudos Epistemológicos e Críticos e Estudos Terminológicos no âmbito da ISKO-Brasil.

- *Definição do Domínio.* A análise realizada refere-se à dimensão epistemológica da Organização do Conhecimento, que está presente em todas as edições já realizadas da ISKO-Brasil a partir de 2012, e está “voltada para e a consolidação de teorias, metodologias, paradigmas, escolas de pensamento” (Guimarães, 2015: 15).
- *Escopo e Alcance.* O *corpus* possui 66 trabalhos (2012-2019) provenientes da dimensão epistemológica da OC na ISKO-Brasil. *Extensão:* Todos os trabalhos da dimensão 1. *Exclusão:* Não houve. *Rótulo:* Dimensão Epistemológica da ISKO-Brasil. *Foco e Especialização:* Esta análise tem como foco descrever a quantidade de trabalhos por ano; quem são os autores e os vínculos institucionais presentes, categorizar as temáticas estudadas e cotejar os termos representativos dos títulos dos trabalhos com dois instrumentos terminológicos da área de CI e OC.
- *Propósito.* O propósito é de caráter descritivo, visa conhecer e mapear o domínio de trabalhos na dimensão epistemológica da ISKO-Brasil e sua comunidade discursiva.

Estudos de Estrutura na Comunicação Científica

Foram recuperados o total de 66 trabalhos que compõe o *corpus* da análise. Com base na quantidade de trabalhos por ano, verificamos que o ano de 2017 se destaca em termos quantitativos com um total de 20 trabalhos, seguido de 2013 com um total de 14 trabalhos, conforme *Figura 1*.

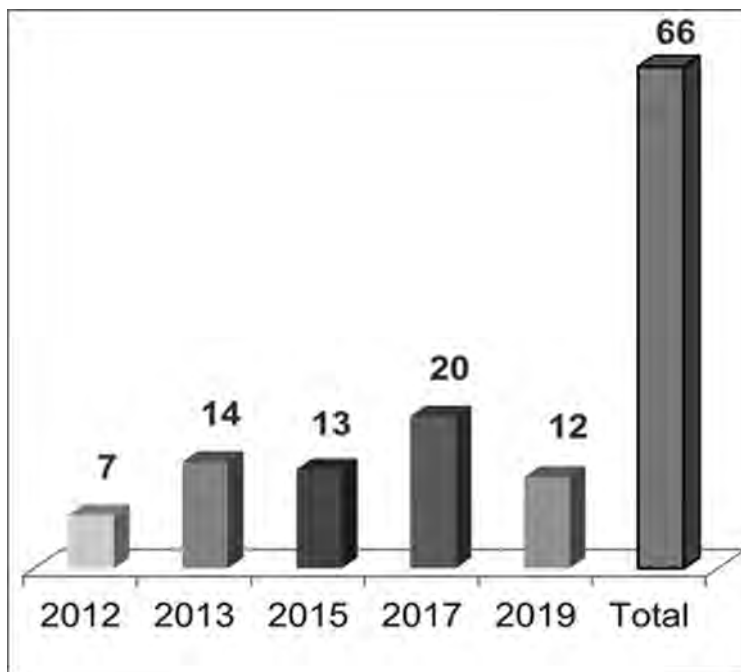


Figura 1. Quantidade de trabalhos por ano da Dimensão Epistemológica da OC – ISKO – Brasil
Fonte: elaborado pelas autoras (2020)

Quanto os autores, há um total de 107, sendo que 71 são autoras e 36 autores. Com análise de frequência, 22 autores/autoras publicam mais de uma vez, sendo: Alves (3); Barros (2); Ortega (2); Martinez-Ávila (4); Sabbag (2); Oliveira (3); Moraes (4); Guimarães (4); Tennis (3); Weiss (2); Bufrem (3); Gracio (3); Fujita (3); Barité (2); Bräscher (3); Silveira (2); Tognoli (3); Dal'Evedove (3); Santos (4); Corrêa (2); Sales (5) e Moreira (2).

Em relação a modalidade de autoria, tem-se com dois autores (29 trabalhos), três autores ou mais (22 trabalhos) e 1 autor (15 trabalhos).

