

SCIRE

Representación y organización del conocimiento

SCIRE

Representación y organización del conocimiento

Vol. 23, n.º 2, julio-diciembre 2017

ISSN 1135-3716

Scire:

knowledge representation and organization
Vol. 23, n. 2, July-December 2017

Ibersid:

Red de Investigación
en Sistemas de Información
y Documentación

© 2017 Francisco Javier García Marco y los autores.

Reservados todos los derechos. Prohibida la reproducción por cualquier medio de la totalidad o parte de la presente edición con fines comerciales sin permiso escrito de los titulares del copyright. Queda, sin embargo, autorizada expresamente la reproducción de los resúmenes y palabras clave de los artículos en inglés y español. También se permite la reproducción de sus textos con objetivos exclusivamente docentes.

Los artículos publicados están sometidos a evaluación científica y editorial en beneficio de la comunidad científica y los autores, pero dicha evaluación se realiza dentro de los límites del desarrollo actual de la disciplina, las limitaciones humanas de editores y revisores, y con una perspectiva ideológica, teórica y metodológica abierta, por lo que la dirección no siempre comparte las opiniones, métodos y procedimientos reflejados en la revista. Los autores se responsabilizan personalmente del cumplimiento de los códigos éticos y la legislación vigente, así como de la obtención de los permisos de copia de materiales ajenos.

Se permite expresamente el autoarchivo de la versión electrónica por parte de los autores en sus repositorios personales e institucionales.

ISSN: 1135-3716 = Scire (Zaragoza)

Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas de la Universidad de Zaragoza

Foto de cubierta: Iuri Rocio Franco Rizzi

Imprime:

Servicio de Publicaciones. Universidad de Zaragoza.

Edificio de Ciencias Geológicas, C/ Pedro Cerbuna, 12.

50009 Zaragoza, España. Tel.: 976 761 330. Fax: 976 761 063.

Scire

representación y organización
del conocimiento

Alcance y objetivos

Scire: representación y Organización del Conocimiento es una publicación semestral de carácter interdisciplinar sobre la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Difusión

Scire tiene difusión internacional. Agradecemos la inclusión en los siguientes servicios de referencia: Scopus, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC y Catálogo Latindex.

Agradecimientos

Ibersid publica en este volumen ponencias y comunicaciones arbitradas presentadas en la edición de 2017 del congreso Ibersid, que contó con ayudas del Vicerrectorado de Investigación de la Universidad de Zaragoza.

Redacción, distribución y canje

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tfno: int + 34 976 762239. Fax: 34 976761506.
E-mail: <mailto:scire@unizar.es>

Suscripciones y números sueltos

Suscripción anual / Annual subscription: 30 €.
Número suelto: 20 €.
(IVA incluido)

Scire

knowledge organization
and representation

Scope and aims

Scire: Representación y Organización del Conocimiento is an interdisciplinary journal published twice a year that is devoted to the representation, standardization, treatment, retrieval and communication of information and knowledge.

Dissemination

Scire has international distribution. We acknowledge its inclusion in the following reference services: Scopus, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC and Catálogo Latindex.

Acknowledgments

In this volume, Ibersid publishes referred invited and contributed papers to the congress Ibersid 2017, which received grants from the Research Vice Rectorate of the University of Zaragoza.

Contact address

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tel.: int + 34 976 762239. Fax: 34 976761506.
E-mail: scire@unizar.es

Subscriptions

Annual subscription: 30 €.
Issue: 20 €.
(VAT included)

Editor

Francisco Javier García Marco, Univ. de Zaragoza. E-mail: jgarcia@unizar.es

Consejo de redacción / Editorial council

| | | |
|---|---|--|
| Mario Guido Barité Roqueta, Universidad de Montevideo, Uruguay | João Batista Ernesto Moraes, Universidade Estadual Paulista, Brasil | Francisco Javier Martínez Mendez, Universidad de Murcia, España |
| José Augusto Chaves Guimarães, Universidade Estadual Paulista, Brasil | Francisco Javier García Marco, Universidad de Zaragoza, España | Álvaro Quijano Solís, Colegio de México, México |

Consejo científico / Scientific council

| | | |
|---|---|---|
| Tomás Baiget, EPI S. A., España Mercedes Caridad Sebastián, Universidad Carlos III de Madrid, España | José María Izquierdo Arroyo, Universidad de Murcia, España María Pilar Lasala Calleja, Universidad de Zaragoza, España | Manuel José Pedraza Gracia, Universidad de Zaragoza, España María Pinto Molina, Universidad de Granada, España |
| Alberto Carreras Gargallo, Universidad de Zaragoza, España | Alfonso López Yepes, Universidad Complutense de Madrid, España | Gloria Ponjuán Dante, Universidad de La Habana, Cuba |
| Emilia Currás Puente, Universidad Autónoma de Madrid, España | José López Yepes, Universidad Complutense de Madrid, España | José Vicente Rodríguez Muñoz, Universidad de Murcia, España |
| Constança Espelt Busquets, Universidad de Barcelona, España | Pedro Marijuán Fernández, Universidad de Zaragoza, España | Adelaida Román Román, CINDOC (Madrid), España |
| María Eulalia Fuentes Pujol, Universidad Autónoma de Barcelona, España | Bonifacio Martín del Brío, Universidad de Zaragoza, España | Juan Ros García, Universidad de Murcia, España |
| Fernando Galindo Ayuda, Universidad de Zaragoza, España | José Antonio Moreiro González, Universidad Carlos III de Madrid, España | Francisco José Ruiz de Mendoza Ibáñez, Universidad de La Rioja, España |
| Blanca Gil Urdicián, Universidad Complutense de Madrid, España | Purificación Moscoso Castro, Universidad de Alcalá, España | Félix Sagredo Fernández, Universidad Complutense de Madrid, España |
| Alan Gilchrist, Cura Consortium, Reino Unido | Félix Moya Anegón, Universidad de Granada, España | Elías Sanz Casado, Universidad Carlos III de Madrid, España |
| Vicente Pablo Guerrero Bote, Univ. de Extremadura, España | María del Carmen Negrete Gutiérrez, Universidad Autónoma de México | Carlos Serrano Cinca, Universidad de Zaragoza, España |
| Víctor Herrero Solana, Univ. de Granada, España | José Luis Otal, Universidad Jaime I de Castellón, España | Emilio Setién Quesada, Asociación Cubana de Bibliotecarios, Cuba |

Revisores externos del número / External reviewers in this issue

Agradecemos enormemente la colaboración altruista y desinteresada de Mónica Izquierdo Alonso, María Luisa Alvíte Díez, Juan Carlos Fernández Molina, Javier Lacasta Miguel, Francisco Javier López Pellicer, José Antonio Martínez Comeche, Luis Miguel Moreno Fernández, Ely Tannuri de Oliveira, Juan Antonio Pastor Sánchez, Ana Belén Ríos

Candidaturas al consejo científico

Se aceptan candidaturas al consejo científico de especialistas del área de Biblioteconomía y Documentación y de otras disciplinas relacionadas (Informática, Ciencias Sociales, Lingüística, Filosofía, Psicología, etc.) con experiencia profesional e investigadora demostrada. En el sistema público de investigación español, suele ser equivalente al doctorado y dos sexenios de investigación o méritos equivalentes.

Scientific council membership policy

Candidatures of researchers from LIS and other related disciplines (Computer Science, Social Sciences, Linguistics, Philosophy, Psychology, etc.) with demonstrated professional and research experience are welcomed. In the Spanish public research system, for example, this usually means having a doctorate and two scientific productivity sexennia or equivalent outputs.

Tabla de contenidos en español

Table of contents in Spanish

| | | | |
|--|----|---|----|
| <i>Tabla de contenidos en español</i> | 9 | <i>Procedimiento de evaluación</i> | 75 |
| <i>Tabla de contenidos en inglés</i> | 11 | <i>Instrucciones para la presentación de trabajos</i> | 77 |
| Artículos | | <i>Índice de autores</i> | 80 |
| <i>La enorme mentira y la gran verdad de la información en tiempos de la postverdad</i> | | <i>Índice de materias en español</i> | 80 |
| <i>Juan Carlos MARCOS RECIO</i> <i>Juan Miguel SÁNCHEZ VIGIL</i> <i>María OLIVERA ZALDUA</i> | 13 | <i>Índice de materias en inglés</i> | 80 |
| <i>Organización automática de documentos mediante técnicas de análisis de redes</i> | | | |
| <i>Carlos G. FIGUEROLA</i> <i>José Luis ALONSO BERROCAL</i> <i>Ángel ZAZO RODRÍGUEZ</i> | 25 | | |
| <i>El análisis de dominio en la Ciencia de la Información: un análisis de la producción científica internacional</i> | | | |
| <i>José Augusto Chaves GUIMARÃES</i> <i>Daniel MARTÍNEZ-ÁVILA</i> <i>Amanda MARTINS</i> <i>Pedro Henrique Carvalho GOMES</i> | 37 | | |
| <i>La no univocidad del concepto de autor desde la perspectiva de Wittgenstein y Foucault</i> | | | |
| <i>Eduardo Graziosi SILVA</i> <i>Daniel MARTINEZ-AVILA</i> <i>Luciana de Souza GRACIOSO</i> | 45 | | |
| <i>Elementos de la teoría semiótica de la percepción de Peirce para el desarrollo de interfaces de recuperación de información</i> | | | |
| <i>Fabrizio BAPTISTA</i> <i>Carlos Cândido de ALMEIDA</i> <i>Edberto FERNEDA</i> | 53 | | |
| <i>Evaluación de las características de TemaTres y Multites para el control de autoridades en bibliotecas universitarias</i> | | | |
| <i>Mariângela Spotti Lopes FUJITA</i> <i>Luciana Beatriz Piovezan dos SANTOS</i> <i>Maria Carolina Andrade CRUZ</i> <i>Walter MOREIRA</i> | 63 | | |