As vinculações dos autores estão a cargo de 26 instituições (sendo 20 Universidades Brasileiras; 1 Diretoria Regional de Ensino e 5 Universidades Internacionais). Para fins quantitativos, destacam-se oito instituições, a UNESP - FFC/Marília (32); UFF (10); UFSC (7); UFSCAR (6); UFPB (6); UFRJ (5); UFPE (5) e USP (5).

Estudos Epistemológicos e Críticos

Para verificar as temáticas estudadas nesse domínio, após as leituras e análises dos trabalhos, foram utilizadas cinco categorias, sendo estas base conceitual;

base histórica; base metodológica; diálogos interdisciplinares, e produção científica, conforme classificações descritas por Dodebei e Guimarães (2013: 17)³.

Nesse sentido, foi constatado que a categoria predominante é a de *base conceitual* com 23 trabalhos, de temáticas relacionadas ao Tratamento Temático da Informação; aos Sistemas de Organização do Conhecimento; Terminologia; Epistemologia da OC; Análise de Domínio; CI; e as relações entre CI e OC.

A categoria de *diálogos interdisciplinares* com 15 trabalhos apresenta predominância de temática sobre questões de Organização do Conhecimento na Arquivística, seguida de temas da Terminologia e Filosofia.

A *base metodológica* compreende 11 trabalhos e os temas referem-se a Ontologia e OC; Ensino em organização e representação do conhecimento Análise documental; Análise Documentária de charge; Autoria e Representação documental; Análise de assunto na representação de fotografias; Representação de fotografias na área de Publicidade; *Autopoiesis* na consolidação de linguagens de indexação; Linguagens de indexação e a análise de domínio; Linguagens documentárias e o tratamento dos recursos audiovisuais e Enriquecimento semântico por meio do Linked Data.

Referente à *produção científica*, há 9 trabalhos, e os temas são referentes a Análise Bibliométrica da revista *Scire* (2006-2010); Análise Bibliométrica sobre Metateoria; Análise Bibliométrica sobre indexação; Análise de Domínio nos anais da ISKO Brasil (2012 e 2013); Análise de Domínio na Produção Científica Brasileira de CI; Tendências temáticas e metodológicas da AKO; Elite e a Frente de Pesquisa nos anais da ISKO-Brasil (2011-2015); Produção científica dos pesquisadores PQ da CI membros da ISKO Brasil; e Perspectivas das pesquisas sobre Folksonomia.

Por último, na *base histórica*, foram identificados 8 trabalhos, e os temas versam sobre abordagem histórico-científica da CI; Representação do conhecimento Bibliografia de Gesner; Índice Poole e Indexação; Classificações de interesse do leitor; Ordenação de documentos nos manuais francófonos de Biblioteconomia; Fundamentos da catalogação descritiva e influências filosóficas nas classificações bibliográficas de artes.

3 “[...] I Congresso Brasileiro de Organização do Conhecimento, com o tema: Desafios e perspectivas científicas para a organização do conhecimento na atualidade e tendo por sub-temas: A dimensão epistemológica da organização do conhecimento (bases conceituais, históricas e metodológicas da organização do conhecimento assim como seus diálogos interdisciplinares e sua produção científica) [...]” Dodebei e Guimarães (2013: 17)

Estudos Terminológicos

Com a busca dos termos nos títulos, chegamos a 175 termos representativos da dimensão epistemológica. Posteriormente analisaram-se aqueles que são sinônimos ou variações terminológicas exatas, chegando a um total de 98, que foram distribuídos em oito categorias: 1) Configurações de Pesquisas em OC; 2) Teóricos da OC e de outros campos científicos; 3) Eventos; Revistas; Associações e Manuais; 4) Métodos de Pesquisas na CI e na OC; 5) Suportes da Informação Registrada; 6) Campos e disciplinas científicas; 7) Tratamento Temático da Informação e Sistemas de Organização do Conhecimento; 8) Temas que dialogam com a OC.