Tabla de contenidos en inglés

Table of contents in English

| | | | |
|--|----|---------------------------------------|----|
| <i>Table of contents in Spanish</i> | 9 | <i>Evaluation process</i> | 75 |
| <i>Table of contents in English</i> | 11 | <i>Instructions for authors</i> | 77 |
| Artículos | | <i>Author index</i> | 80 |
| <i>The huge lie and the great truth of information in times of after truth</i> | | <i>Subject index in Spanish</i> | 80 |
| <i>Juan Carlos MARCOS RECIO</i> <i>Juan Miguel SÁNCHEZ VIGIL</i> <i>María OLIVERA ZALDUA</i> | 13 | <i>Subject index in English</i> | 80 |
| <i>Automatic organisation of documents using network analysis techniques</i> | | | |
| <i>Carlos G. FIGUEROLA</i> <i>José Luis ALONSO BERROCAL</i> <i>Ángel ZAZO RODRÍGUEZ</i> | 25 | | |
| <i>Domain Analysis in Information Science: an analysis of the international scientific production</i> | | | |
| <i>José Augusto Chaves GUIMARÃES</i> <i>Daniel MARTÍNEZ-ÁVILA</i> <i>Amanda MARTINS</i> <i>Pedro Henrique Carvalho GOMES</i> | 37 | | |
| <i>The non-univocity of the concept of “author” from the perspective of Wittgenstein and Foucault</i> | | | |
| <i>Eduardo Graziosi SILVA</i> <i>Daniel MARTINEZ-AVILA</i> <i>Luciana de Souza GRACIOSO</i> | 45 | | |
| <i>Elements of Peirce’s semiotic theory of perception for the development of information retrieval interfaces</i> | | | |
| <i>Fabricao BAPTISTA</i> <i>Carlos Cândido de ALMEIDA</i> <i>Edberto FERNEDA</i> | 53 | | |
| <i>An assessment of TemaTres and Multites characteristics for authority control in academic libraries</i> | | | |
| <i>Mariângela Spotti Lopes FUJITA</i> <i>Luciana Beatriz Piovezan dos SANTOS</i> <i>Maria Carolina Andrade CRUZ</i> <i>Walter MOREIRA</i> | 63 | | |

La enorme mentira y la gran verdad de la información en tiempos de la postverdad

The huge lie and the great truth of information in times of after truth

Juan Carlos MARCOS RECIO (1), Juan Miguel SÁNCHEZ VIGIL (2), María OLIVERA ZALDUA (3)

(1) Facultad de Ciencias de la Información de UCM, Av. Complutense s/n 28040, jmarcos@ucm

(2) Facultad de Ciencias de la Documentación de UCM, C/ Stma. Trinidad 37 28010, jmvigil@ucm.es (3) molivera@ucm.es

Resumen

Los tiempos presentes en los que la producción de contenidos es tan alta, dispersa y divergente, no permite a los ciudadanos reflexionar sobre lo que se escribe, quién lo hace o dónde se originó la idea principal. Acostumbrados al placer de disfrutar una información por más de 24 horas, que era el tiempo de valor de las noticias en el siglo XX, y pasar al siglo actual donde apenas dura unas horas, conlleva un ejercicio el determinar que toda la información que circula es verdad. Estamos en tiempos de la postverdad, donde las redes sociales mandan porque el consumo mayoritario se logra en ellas. Los medios han cedido un testigo que les convertía en referentes informativos y además generadores de opinión pública. El reto no es conseguir información, ahora el valor está en las fuentes, aunque la mayoría de los usuarios prefieren la rapidez que les ofrecen las redes sociales y en organizar el conocimiento en ellos, frente a la veracidad de los medios tradicionales.

Palabras clave: Postverdad. Medios de comunicación. Redes sociales. Control de información. Mentira. Facebook.

1. Introducción

Una industria de noticias fuerte también es crítica para construir una comunidad informada. Dar a la gente una voz no es suficiente sin tener personas dedicadas a descubrir nueva información y analizarla. Hay más que debemos hacer para apoyar a la industria de las noticias y asegurarnos de que esta función vital y social sea sostenible. (Mark Zuckerberg, 2017)

Los grandes políticos han sido quienes mejor han entendido el concepto de verdad. Incluso, por encima de los filósofos y de los comunicadores. Para ellos, la verdad, su verdad, la que hacen creer al resto de ciudadanos, es la que importa. Y dedican grandes esfuerzos a demostrarlo, con infinidad de teorías y dando las suficientes vueltas como para que sea creíble. Una mentira solitaria queda en eso. Una mentira contada miles, millones de veces se convierte en una verdad, creíble cuando la confianza en determinados medios es total. Y aquí ya entran en juego las redes sociales. Antes presentamos la situación anterior a su llegada.

Abstract

The present times, in which the production of content is so high, dispersed and divergent, does not allow citizens to reflect on what is written, who does or where the main idea originated. Accustomed to the pleasure of enjoying information for more than 24 hours, which was the value time of the news in the twentieth century, and moving to the current century where it lasts only a few hours, entails an exercise to determine that all the information that circulates is true. We are in the times of the future, where social networks rule because the majority consumption is achieved in them. The media have yielded a witness that made them informative referents and also generators of public opinion. The challenge is not to get information. Now the value is in the source, although the majority of the users prefer the speed that they offer the social networks against the veracity of the traditional means.

Keywords: Post truth. Media. Social networks. Information control. Lies. Facebook.

Nadie está en poder de la verdad absoluta. Y menos los medios de comunicación. La llegada de tecnologías de uso y control masivas está permitiendo importantes cambios en la manera de comunicar que tienen las personas. Deberíamos preguntarnos y aproximarnos en este texto a esta cuestión: ¿Qué vale más un consejo de alguien próximo a la persona o miles y miles de comentarios de gente que no conoces de nada pero que te apoyan con un “me gusta”? La respuesta parece estar más cerca de la segunda versión. El narcisismo es un camino galopante que recorre la sociedad actual. Los más jóvenes, muchos de ellos aislados en su propio mundo, en una sociedad que tiene otros intereses, descubren con gran satisfacción que hay otras realidades fuera, que además atienden sus peticiones, esas que en sus círculos próximos no toman en serio. Se sienten identificados en los mundos paralelos de las redes sociales, donde buscan la reafirmación de su yo y la constatación de que otros viven también esas experiencias. En algunos casos, esas propuestas terminan en una visión trágica de la

vida. Para identificarse requieren de extremismos que graban como testimonio de su propia necesidad, antes de tomar una drástica decisión.

Hasta la llegada de las redes sociales fueron los medios de comunicación el gran espejo de la sociedad. Aparecer o no en ellos era símbolo de estatus. Lo que hacían los medios, la forma en que construían la información y los métodos empleados para hacer una sociedad más democrática, justa y participativa era una prerrogativa de los periódicos impresos, primero y, luego, de la televisión. El siglo XX, sobre todo después de la segunda Guerra Mundial, se edificó en base a la sociedad que descubre la televisión. Un mundo irreal que invita a los ciudadanos a que lo tomen como real. Es una mentira rodeada de verdad que crea mundos paralelos y que los ciudadanos están dispuestos a validar. Y el cine vino a refrendar esas teorías, añadiendo una gran parte de glamour que llevaba a los espectadores, por un par de horas y unos cuantos dólares al paraíso, del que luego eran expulsados a su propia realidad.

Sin embargo, está en la naturaleza del ser humano ser cambiantes y oponerse a situaciones que no son tan auténticas como se ven reflejadas en los medios de comunicación. Y apostaron entonces por creer que esa tecnología que es tan buena para la ciencia, los hospitales, el ocio y el entretenimiento, no lo era tanto para los medios de comunicación que se vieron desplazados por las redes sociales. Éstas vinieron a ser el espejo en el que se miran desde entonces las generaciones más jóvenes y, también porque no decirlo, los que antaño soñaron con verse dentro del paraíso de la TV y del cine.

Pero nadie calculó las consecuencias. Solo importaba el número. Cuantos más mejor. Todo era gratis, con la condición de que se añadiera lo que importaba para que otros hicieran lo mismo y se construyera una gran colmena. Eso son las redes sociales, una colmena, manejada por una reina a la que todos alimentan con su esfuerzo. Te ofrecen el paraíso para que vivas en él, siempre y cuando aportes fotos, textos, etc., propios y o extraños que puedes compartir.

La democracia, más o menos controlada desde la televisión en las campañas políticas, cedió el testigo a esas redes; incluso, se especuló que la victoria del candidato Trump se debió a la información privilegiada en el manejo de esas redes. La economía se volvió más centralizada que nunca y su poder creó tentáculos tan poderosos que ya no se sabe bien quién mueve los hilos de las empresas dentro de los países, ya que muchas de ellas se han deslocalizado, aprovechando otros salarios en lugares donde la sociedad permite que el pago sea escaso. La vida en

toda su dimensión se adaptó a lo que las grandes corporaciones (casi todas de Estados Unidos) nos invitaban a probar. Así, el mundo pasó de un Plan Marshall a un Plan Zuckerberg. El producto era diferente porque las necesidades también lo eran.

También este planteamiento de buscar la verdad es necesario para organizar el conocimiento. No se puede dar como válida la información y al tiempo decir lo contrario. Los errores que se producen en los medios de comunicación tienen una especial significación; aunque más tarde se obligue por ley a la rectificación, no todos los usuarios de ese medio se dan cuenta. En este sentido es clave la función de la hemeroteca (tradicional o digital) y en consecuencia la técnica para organizar los documentos conservados con el fin de facilitar el acceso, es decir de globalizar.

En este orden del mundo se ha elaborado un calendario de mentiras y verdades que se manejan al antojo de los políticos y las corporaciones que los respaldan. No es nada nuevo. Se trata de seguir creando mundos para que el ser humano se identifique con su propia felicidad y viva inmerso en ella. El reto es para los medios de comunicación. Ellos deben destapar las verdades ocultas y las mentiras ciertas. Cuando de verdad aportan valor y construyen un mundo de igualdad es cuando están respondiendo a este enunciado: la enorme mentira y la gran verdad deben resolverse para ofrecer al ciudadano lo que el mundo está construyendo, lo que sucede a escasos kilómetros, pero también lo que les afecta cuando a lo lejos una empresa quiebra y no tarda en pasar lo mismo en su ciudad.

2. Las redes mandan, la información ya no tanto

Más de un siglo de dominio de los medios de comunicación escritos sirvieron para ir construyendo sociedades democráticas, que sin embargo no impidieron dos guerras mundiales y otras tantas en países que nunca habían tenido un conflicto. Los periódicos, que también ayudaban a crear guerras y luego luchaban por detenerlas, estaban por encima de la verdad y de la mentira. Y así siguen actuando en la actualidad, aunque menos. Buscaban siempre la cara amable de la sociedad, el referente de una razón fundamentada en la verdad. Mientras que los periódicos impresos dominaron el mundo, con millones de ejemplares saliendo de las rotativas, la sociedad lectora servía para justificar los condicionantes de la información. Situémonos en 1898, cuando William Random Hearst, uno de los grandes magnates de la prensa de entonces, que había enviado a uno de sus dibujantes a Cuba y

éste quería regresar a Estados Unidos porque no encontraba nada interesante ni noticioso que dibujar, recibió un telegrama de su jefe: "Tú haz los dibujos, que yo pondré la guerra". Y así fue. Poco tiempo después, el *New York Journal* anunciaba un nuevo conflicto entre España y Estados Unidos con un claro vencedor desde el comienzo.

Si se analizan los principales periódicos del siglo XX, a lo largo de las confrontaciones que hubo, se descubre con mediana claridad que las enormes mentiras y las grandes verdades conviven con cierta naturalidad. Lo que se viene a denominar propaganda de guerra, que incluye también en el siglo XX a la radio y a los pasquines que se lanzaban desde los aviones, se fue transformando en otra manera de contar lo que sucedía en esos conflictos. Y al finalizar los mismos se puede valorar cuánto de verdad hubo de información y cuanto de manipulación o propaganda a favor de la causa.

Mientras los periódicos impresos perdían fuerza, la televisión se convirtió en la "reina" de las noticias que el mundo deseaba recibir en conflictos más recientes, como la Guerra de Irán e Irak, o la caída de Sadam Hussein años después. Y fue entonces cuando la cadena CNN tuvo el privilegio de informar, una vez que todos los medios fueron obligados a salir del país. El control absoluto del régimen político solo permitía obtener una única versión, por muy independientes que fueran sus periodistas. Algo no muy diferente es lo que ha pasado estos días cuando este mismo medio fue expulsado de Venezuela.

Este tiempo de la televisión está pasando a mejor vida. La mayoría de los ciudadanos que siguen la información en este medio no tienen como prioridad estar informados, sino profundizar en alguna información que la televisión ha preparado y pone en su sitio web como contenido informativo. Ya no es un referente de actualidad. ¿Dónde se informan los ciudadanos? ¿Cuál es el medio predominante? ¿Qué tipo de informaciones se ofrecen? ¿Qué hay de verdad y que no? ¿Existe un manejo de fuentes que nos permitan creer en la verdad de la información? ¿Qué tipo de documentos se generan? ¿Cómo se organizan los nuevos documentos?

No hay respuestas milagro para este nuevo entorno. Estamos en tiempos de la postverdad. Pensemos en que antes una información tenía al menos una vigencia de 24 horas. Tiempo suficiente como para contrastar bien lo que se iba a publicar. Buscar las causas y consecuencias de la información, reflexionar sobre los hechos y madurar a la hora de escribir las mejores ideas. En la actualidad, esa misma información se viene

cambiando, mejorando y actualizando varias veces al día. En ese caso, con más razón. Si los periodistas disponen de herramientas que les permiten cambiar, actualizar y ofrecer nuevos contenidos de una misma información, ¿cómo es que en el mundo cada vez hay más noticias falsas? El reciente triunfo de Trump en la presidencia de los Estados Unidos es un ejemplo claro de lo que la verdad esconde y de lo que la mentira afirma.

El mundo es de las redes sociales. Una gran crisis de identidad viene afectando a los periódicos y al resto de medios desde hace dos décadas. Hasta entonces, la mayoría de las noticias procedían de un pequeño núcleo de corporaciones de medios que eran capaces de llegar a un gran público. Se trataba de unas cuantas agencias de noticias a nivel global y varios grupos de comunicación con poder para obtener informaciones en todo el mundo y construir su propia verdad. Lo hacían sin competencia, o al menos con un Estado en el que todos tenían acceso a la publicidad que les permitía, además de las subvenciones, seguir adelante. Pero también tuvieron que asumir la crítica de la sociedad o de algunos sectores y/o personas influyentes. Así Chomsky y Herman fungieron con su espíritu crítico: "la materia prima de la noticia debe pasar a través de filtros sucesivos, dejando sólo el residuo limpio y apropiado para imprimir". El resultado, como señala Tanz, fue un falso consenso nacional, que ignoró hechos, voces e ideas periféricas. (Tanz, 2017).

Hasta entonces, y este entorno es la parte fundamental de lo que está sucediendo en la actualidad, los medios estaban al servicio, más o menos, de los dirigentes políticos que los subvencionaban o apoyaban con informaciones que hacían crecer sus audiencias/lectores. Los grandes debates de cierre de elecciones, aplicados primero en Estados Unidos, y luego replicados en otros países occidentales, llegaron a su fin (aunque se sigan haciendo) en la primera década del siglo XXI. A modo de resumen, las redes sociales se fueron apoderando del entorno político de la siguiente manera: 2008, "Elección de YouTube" y "Facebook", ya que las dos plataformas fueron muy utilizadas por el personal de Barack Obama para transmitir sus mensajes; 2012 pasó a la historia como la elección Twitter, gracias a los tweets como los legendarios "Cuatro años más"; y 2016, ascenso de todas las plataformas sociales, llega Instagram, aplicaciones en vivo (Hillary Clinton y Ted Cruz) y Snapchat.

Lo que algunos medios de comunicación no han entendido es que esa cadena se ha roto. No son tan importantes como antaño. Los políticos ya no necesitan a los medios, ya no dependen de los

periodistas para llegar a sus audiencias. Ahora pueden hablar directamente con ellos o comunicar desde las webs informativas de las instituciones públicas y privadas (Salvador Benítez, Sánchez Vigil 2016). En estos momentos, la comunicación es directa con los votantes, sin el tamiz de la televisión. Y lo más importante: reciben una respuesta directa, la estudian y saben qué tienen que ir cambiando en su estrategia comunicativa. Al mismo tiempo perciben qué otras personas están alejadas de su perfil votante y pueden construir de manera directa para ellos, para convenirlos desde una misma posición.

Mientras que los periódicos y la televisión eran un filtro difícil de superar para los votantes, ahora, las redes sociales mandan porque son interactivas y porque el seguidor de esas redes se siente en constante comunicación con el candidato. Y esto que se aplica en los procesos electorales también se utiliza en el mundo de la empresa, el entretenimiento, el comercio, etc. No es arriesgado decir que esta revolución de las redes sociales donde el poder está abajo, frente al poder de la pirámide que controlaban los grupos económicos que movían los hilos políticos, ha perfilado una nueva sociedad que aún no se puede catalogar porque los resultados están por verse. En todo caso, y en relación a los medios de comunicación, perdieron el poder y sólo sobrevivirán si saben adaptarse a los tiempos y utilizan las redes sociales en beneficio propio. Los lectores de papel se reducen considerablemente. Un periodista publica una historia y aunque la lean miles de personas, nunca llegará a crear estado de opinión porque serán una minoría frente a la gran masa que ha leído en las redes sociales otro texto y ese será el que dé valor a la sociedad cuando se mueva por las redes. Si un periodista no es capaz de compartir su texto y de que las redes sociales como un espejo lo ofrezcan, aquél no pasará a la historia. Así pues, las redes mandan, la información no tanto.

Cuestión fundamental en el cambio es la modificación de los documentos, los nuevos modelos que circulan por las redes, generalmente multimedia, lo que influye en la organización del conocimiento, y en el diseño de herramientas y técnicas para el tratamiento documental; es decir, que debe prestarse mayor atención al tema debido a la profusión de contenidos.

3. La influencia del periodismo no es la que era

El constante trasvase de lectores hacia las redes sociales, de aquellos que abandonan la televisión por YouTube, de los que no consumen revistas

porque han encontrado en Snapchat y/o Instagram los recursos que necesitan para ser “glamurosos” o de los que escuchan radios nativas en digital frente a las grandes cadenas, dibujan un periodismo diferente en su concepción. El camino se ha recorrido dando pasos de gigante por parte de las redes sociales y muy tímidos por el resto de los medios. El empuje, lejos de venir de periódicos con gran poder, enraizados en la sociedad; de televisiones con audiencias millonarias o de radios con miles de seguidores, lo dieron las propias redes. En muchos casos ha sido un proceso autoproclamado al ver que los medios fueron dejando un campo abierto al consumo de noticias.

Casi sin darse cuenta y pero conscientes de la realidad, algunas redes sociales se encontraron con millones de personas que aceptaban contenidos informativos que los propios usuarios ofrecían como suyos. Mientras el periodismo digital jugaba al periodismo ciudadano, las redes sociales afrontaban el reto de aceptar lo que sus usuarios querían mostrar a otras personas. Entonces, el periodismo enrabietado comenzó una guerra contra ellas. No es periodismo si no tiene suficientes datos, no lo es cuando no está contrastado, cuando no tiene declaraciones de los protagonistas, etc. Pero para los usuarios de las redes era información. ¿Qué separa/une a la información y el periodismo? El periodismo sin información no es periodismo. Los medios tenían los datos, las declaraciones, la profesionalidad del periodista, pero no encontraban lectores que consumieran esa información. Las redes carecían de todo lo anterior, pero ofrecían informaciones que para la mayoría de los usuarios eran suficientes.

En este compromiso se vieron los lectores de Google News. Estaban suficientemente informados. Cuando de verdad querían ampliar la información, a través de un enlace llegaban a los propios medios para completar la lectura. En este caso, ambos medios se complementaban. Google News arrastraba su audiencia hacia el medio y éste obtenía más lectores y consecuentemente más publicidad.

Facebook es algo diferente, si bien nunca se ha posicionado como un medio de comunicación, si que han valorado que algunas propuestas de sus usuarios se podrían comparar con contenidos informativos. Pero, lejos de enfrentarse a los grandes medios de comunicación, su CEO y fundador pasó a desmentir esa propuesta cuando algunos medios informaron que estaban haciendo información: “Cuando piensas en una compañía de medios, ya sabes, la gente está produciendo contenido, la gente está editando contenido, y eso no es lo que nosotros hacemos” (Zuckerberg, 2017).