A categoria 7 de TTI e SOC, possui um total de 44 termos (Análise documental; Análise documental de obras de ficção; Análise documentária de charge; Assunto; Autoria; Catalogação de fotografias; Clasificación Decimal Dewey (CDD); Classificação; Classificação de documentos de arquivo; Classificação de interesse do leitor; Conceito; Controle de Estoque; Controle de Vocabulário; Descrição; Enriquecimento Semântico; Ensino da Análise de Assunto; Ensino em organização e representação do conhecimento; Especificidade; Exaustividade; Folksonomia; Indexação; Indexação Social; Índice; Informação Musical; Linguagens de indexação; Linguagens documentárias; Linguística; Mapeamento conceitual; Metadados de assunto; Número de chamada; Ontologia; Ordenação de documentos; Organização do conhecimento arquivístico; Organização do Conhecimento; Organização e Representação do Conhecimento; Representação Colaborativa da Informação; Representação da informação; Representação de fotografias; Representação do conhecimento; Representação do conhecimento arquivístico; Representação Documental; Sistemas de organização do conhecimento; Teorias Documentárias, e Usuário).

Por razões quantitativas, elegeu-se a categoria 7 para verificar o nível de compatibilização dos termos no Tesouro Brasileiro de Ciência da Informação (TBCI) e no DOCCIT (*Diccionario de organización del conocimiento: Clasificación, Indización, Terminología*) para verificar o nível de compatibilização (Figura 3).

O nível de compatibilização se apresenta mediano, uma vez que contempla com exatidão metade dos termos em ambos os instrumentos terminológicos. Um aspecto que se destaca é que o DOCCI possui maior compatibilização, aspecto que pode ser compreendido, pois é um instrumento que se dedica somente as questões de OC. Os termos que não aparecem no dicionário e no tesouro são: controle de estoque; informação musical; organização do conhecimento arquivístico; organização e representação do conhecimento; representação de fotografias; representação do conhecimento arquivístico; representação documental; teorias documentárias.



Figura 2. Descrição e quantidade de termos extraídos dos títulos dos trabalhos
Fonte: elaborado pelas autoras (2020)

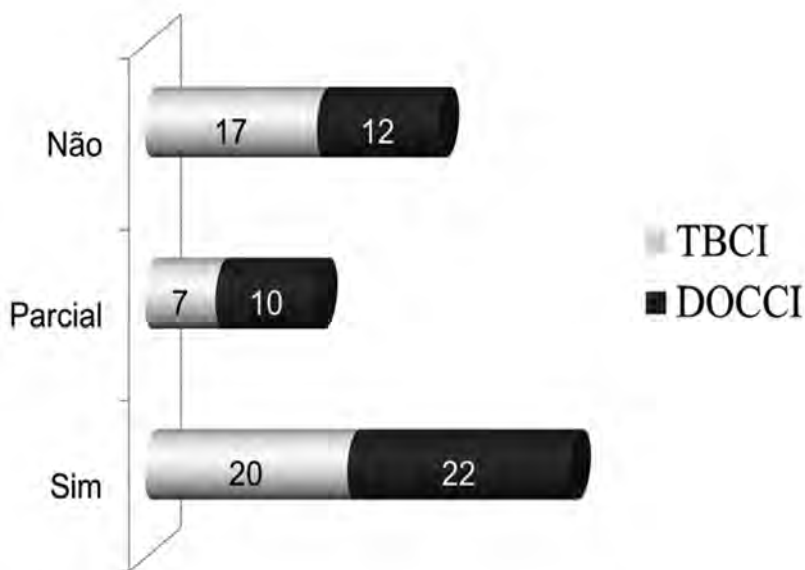


Figura 3. Compatibilização dos termos de TTI e SOC
Fonte: elaborado pelas autoras (2020)

CONSIDERAÇÕES FINAIS

Considerando a importância de uma comunidade científica internacional para o cultivo de reflexões e pesquisas, foi selecionada para ser mapeada a ISKO-Brasil.

Ressaltamos que essa comunidade discursiva ao ter um capítulo no Brasil contribui para a própria construção teórica e metodológica da organização do conhecimento brasileira a partir dos trabalhos apresentados, e posteriormente publicam e disponibilizam os anais de forma gratuita e aberta.

Dessa forma, este trabalho considerou os anais da ISKO Brasil, em específico da dimensão epistemológica – publicações estas que constituíram o *corpus* de coleta e de análise do trabalho. Fizemos uso da metodologia de Análise de Conteúdo e de 3 abordagens de AD, 1) Epistemológica - para identificar as temáticas das pesquisas; 2) Estrutura da comunicação científica - cujo foco está em descrever ano, autores e instituições; e por fim, 3) Estudos terminológicos.