Más recientemente aclaró esta idea en un memorándum que publicó en Facebook; o más bien, abrió nuevas brechas al señalar que era difícil para los medios sobrevivir en un mundo que domina Facebook (LaFrance, 2017):

Ahora, Zuckerberg está dejando claro que quiere que Facebook tome el control de muchas de las funciones reales —no sólo los dólares de publicidad— que las organizaciones de noticias tradicionales tuvieron una vez. Zuckerberg está construyendo una organización noticiosa sin periodistas.

Algunos periodistas han visto en este manifiesto una manera de destruir el periodismo; al menos, el que ha venido apoyando a la democracia, a la construcción de sociedades más justas y a una mejor convivencia entre las personas.

Lejos de tranquilizarse, los medios entendieron que era una manera de acercar posturas, pero nada se clarificaba porque mientras Facebook sigue creciendo por millones de usuarios, los lectores y las redacciones están descendiendo a pasos agigantados. Los medios de comunicación están en un problema. Es un problema propio en el que las redes sociales no tienen la culpa, como explica Tanz (2017):

El modelo de negocio impulsado por la publicidad está al borde del colapso. La confianza en la prensa está en uno de los puntos bajos de todos los tiempos. Y ahora esas dos preocupaciones largas se han sumado a una crisis existencial aún mayor. En una era post-hecho de noticias falsas y burbujas de filtro, en las cuales el público selecciona la información y las fuentes que coinciden con sus propios sesgos y descarta el resto, los medios de comunicación parecen haber perdido su poder de moldear la opinión pública.

Y lo que realmente quedaba a los medios del siglo XX era la manera en que podían crear y cambiar la opinión pública. Entre los medios y los usuarios había una excelente comunicación. Para los usuarios, porque respaldaban su propio pensamiento cuando se apoyaban en lo que habían leído en el periódico o visto en la televisión. Con frecuencia se oían expresiones como lo ha dicho tal periódico o lo he visto en los informativos de la televisión. Esa comunión funcionó por el interés y compromiso de ambas partes.

En el siglo XXI los consumidores de noticias no tienen deudas ni compromisos con los grandes medios. Entraron a convivir con las noticias digitales y muy pocos han tenido contacto diario con los medios cada día. Fue una puesta en común total. No necesitaban leer secuencialmente el periódico para estar informados. Lo estaban porque tenían abiertas sus principales redes sociales de las que obtenían las noticias. En ocasiones, cuando la noticia lo requería y después de pedir opinión a otros amigos o personas que estaban

en su mismo círculo, pasaban a completar esa información en algún medio tradicional. Pero esa es una minoría insuficiente para mantener una tasa de usuarios alta que interese a las marcas que pueden poner publicidad en el medio. Y hay que resolver otra cuestión que ya empieza a preocupar a los editores de prensa: generaciones que no han convivido nunca con los medios impresos y que ya están en el mercado (Tyner, 2017):

Para las personas como nosotros, que no crecimos con los periódicos y que hemos utilizado la Internet desde que éramos niños, muchas de las convenciones de las noticias de hoy no tienen sentido. Las noticias en 2017 no tienen que seguir el ciclo de producción de noticias en 1987, y tampoco deben venir en la misma forma. Adaptar un periódico para la web moderna no es lo suficientemente bueno para las personas que nunca leen los periódicos en primer lugar. Nos merecemos algo nuevo.

El cambio hacia un mundo más completo es un reto que no han resuelto los medios, especialmente los periódicos. Se ha intentado de varias maneras. Abrir el contenido del periódico, abrir una parte del mismo; combinar gratis con otra parte de pago; apostar por excelentes contenidos que atraigan lectores que hasta entonces solo buscaban la información en las redes sociales, crear muros de pago y zonas gratuitas, etc. En estos últimos veinte años hay una cuestión prioritaria: un buen producto informativo solo se consigue cuando un periodista logra hacer información de calidad. Otra cosa es que ese tipo de información termine por interesar a unos pocos, muy pocos. En ese caso, ese periódico, radio y/o televisión han de apostar por productos muy bien elaborados para un público selecto.

Esta información “gourmet” tiene su recorrido en determinados nichos de mercado que no llegan a cubrir los grandes medios. En ningún caso se puede comparar con las informaciones globales que tienen un gran recorrido en las redes sociales. Por ejemplo, la noticia de la muerte de un cantante de éxito o la situación de una “celebrity” da mucho juego como contenido noticioso en las redes sociales; pero quizás no tanto en un medio digital. A cambio, un gran descubrimiento científico sería muy valorado por los medios tradicionales que ampliarían la información preguntando a otros científicos y tomando información de prestigiosas revistas de esa área. El resultado sería de mucho valor para los seguidores de ese medio, pero muy pocos serían los que pasarían a comentarlo con sus amigos y conocidos en las redes.

No todo sucede exactamente así. Este es un ejemplo. El periodismo tradicional, con fuentes de información fiables, con métodos para contrastar la información y con rigurosidad a la hora de informar ha quedado al margen de lo que quieren los

nuevos lectores. Frente a los lectores pasivos, con muy poca participación en los medios escritos del siglo XX, como mucho a través de las cartas al director o de algún artículo de opinión, se ha pasado a un lector crítico (Marcos et al., 2008) y actuante que comenta aquellas noticias que a su juicio tienen un valor para él y sus conocidos. Mientras, el resto de noticias, las que de verdad cuesta crear, esas no suelen despertar un interés en los miembros de una red social. No se trata de discriminar, ni de comparar. No son los buenos, los intelectuales, los comprometidos, etc., los que leen los periódicos tradicionales y ven programas de calidad en la televisión; ni a la contra, los que solo asumen la información a través de las redes son los mal informados, los que se dejan influir, los que creen que la verdad es esa y solo esa.

El cambio no ha resuelto esta dicotomía. Nunca antes (y eso que también se cometían errores) se han producido tantas noticias falsas (algunos artistas han muerto varias veces) ni se han juzgado de forma paralela a protagonistas de la información como cuando su historia aparece en alguna red social. La vida cambia y con ella el periodismo. Y aunque haya perdido influencia en la sociedad, gracias a él se pueden entender muchos cambios y transformaciones de la sociedad actual. Solo hay que estar atento a las noticias con rigor que traen los medios tradicionales en formato impreso y/o digital.

Este cambio al que todavía la sociedad se está acostumbrando muestra aristas muy peligrosas como el cierre de determinados medios que fueron significativos en algunas ciudades, y con ellos la documentación recopilada durante años. Cada vez que un medio desaparece, la sociedad es más frágil y el mundo más peligroso. Así lo ha constatado en un estudio el profesor Lee Shaker, asistente de comunicaciones en la Universidad Estatal de Portland: "En el pasado, las muertes de las organizaciones noticiosas han puesto en peligro la perspectiva de una comunidad segura, bien informada y comprometida". Un documento de 2014 encontró una caída sustancial en el compromiso cívico tanto en Seattle como en Denver entre 2008 y 2009, después de que ambas ciudades vieron el cierre de diarios de larga duración. En Seattle, el *Post-Intelligencer* detuvo su edición impresa, pero continuó produciendo noticias en línea; en Denver, *Rocky Mountain News* se cerró (LaFrance 2017). Por tanto, los medios que nacieron para ser un contrapeso al poder, si desaparecen y ese poder pasa a estar en manos de corporaciones que no reconocen el valor de la información, excepto en el tema de las noticias falsas, entonces el mundo tendrá otra deriva.

4. La verdad no se crea ni se destruye... aparece cuando nadie la espera

El punto de partida podría ser esta propuesta del escritor y periodista George Orwell. En 1946 publicó el ensayo: *Politics and the English Language* en el que advierte a los lectores (Orwell, 1946):

El lenguaje político -y con variaciones esto es cierto para todos los partidos políticos, desde los conservadores hasta los anarquistas- está diseñado para hacer que las mentiras sean veraces y asesinas respetables, y para dar una apariencia de solidez al viento puro.

Setenta años después, unas elecciones a la presidencia de Estados Unidos en 2016 han venido a reorientar la relación de los políticos con los medios. Desde su atalaya, nada más tomar el poder y coincidiendo con manifestaciones en varias ciudades del mundo contra su política, su respuesta fue que eran los medios los culpables de haber orquestado esa situación. Los primeros días de su mandato los pasó, precisamente, atacando a la prensa y el resto de medios para marcar una línea diferente, rompiendo protocolos establecidos en la sala de prensa de la Casa Blanca y permitiendo que fueran otros medios los que tomaran el protagonismo en las ruedas de prensa.

Marcado el territorio, la propuesta fue reunirse con algunos de los grandes medios para explicarles el nuevo funcionamiento; pero sus propuestas no calaron en los periodistas. La llegada de este presidente ha sido más rupturista que la de otros. Y así lo han atestiguado algunos de los grandes periodistas de ese país. Por ejemplo, Margaret Sullivan comentó que la forma tradicional de informar sobre un presidente había muerto; al menos, como se venía haciendo hasta ahora. Para calmar las aguas, el Secretario de Prensa de la Casa Blanca, Sean Spicer comentó a un nutrido grupo de periodistas que "nuestra intención nunca es mentirle". Pero el periodismo no debe ser partidista. Debe buscar la verdad con honestidad y transparencia, descubrir hechos noticiosos y alternativos y presentarlos como verdaderos. Eso es lo normal, lo que cada día debe plantearse un periodista.

Pero hay un iceberg que apenas está mostrando una parte de lo que oculta. El férreo control que algunas corporaciones ejercían sobre los medios, mediante la compra de acciones, ahora es más claro porque se conoce la identidad de quien controla ese grupo de accionistas. Por tanto, determinados contenidos están marcados por la propia línea editorial, por ejemplo, el pensamiento de Jeff Bezos en *The Washington Post* o de la familia Sulzberger en *The New York Times*. El torrente de información actual ayuda a crear

climas de dudosa reputación y es entonces cuando aparecen las noticias falsas; las noticias montadas con un propósito o las noticias que deliberadamente apoyan una tendencia política, económica y/o religiosa.

En estos tiempos en los que falta una reflexión crítica de las cosas, incluida por supuesto la lectura pausada de los contenidos que ofrecen los periódicos, el auge de mentiras en la sociedad de la postverdad ha venido a alertar a los ciudadanos de que no sirve cualquier información que encuentren en las redes sociales; ni siquiera en los grandes periódicos donde se han “colado” determinados contenidos que tenían un variado interés.

¿Por qué la sociedad de la postverdad? ¿Por qué la verdad no siempre aparece como antes se la esperaba? ¿Quién maneja los hilos de la información? ¿Cómo detectar lo que es verdad de lo que no lo es en las informaciones? El punto de partida es claro: todos mentimos alguna vez. Se empieza de niños, en algún juego, y con ese sabor dulce del triunfo se aplica más tarde en la vida diaria. ¿Quién no ha mentido alguna vez? La cifra debe ser muy pequeña porque en la actualidad la mentira se ha empoderado y actúa con determinada libertad porque tiene un público que la acepta con demasiada facilidad. Lo cierto es que hay tres ejemplos claros que han constatado que la verdad ya no se presenta como antes: la votación del Brexit, en Inglaterra; el referéndum de Colombia para el proceso de Paz y las elecciones 2016 de Estados Unidos. En los tres, lo que parecía ser verdad resultó que no lo era.

Si los medios de comunicación son un reflejo de lo que hace y piensa la sociedad, las mentiras tienen su espacio. Según una información del diario británico *Daily Mail*, citado por Alpañés (2016), las mujeres mienten una media de tres veces por jornada mientras que los hombres lo hacen seis. El problema es que con las redes sociales todo se disemina más. Son un gran altavoz que llega a millones de personas. Y cada vez que un mensaje falso se distribuye a esos niveles, muchos empiezan a considerarlo verdadero.

Si tomamos como ejemplo, nuevamente las elecciones de Estados Unidos, descubrimos que los mensajes del actual presidente iban cargados de mentiras que los ciudadanos asumieron como si fueran verdades. Un seguimiento de sus discursos realizados por la agencia independiente PoliFact estudió 150 aseveraciones de Donald Trump: “El resultado: el 78% de sus afirmaciones eran falsas o inexistentes. En nuestro país –España– el partido del gobierno está procesado por destrucción de pruebas, es decir, por ocultar la verdad” (Alpañés, 2016, p. 23-24). Así pues, si

los gobiernos mienten, los ciudadanos toman como referente ese modelo y consideran que mentir es algo normal, puesto que además de los gobiernos y los políticos, los periódicos no dicen toda la verdad, la sociedad la oculta y hasta los niños cuando pueden, disfrutan contando alguna mentira. Alguna solución se debería proponer. De momento, un ex presidente de los Estados Unidos, George W. Bush ya ha elevado su voz cuando señala que los medios de comunicación son “indispensables para la democracia” y pide su apoyo en estos tiempos diferentes en su país (Nelson, 2017):

Considero que los medios de comunicación son indispensables para la democracia. Necesitamos los medios de comunicación para que la gente como yo pueda rendir cuentas. El poder puede ser muy adictivo y puede ser corrosivo, y es importante que los medios de comunicación llamen a orden a las personas que abusan de su poder, ya sea aquí o en otro lugar.

5. ¿Quién gana con las noticias falsas? El valor de las fuentes/documentación

Hay grandes mentiras en los gobiernos actuales. Algunas se tratan de justificar en beneficio de la sociedad, o pensando que la mejor manera de atajar futuros problemas es poniendo remedio antes. El ejemplo de las guerras, citado con anterioridad, avala esta idea. La pregunta quizás no esté bien formulada, aunque políticamente siempre hay un vencedor. Lo correcto sería: ¿Por qué las sociedades actuales se dejan influenciar tanto no ya por los medios, sino por las redes sociales? ¿Por qué una simple información puede hacer tanto daño a una sociedad cuando se extiende como la pólvora en las redes sociales? También hay espacio para construir una sociedad más solidaria cuando los ciudadanos utilizan las redes para por ejemplo compartir datos sobre una enfermedad rara y eso permite salvar la vida de otras personas. No se pueden demonizar las redes sociales, pero en el tema de la información que difunden, unas veces creadas a propósito y otras sin querer, lo que se dice no es la verdad; y una vez establecida es muy difícil de dar marcha atrás.

El poder de los medios se ve debilitado, pero aún tiene un margen para mejorar. Por supuesto no se puede comparar con el crecimiento que experimentan las redes sociales y la velocidad con la que lo hacen, aunque ya hay quien viene anunciando que algunas, como Facebook, han tocado su techo. Desglosemos algunas cifras para situar este entorno (Lafrance, 2017):

El diario japonés *Yomiuri Shimbun* afirma que su circulación de 9 millones de copias al día la convierte

en la más grande del mundo. En los Estados Unidos, la audiencia promedio diaria en tiempo real de *CNN*, *Fox News* y *MSNBC* fue de 3.1 millones de personas en 2015, según el análisis de Pew Research Center de los datos de Nielsen. Y el *New York Times* tenía alrededor de 1,6 millones de suscriptores digitales desde el otoño pasado.

Y estas cifras tienen que competir con los 1.900 millones de usuarios activos que conviven por varias veces al día en Facebook.

Sin embargo, frente a esta avalancha de noticias falsas, tanto Facebook como Google se han implicado para frenarla. De alguna manera, ambas pueden forzar a quien trata de introducir contenidos falsos, bien frenando su avance o bien alejándolos de los posibles beneficios a través de la publicidad. No parecen ser ellos los malos de la película. Esta nueva situación debe conducir a repensar lo que los medios de comunicación quieren. Deben saber en qué lado ponerse. Y las grandes empresas tecnológicas no tienen claro en qué nicho se ubicarían, como se pregunta Bell (2017):

¿Por qué las empresas de tecnología o sus fundadores millonarios quieren hacerlo? Ellos, más que cualquier otro interesado, tienen un interés en la integridad de la información, y la sostenibilidad del periodismo como un proyecto independiente. Ni Facebook ni Google realmente quieren emplear a periodistas, pero las instituciones de periodismo seguirán fallando debido a las condiciones adversas del mercado. Zuckerberg expresa su preocupación por el 'periodismo local' y la Iniciativa de Noticias Digitales de Google ha hecho lo mismo.

Este es un proceso con muchos altibajos. Hay quienes, incluso, proponen que cada millonario tecnológico compre alguna de las cabeceras que están pasando por malos momentos y la relancen hacia lo digital; más o menos, la experiencia de Jeff Bezos con *The Washington Post*. Pero no es lo mismo gestionar un periódico donde se abren muchos frentes, que en una empresa tecnológica donde los retos van en otras direcciones.

Aún está por resolver quién gana o a quién le interesan las noticias falsas. El proceso pasaría por combatir constantemente la desinformación o ver el grado de dependencia que tienen los medios de Facebook y viceversa. En ese proceso andan autores como Evangeline (2017), Filloux (2016) o Silverman (2016) que sostienen que en los últimos meses de la campaña 2016 a las elecciones en Estados Unidos, algunas noticias falsas de Facebook recibieron una mayor participación que 19 grandes medios de comunicación de ese país. Por ejemplo, Silverman (2016) apunta:

Durante los meses críticos de la campaña, 20 historias de elecciones falsas de alto rendimiento provenientes de sitios de engaño y blogs hiperpartidistas

generaron 8.711.000 acciones, reacciones y comentarios en Facebook

Por su parte, Filloux considera que esta red social hasta podría prescindir de las noticias, algo que nunca han dicho sus directivos: "Debemos afrontar el hecho de que Facebook no se preocupa por las noticias en el sentido del periodismo. Las noticias representan alrededor del 10% de las fuentes de noticias del usuario promedio y las noticias pueden ser cortadas de la noche a la mañana si las circunstancias lo dictan sin ningún impacto significativo para la plataforma (Filloux, 2016). Pero no hay que olvidar que el negocio de Facebook pasa porque los usuarios permanezcan cuanto más tiempo mejor dentro de la red social. Pero, a la larga, al propio Facebook no le interesa tener que desmentir tantos errores y por eso junto con Google lanzó una campaña para frenar las noticias falsas. Hasta el propio Zuckerberg (2017) se encargó de presentar un Manifiesto que defiende otras formas de comunicar, entre ellas, frenar el avance de noticias falsas.

Hay una división, que no es tanto un enfrentamiento radical, sino una manera de plantear la información en diferentes niveles. En el fondo, todos quieren lo mismo: tener lectores, disfrutar de usuarios activos, disponer de personas en una red social. El final es el mismo cuando los medios impresos se dieron cuenta de que si conseguían más lectores, la publicidad aumentaría. Las redes sociales luchan no sólo por tener más usuarios sino porque permanezcan más tiempo. Sin publicidad, quizás estos movimientos de informaciones falsas para lograr que los usuarios estén un mayor tiempo en la red, no llegarían a producirse.

Por otra parte, hay que preguntarse si en las redes sociales hay un procedimiento que ayude a resolver estas cuestiones: ¿cuál es el proceso que se sigue para verificar una información? ¿cómo se ejecuta? ¿quién es el responsable? ¿se hace un seguimiento? No hay que olvidar que estamos en los tiempos del "todo vale". Se toma un texto de algún lugar de internet, se adapta, se mejora, o se empeora y se vuelve a publicar. Tantas veces y con tantas versiones, que el usuario ya no acierta a saber si es verdad o no. ¿Dónde están los filtros necesarios? Es posible que no existan porque la grandeza de Internet, en general y de las redes sociales, en particular, es dejar entera libertad (salvo en contadas excepciones hay censura y suelen ser sobre imágenes, no tanto sobre textos). La libertad nadie duda de que es buena, pero alguna manera de dosificarla es necesaria para evitar que se sigan creando noticias falsas y que tengan intención de beneficiar a determinadas personas y/o partidos políticos o empresas.

Nadie duda que en los periódicos, en la radio y en la televisión se han “colado” informaciones falsas. En ocasiones, intencionadamente. Pero no ha sucedido al nivel de los últimos meses de 2016, coincidiendo con las elecciones en Estados Unidos. En parte, por el propio interés de las redes sociales para dar una versión informativa de lo que sucedía y estar mejor posicionados que el resto de medios tradicionales; y, en parte, por sumar usuarios y tiempo de permanencia en esas redes. Por el contrario, la seriedad de la información, el rigor periodístico, la noticia con datos, el análisis de los entornos, etc., no tuvieron grandes lectores, ni seguidores en radio y televisión. Es una muestra más de que las nuevas generaciones, algunos ya ni conocen las versiones impresas de los medios, quieren informaciones breves, dinámicas, actualizadas y sin muchos datos. Es un tipo de periodismo que no se puede llamar como tal. Al menos, como el producto que colaboró en la creación de sociedades democráticas y que ayuda a la libertad de pensamiento y de acción entre las naciones.