Foi possível verificar que alguns domínios se destacaram recentemente na ISKO-Brasil, como a Arquivística na Organização do Conhecimento em 2019; além das temáticas de TTI, SOC, Terminologia e Filosofia, nos anos analisados. Entretanto, notamos uma carência de estudos de base histórica, se comparado com as demais categorias.

Agradecimentos

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Demanda Social).

REFERÊNCIAS

- Abbagnano, Nicola. 1998. *Dicionário de Filosofia*. 2a. ed. São Paulo, Brasil: Martins Fontes.
- Araújo, Paula Carina. 2019. “Epistemology of knowledge organization: a metatheoretical investigation”. Tese de Doutorado. Universidade Estadual Paulista Júlio de Mesquita Filho, Programa de Pós-graduação em Ciência da Informação.
- Araújo, Paula Carina, Joseph Tennis e José Augusto Chaves Guimarães. 2017. “The concept of epistemology in knowledge organization” em *Memória, tecnologia e cultura na organização do conhecimento*, coordenado por Fabio Assis Pinho e José Augusto Chaves Guimarães, 71-78. Recife: Ed. UFPE.
- Bardin, Laurence. 1977. *Análise de conteúdo*. Lisboa: Edições 70.
- Castro, Gustavo de. 2015. *O enigma Orídes*. Brasil: Página Itaú Cultural.
- Dodebei, Vera e José Augusto Chaves Guimarães. 2013. “Prefácio”, em *Complexidade e Organização do Conhecimento: desafios de nosso século*, coordenado por Vera Dodebei e José Augusto Chaves Guimarães, 13-15. Rio de Janeiro: Fundepe.

- Fontela, Orides. 1996. *Teia: poemas*. São Paulo: Geração Editorial.
- Francelin, Marivalde Moacir. 2005. “Abordagens em epistemologia: Bachelard, Morin e a epistemologia da complexidade”. *Transinformação* 17 (2): 101-109.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862005000200001&lng=en&nrm=iso
- Guimarães, José Augusto Chaves. 2015. “Organização do conhecimento: passado, presente e futuro em um contexto de diversidade cultural”, em *Organização do Conhecimento e Diversidade Cultural*, coordenado por José Augusto Chaves Guimarães e Vera Dodebei, 13-19. Marília: Fundepe.
- Hjørland, Birger. 2002. “Domain analysis in information science: eleven approaches: traditional as well as innovative”. *Journal of Documentation* 58 (4): 422-462.
- Hjørland, Birger. 2017. “Domain analysis”. *Knowledge Organization* 44 (6): 436-464.
- Hjørland, Birger e Hanne Albrechtsen. 1995. “Toward a new horizon in information science: domain-analysis”. *Journal of The American Society for information Science* 46 (6): 400-425.
- Hjørland, Birger e Jenna Hartel. 2003. “Afterword: ontological, epistemological and sociological dimensions of domains”. *Knowledge Organization* 30 (3/4): 239-245.
- Japiassú, Hilton e Danilo Marcondes. 2006. *Dicionário Básico de Filosofia*. 4a. ed. Rio de Janeiro: Jorge Zahar.
- Santos, Andreia dos, Bianca Ferreira Hernandez e Luciane Paula Vital. 2019. “Mapeamento da produção científica da Organização do Conhecimento em sua dimensão política e social no Brasil”, em *Organização do Conhecimento responsável: promovendo sociedades democráticas e inclusivas*, coordenado por Thiago Henrique Bragato Barros e Natalia Bolfarini Tognoli, 126-136. Belém: ED. da UFPA.
- Smiraglia, Richard. 2013. “The epistemological dimension of Knowledge Organization”, em *Complexidade e Organização do Conhecimento: desafios de nosso século*, coordenado por Vera Dodebei e José Augusto Chaves Guimarães, 17-25. Rio de Janeiro: Fundepe.
- Tennis, Joseph. 2012. “Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero?” *Brazilian Journal of Information Science: research trends* 6 (1): 3-15.

Para citar este texto:

- Alencar, Máira Fernandes e Brígida Maria Nogueira Cervantes. 2021. “Olhando a dimensão epistemológica da organização do conhecimento brasileira com três lentes da análise de domínio: epistemológica, estruturação científica e terminológica”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 35 (89): 195-211.
<http://dx.doi.org/10.22201/iibi.24488321xe.2021.89.58457>