La pregunta clave sería ¿están interesadas las redes sociales y el resto de medios en publicar contenidos noticiosos que no sean verdaderos? ¿Cómo lo harían? ¿Qué consecuencias tendría? Ya refiere Bell, líneas atrás, que ni Google ni Facebook contratan periodistas. Por tanto, la respuesta es clara: hay un nuevo tipo de información que interesa a las nuevas generaciones de lectores/participantes en las redes sociales, y que no están dispuestos a recibir más ni mejores contenidos. Tampoco les interesa que esa información se ajuste a la realidad, ni que sea verdadera o no. La velocidad está por encima del valor real de la información. Existen propuestas por parte de los medios de comunicación para frenar el avance de las noticias falsas. Una de ellas —creada y gestionada por *Le Monde: Le Décodeurs* (Imagen 1)— pretende gestionar y estrangular la difusión de noticias falsas en línea (Daves, 2017):

‘Decodex’ es el nombre de los tres productos de comprobación de hechos del periódico francés alimentados por una base de datos de 600 sitios web considerados poco confiables y compilados por *Le Décodeurs*, la unidad de comprobación de hechos de *Le Monde*, durante el último año.

Otra más reciente reunió en las jornadas Misinfo-Con a un grupo de expertos con el fin de “llegar a la raíz del problema de las noticias falsas” y acabar con “La propagación de la desinformación, así como el contenido de noticias deliberadas y fabricadas en línea, ya que tienen muchas cabezas, y no existe una sola arma para defenderse de ellas”. Para ello proponen arreglar el quebranto del periodismo. Además, crearon una

serie de propuestas para aplicar de manera inmediata en los medios: “listas de verificación, campañas educativas, soluciones tecnológicas, compromiso con los proyectos de la comunidad, esfuerzos de diversidad y mejora de los modelos de negocio. Después de todo, las ‘noticias falsas’ han evolucionado para encontrar muchas cosas, incluyendo las historias que a uno no le gustan” (Wang, 2017).



Imagen 1. Portal de acceso a Le Décodeurs

Se trata de una generación que fomenta otros valores. Su templo de saber ya no es la biblioteca sino Google, sus libros no están en la librería del barrio, sino en Amazon; su música, sus fotografías, sus juegos, su propia identidad es virtual y con ella firman parte de su legado. No esperan que una información tenga o no buenas fuentes o que el redactor haya utilizado un centro de documentación, o, más lejos aún, que haya consultado varias bases de datos. Lo que demandan las nuevas generaciones es una información que les permita estar al día en los contenidos que ellos previamente han seleccionado. Ni siquiera conciben un medio como una totalidad. Antes se leía uno el periódico, primero las secciones que más interés tenían y luego el resto.

Hoy los medios permiten suscripciones a una serie de contenidos, por lo que estamos generando una sociedad relativamente informada en una materia y con escasas perspectivas del resto. Mientras no se dé valor a las bibliotecas, la lectura, los buenos programas de radio y televisión, el mundo tomará una deriva limitada: la que interesa a las redes sociales para seguir atrayendo a sus usuarios. Mientras ellos estén cómodos y el negocio funcione, para qué cambiarlo. Esta es una sociedad democrática, liberal, participativa... Mejor en palabras del propio Zuckerberg (2017):

La exactitud de la información es muy importante. Sabemos que hay información falsa e incluso contenidos engañosos en Facebook, y tomamos esto muy en serio. Hemos presentado batallas contra el spam, pero tenemos más trabajo que hacer. Estamos procediendo con cuidado porque no siempre hay una línea clara entre los engaños, la sátira y la

opinión. En una sociedad libre, es importante que la gente tenga el poder de compartir su opinión, aunque otros piensen que están equivocados. Nuestro enfoque se centrará menos en la prohibición de la desinformación, y más en la aparición de perspectivas adicionales, incluyendo la disputa por la precisión de un artículo.

Queda, mientras tanto, pedir que los medios tradicionales, en formatos impresos o digitales, no se rindan y sigan ofreciendo información contrastada, lo más próxima a la verdad, con apoyos documentales, con aportaciones de especialistas, con análisis en profundidad y con el rigor que siempre ha caracterizado a estos medios. La verdad no tiene dueño, pero en estos tiempos de la postverdad se agradece más que nunca a quienes están comprometidos con ella.

El planteamiento de la información requiere de una organización general y específica. Se trata de estructurar la documentación para apoyar a los medios en su análisis y gestión. Cada paso que da el redactor, pero sobre todo el documentalista, será para tener una visión mayor en su organización y en consecuencia en su gestión. No se puede olvidar este proceso; en caso contrario, los resultados del consumo de noticias no serían productivos para las empresas.

Debemos plantear si la organización y gestión del conocimiento en la comunicación cambia sus parámetros en tiempos de la postverdad. Es este un texto para la reflexión y el análisis, y aquí una cuestión interesante a añadir sería si deberían construirse archivos de la mentira

6. A modo de conclusiones

La información siempre ha estado unida a tiempos de convulsión. Una redacción tranquila es sinónimo de que no pasa nada en el mundo. Pero las cosas suceden y hoy es fácil de contarlas, mucho más que en el siglo XX. Tenemos, sí, más información, pero no significa que estemos mejor informados. De hecho, una gran mentira recorre y recoge las propuestas que el actual presidente de los Estados Unidos ofrece a los medios. Dejó claro el primer día qué pensaba en los medios y cómo iba a actuar. Pasó a atacar directamente a los medios y a hacerles culpables de muchas situaciones que él mismo había generado. Ya hay en su propio partido quien está en contra de esa forma de hacer política y pide a los medios que refuercen la democracia contando la verdad.

Otra gran potencia, Rusia, viene utilizando parecidos métodos. ¿Qué cambió para llegar a esta situación? No se puede culpar a las redes sociales de que hayan tomado un camino en el que sus informaciones son suficientes para muchos de sus usuarios, y eso que para algunos, incluido

el propio Zuckerberg, no es periodismo, ni Facebook es un medio ha servido para propagar con mayor rapidez informaciones falsas. Los políticos saben que las redes les facilitan una tarea que antes les costaba mucho tiempo. Pueden tomar un tema y hacerlo viral, aunque lo que se cuente no sea cierto. El caso es llegar al público masivo. ¿Qué dicen las redes sociales? Mientras tengamos usuarios que estén interesados en esos temas, adelante. Pero esta situación tampoco es buena para las redes, de ahí que también estén persiguiendo contenidos que no se ajustan a la realidad. Algunos van más lejos, como *Le Monde*, que ha creado un sitio para determinar qué otros sitios de Internet “trafican” con informaciones falsas. La mejor solución para frenar las noticias falsas es la que están proponiendo algunos editores de prensa a través de los *fact-checkers* o verificadores de hechos. Para ellos, los verificadores desempeñan un papel importante y necesario a la hora de comprobar la información. Ese es el reto. Difícil porque los recortes en los medios han dejado que se pierdan algunas funciones, como la del revisor de los textos.

La sociedad de la postverdad ya no es una responsabilidad de los medios, como en el siglo XX. Ahora muchas mentiras se mueven por Internet con el peligro de que otros las copien, las difundan y las hagan virales. La mejor manera de estar informados de verdad siguen siendo los medios tradicionales de toda la vida y los nuevos medios digitales que se han ganado el derecho a contar la verdad. Las redes sociales son casi más rápidas que el periodismo digital, pero no tan seguras. Y esto parece que no va a cambiar de inmediato.

Sin un control por parte de los documentalistas y gestores de la información será difícil que el conocimiento que se genera en estas redes sociales tenga un valor de futuro. La inmediatez de los hechos y las informaciones que publican estas redes es su principal motivación. El consumo de noticias decrece a pasos agigantados en los medios y se hace más activo en las redes sociales, por este motivo el conocimiento que antes ofrecían los medios terminará por perderse si la tendencia de consumo de información continua dentro de las redes sociales.

Referencias

- Alpañés, E. (2016). Bienvenido a la sociedad postverdadera. // *Yorokobu*. 79, 22-24.
- Bell, E. (2017). How Mark Zuckerberg could really fix journalism. 21/02/2017. http://www.cjr.org/tow_center/mark-zuckerberg-facebook-fix-journalism.php?utm_source=Daily+Lab+email+list&utm_campaign=81f5b61e0d-dailylabel3&utm_medium=email&utm_term=0_d68264fd5e-81f5b61e0d-395887489 (2017-04-10).

- Davies, J. (2017). Le Monde identifies 600 unreliable websites in fake-news crackdown. 25/01/2017. <http://digiday.com/uk/le-monde-identifies-600-unreliable-websites-fake-news-crackdown/> (2017-04-10).
- Evangeline (2017). Hacking the Facebook news problem. 16/02/2017. <https://medium.com/global-editors-network/hacking-the-facebook-news-problem-85eb31f823a1#t1dbk1g01> (2017-04-10).
- Filloux, F. (2016). Facebook's Walled Wonderland Is Inherently Incompatible With News. 05/12/2016. <https://mondaynote.com/facebooks-walled-wonderland-is-inherently-incompatible-with-news-media-b145e2d0078c#.ro8n035mq> (2017-04-10).
- Herman, E. S.; Chomsky, N. (1998). *Manufacturing consent the political economy of the mass media*. London: Arrow Books.
- Lafrance, A. (2017). The Mark Zuckerberg Manifesto Is a Blueprint for Destroying Journalism. 17/02/2017. <http://www.theatlantic.com/technology/archive/2017/02/the-mark-zuckerberg-manifesto-is-a-blueprint-for-destroying-journalism/517113/> (2017-04-10).
- Marcos Recio, J.C.; García-Alonso, P.; Parra, D.; Edo, C.; Flores Vivar, J. (2008). *Estrategias y perspectivas documentales en la información digital*. Sevilla: Ámbitos, 2008. <http://www.redalyc.org/articulo.oa?id=16812702001> (2017-04-10).
- Nelson, L. (2017). Bush breaks with Trump, calls media 'indispensable to democracy'. 27/02/2017. http://www.politico.com/story/2017/02/george-w-bush-trump-media-235430?utm_source=Daily+Lab+email+list&utm_campaign=386d129d80-dailylabemail3&utm_medium=email&utm_term=0_d68264fd5e-386d129d80-395887489 (2017-04-10).
- ORWELL, G. (1946). *Politics and the English Language*. London: Horizon, april 1946.
- Salvador Benítez, A; Sánchez Vigil, JM (2016). La fotografía como documento informativo en las webs de los Ministerio Españoles: Acceso, visibilidad, funcionalidad y políticas de uso. // *Revista Española de Documentación Científica*, 39:2.
- Silverman, G. (2016). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. 16/11/2016. http://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.xe4W2DGeb#.ng36eAQI2 (2017-04-10).
- Tanz, J. Journalism Fights for Survival in the Post-Truth Era. 14/02/2017. <http://www.wired.com/2017/02/journalism-fights-survival-post-truth-era/> (2017-04-10).
- Tyner, J. (2017). Dear news media: Create news for people who have never read a newspaper. 21/02/2017. <https://medium.com/media-center-lab/dear-news-media-create-news-for-people-who-have-never-read-a-newspaper-8f94845f9008#.mpds2s0ov> (2017-04-10).
- Wang, S. (2017). Shan Wang. Getting to the root of the "fake news" problem means fixing what's broken about journalism itself. 27/02/2017. http://www.niemanlab.org/2017/02/getting-to-the-root-of-the-fake-news-problem-means-fixing-whats-broken-about-journalism-itself/?utm_source=Daily+Lab+email+list&utm_campaign=386d129d80-dailylabemail3&utm_medium=email&utm_term=0_d68264fd5e-386d129d80-395887489 (2017-04-10).
- Zuckerberg, M. (2017). Building Global Community. 16/02/2017. <http://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634> (2017-04-10).

Enviado: 2017-04-10. Segunda versión: 2017-06-21.

Aceptado: 2017-09-14.

Organización automática de documentos mediante técnicas de análisis de redes

Automatic organisation of documents using network analysis techniques

Carlos G. FIGUEROLA, José Luis ALONSO BERROCAL, Ángel ZAZO RODRÍGUEZ

Universidad de Salamanca, Instituto de Estudios en Ciencia y Tecnología, c/ Espejo, sn 37008 Salamanca (España),
figue@usal.es, berrocal@usal.es, zazo@usal.es

Resumen

La organización automática de documentos permite conocer la estructura temática de grandes colecciones documentales. En este trabajo se plantea modelar la colección de documentos mediante una red o grafo de nodo y enlaces y aplicar técnicas de Análisis de Redes Sociales. Se describe un experimento llevado a cabo con una colección de noticias de prensa, analizando la estructura temática resultante tras aplicar técnicas de detección de comunidades de nodos en dos niveles de agrupamiento. Los resultados parecen bastante prometedores, planteando como trabajo futuro la implementación y comparación de diferentes algoritmos de detección de comunidades.

Palabras clave: Clasificación automática. Redes sociales. Clustering.

1. Introducción

Uno de los retos de la llamada Sociedad del Conocimiento es la organización de documentos, de forma que sea posible encontrar una estructura lógica que permita un acceso más eficaz a la información contenida en ellos. En el territorio de la archivística clásica, el Principio de Procedencia (Martín-Pozuelo, 1996) ha permitido solventar esta cuestión, buscando una organización documental basada en la estructura interna de las organizaciones productoras de documentos.

La descripción a partir de campos, más o menos explícitos, y el uso de lenguajes documentales (controlados o no) se han aplicado durante décadas en el terreno de bibliotecas y centros de documentación; están orientados, no obstante, a la aplicación con documentos con soporte físico.

El desarrollo de las Tecnologías de la Información y, especialmente, Internet ha desestabilizado por completo lo que conocíamos como producción de documentos y, por ende, todo el proceso documental.

El problema ahora no es sólo la cantidad de documentos —también— sino otra serie de factores añadidos como la volatilidad, la falta de tipología, las autorías de nuevo tipo (colaborativas, y

Abstract

Automatic organization of documents can show the semantic structure of broad collections of documents. This paper proposes to model a document collection using a graph or network and then applying the so-called Social Networks Analysis techniques. We describe a practical experiment carried out with a collection of newspaper articles, and then we analyze the topic structure resulting after applying community discovery techniques. Results look enough promising; we envisage as future work the application and comparison of different communities discovery algorithms.

Keywords: Automatic classification. Clustering. Social Network Analysis.

otras), nuevas estructuras documentales, y un largo etcétera.

Una vía de abordar esta situación es la automatización de procesos; sin embargo, ésta está lejos de poder enfrentarse de forma eficaz a estos problemas. Una de las cuestiones que tiene planteadas es la organización automática de los documentos en función de su contenido temático. Compatible con otros modelos de organización y acceso, la organización automática puede favorecer la recuperación, al colocar documentos de contenidos similares juntos; o el análisis cuantitativo de las colecciones documentales; o el desarrollo y perfeccionamiento de aplicaciones de extracción de información basados en *machine learning*, al poder ofrecer grandes cantidades de contextos similares para el entrenamiento de dichas aplicaciones.

En este trabajo se describe un sistema de clasificación automática de documentos basado en técnicas de Análisis de Redes Sociales; se describe también su aplicación en un caso concreto y se analizan los resultados. Es importante remarcar que se trata de una aplicación real en un caso real; no se trata, por consiguiente, de un experimento de laboratorio en el cual el entorno y los parámetros de aplicación están controlados. Antes bien, al tratarse de un caso específico en una situación

real, el contexto viene dado por esa situación real; esto implica, por ejemplo, que las características de los documentos con los que se ha trabajado escapa a nuestra elección; o que la evaluación de los resultados carece de puntos de referencia con los cuales establecer comparaciones.

Este trabajo está organizado de la siguiente manera: en la sección siguiente se introduce la clasificación automática de documentos y algunos de los sistemas clásicos más utilizados; se introduce también el Análisis de Redes Sociales. En la siguiente sección se detalla la metodología aplicada, la colección de documentos utilizada, el procesamiento a que ha sido sometida, así como los algoritmos aplicados. A continuación, se describen los resultados obtenidos y se evalúan, discutiendo las métricas obtenidas. Finalmente, se ofrecen unas conclusiones.

2. La clasificación automática de documentos

Es frecuente distinguir entre sistemas de clasificación supervisada (a veces denominada también categorización) y clasificación no supervisada o clustering (Campos Ibáñez y Romero López, 2011; Ares Brea et al., 2011). En el primero de los casos los documentos se clasifican en una estructura creada *ad-hoc* por personas; la mayor parte de los sistemas aplican técnicas de *machine learning* y consisten, de una forma u otra, en construir patrones o modelos de las diferentes categorías; y medir después el parecido o similitud de cada documento con cada uno de esos patrones (Baharudin y otros, 2010).

Existen diferentes sistemas de clasificación supervisada, como los probabilísticos (Langley y otros, 1992; McCallum y Nigam, 1998), el llamado *Nearest Neighbour* (Yang, 1999) o los SVM (Joachims, 2002). Estos sistemas se vienen empleando con éxito en diferentes aplicaciones; esto incluye, obviamente, el campo documental, en el que se han aplicado con el fin de clasificar documentos o texto (Eyheramendy, Lewis y Madigan, 2003; Kim et al., 2006; Joachims, T., 1998). Por nuestra parte, hemos podido documentar tasas de éxito superiores al 94 % (Figuerola, 2013; Quintanilla, Figuerola y Groves, 2015; Figuerola et al. 2017).

La clasificación no supervisada o *clustering* carece de estructuras clasificatorias pre-definidas y es el mismo sistema el que, en función de las características de la colección documental, las crea. Se trata, pues, de un escenario más cercano a la idea de organización automática o autoorganización de documentos.

Diversos métodos pueden utilizarse para clasificar temáticamente, de manera automática, colecciones amplias de documentos; una revisión de los más importantes puede encontrarse en Aggarwal y Zhai (2012). Así, dentro de lo que se conoce como clasificación no supervisada, es relativamente frecuente el uso de algoritmos de clustering como *k-means* (Jain, 2010). La versión clásica de éste tiene el inconveniente de producir clusters planos, de un solo nivel, además de necesitar fijar de antemano el número de clusters deseados, lo cual no siempre es fácil y requiere, con frecuencia, de un proceso de prueba y error.

Diferentes sistemas de modelado de temas (*topic modeling*) (Hidayat et al., 2015; Griffiths and Steyvers, 2004) se utilizan también para descubrir de forma automática los temas tratados en una colección de documentos (Figuerola, García Marco y Pinto, 2017). Obviamente, delimitados los temas, es posible agrupar o clasificar los documentos en función de esos temas. El procedimiento más conocido es el denominado *Latent Dirichlet Allocation (LDA)* (Blei, Ng and Jordan, 2003), que goza de cierto prestigio en el terreno de las *Digital Humanities* (Shawn and Milligan, 2012). Aunque existen implementaciones fáciles de utilizar (de ahí, tal vez, su relativa popularidad), tiene el inconveniente de que es preciso elegir de antemano el número de temas o topics deseados, lo cual no siempre es sencillo (Arun et al., 2010). Además, el etiquetado o identificación de los temas detectados suele requerir una fase de análisis por parte de expertos, no exenta de subjetividad e inconsistencias. Este problema del etiquetado es, en realidad, común a todos los sistemas de clustering; el sistema agrupa documentos afines pero es responsabilidad de los utilizadores etiquetar cada uno de tales agrupamientos con una expresión realmente significativa sobre el contenido de ese agrupamiento.

En este trabajo se plantea abordar la clasificación automática de documentos mediante la aplicación de técnicas de *Análisis de Redes* (Otte and Rousseau, 2002), efectuando una aplicación sobre una colección de documentos y analizando sus resultados. La aplicación de estas técnicas no requiere prefijar el número de *clusters* deseados, puede producir estructuras jerárquicas y, como se verá, los resultados de su aplicación parecen bastante prometedores.

3. Metodología

El Análisis de Redes tiene su origen en la teoría matemática de Redes o Grafos; una de sus aplicaciones en Ciencias Sociales más tempranas es la efectuada en los años 60 del siglo pasado, por

sociólogos intentado modelar las relaciones entre personas y grupos sociales (Scott, 2013). Brevemente, una red es un conjunto de nodos o vértices conectados por arcos o enlaces. Los nodos pueden tener una serie de características o atributos arbitrarios, definidos por quien aplica este artefacto. Los enlaces o arcos conectan dos nodos entre sí; los enlaces pueden tener dirección (parten de un nodo y apuntan a otro) o no (simplemente conectan dos nodos en una relación ambivalente); pueden existir también arcos reflexivos (parten y llegan al mismo nodo). Los arcos o enlaces también pueden tener atributos arbitrarios a gusto del usuario, pero uno de los más habituales es el peso: un valor numérico que intenta expresar la fortaleza de la relación que representa ese arco.

Lo interesante de las Redes es que se han desarrollado métodos y procedimientos para analizar la estructura interna de una red. De manera que, si conseguimos modelar un determinado fenómeno mediante una red, podemos utilizar esas técnicas de análisis para estudiar la estructura interna de ese fenómeno. Una de las cosas que es posible hacer con una red es descubrir o detectar las posibles comunidades de nodos que haya en ella (Plantíe and Crampes, 2013).

Es posible modelar una colección de documentos como una red, en la cual los documentos pueden ser representados por nodos. Dos documentos o nodos pueden estar conectados entre sí por un arco o enlace si ambos tienen un contenido parecido o similar; y el peso de ese arco podría ser al grado de similitud entre esos dos documentos.

Desde finales de los años 60 del pasado siglo disponemos de formas para medir la similitud entre dos documentos, gracias a la formulación del modelo vectorial de recuperación de la información por G. Salton (Salton, 1983). Naturalmente, ha habido desde entonces nuevos modelos y nuevas propuestas y podemos aplicar cualquiera de ellas; pero las más utilizadas hoy día siguen estando basadas en el modelo vectorial de Salton, aún cuando no haya estado exento de polémicas (Leydesdorff, 2008). Básicamente, la similitud entre dos documentos se mide en función de la cantidad de palabras que éstos tienen en común; sucede que no todas las palabras tienen la misma importancia o peso y, en consecuencia, no cuentan lo mismo. Y sucede también que los documentos tienen tamaños diferentes (en número de palabras) y esto requiere aplicar algún sistema de normalización.

Así que, de un modo u otro, es posible calcular la similitud entre cada pareja de documentos de nuestra colección, trazando una red de nodos y arcos.

A esta red es posible aplicar alguno de los varios sistemas disponibles de detección de comunidades de nodos. Una comunidad de nodos es un conjunto de éstos que enlazan fuertemente entre sí, y débilmente con los no pertenecientes a esa comunidad. Dado que los enlaces de nuestra red están basados en las similitudes de contenido entre los nodos, una comunidad debería agrupar a documentos de temática similar (Lee and Cunningham, 2014; Pons and Latapy, 2005).

4. La colección de documentos

Hemos aplicado estas ideas a una colección de noticias de prensa extraídas de algunos periódicos. Esta colección está formada por 50.000 noticias, todas sobre Ciencia y Tecnología y es conocida como Spanish Corpus of Scientific Culture (SCSC) (Figuerola, Quintanilla et al., 2017). Se trata de noticias de tres diarios españoles de ámbito nacional en su versión digital, publicadas entre 2002 y 2011, a texto completo. A la totalidad de noticias extraídas de la hemeroteca digital de cada uno de los periódicos se aplicó un sistema de categorización para filtrar u obtener todas las noticias relacionadas con la Ciencia y/o la Tecnología.

El sistema de categorización automática aplicado, basado en SVM (*Support Vector Machine*), consiguió, en este caso, una precisión bastante elevada (94.5 % de aciertos), dejando la colección con una cantidad muy pequeña de ruido (Groves, Figuerola y Quintanilla, 2015). Las noticias se convirtieron a texto plano, desde su formato web original.

Pese a ser noticias sobre Ciencia y/o Tecnología, es obvio que, dentro de ese amplio campo, es posible preguntarse por su estructura temática: qué campos o disciplinas científicas están presentes, y en qué medida; aquí es donde entra la aplicación de las técnicas de Análisis de Redes.

Se calculó la similitud entre cada par de noticias o documentos aplicando un sistema clásico de *tf-idf* para estimar los pesos de cada palabra y el bien conocido coeficiente del coseno (como el que aplican muchos sistemas de recuperación de la información) para medir la similitud. El resultado es una semi-matriz cuadrada de 50.730 por 50.730 elementos. Esta matriz es bastante densa, puesto que casi todos los documentos tienen alguna palabra en común, aunque sean palabras muy poco significativas, de peso muy bajo.

Esta matriz de similitudes nos permite construir una red de 50.730 nodos (cada documento). Y si enlazamos cada documento con los que mantiene alguna similitud, por baja que ésta sea, obtenemos una cantidad de enlaces cercano a 1.200 millones; una cantidad intratable debido a

su tamaño. Se impone podar enlaces aplicando un umbral de similitud; tras varias pruebas, se aplicó un umbral de 0,09, eliminando todos los enlaces inferiores a este valor. El resultado fue una red de 50.730 nodos y 23.612.149 enlaces.

Una representación visual de esa red puede observarse en la figura 1.

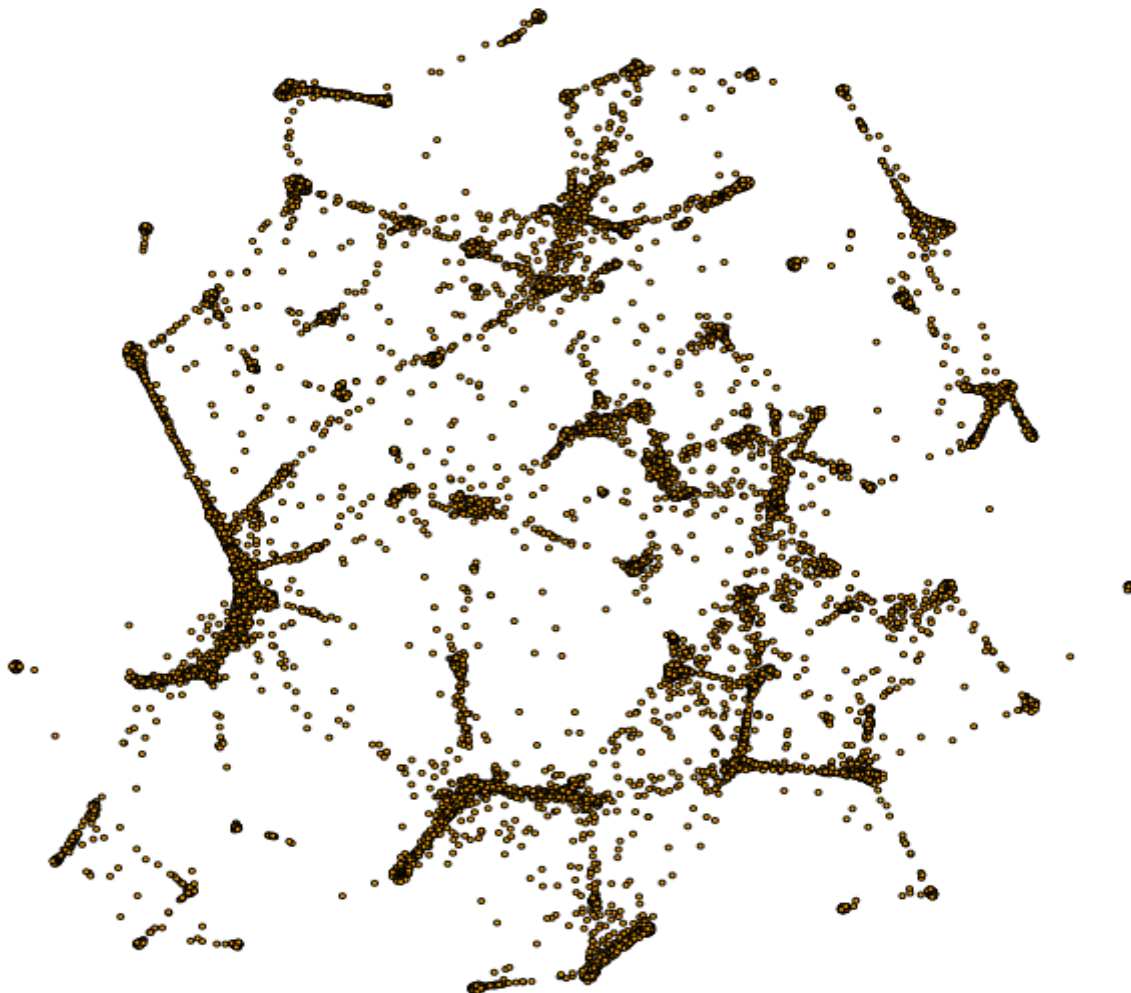


Figura 1. Red de documentos

Esta representación está realizada mediante un algoritmo *force directed* conocido como *Open Ord*, (Martin et al., 2011), que intenta colocar los nodos más fuertemente enlazados más próximos entre sí. Sin entrar en las dificultades de conjugar un número tan alto de nodos y enlaces, claramente podemos observar como hay conjuntos de nodos (documentos) más o menos agrupados. Estos grupos es lo que conocemos como comunidades de nodos de una red y, dado que la relación entre los nodos, en nuestro caso, está basada en la similitud semántica; podemos pensar que esas comunidades agrupan documentos que tratan sobre los mismos o parecidos temas.

Podemos aplicar, pues, alguno de los algoritmos de detección de comunidades para obtener la estructura temática de nuestra red de documentos. Sin embargo, uno de los problemas de la detección de comunidades es que suele requerir un tiempo de procesamiento considerable (Lancichinetti and Fortunato, 2009); tras algunas pruebas, aplicamos el algoritmo de detección de comunidades conocido como *Infomap* (Rosvall, Axelsson and Bergstrom, 2009; Bohlin et al., 2014; Edler and Rosvall, 2015).

Una de sus características es su eficacia con redes grandes.

5. Resultados

Como resultado obtenemos 23 comunidades, muchas de ellas con un segundo nivel de subcomunidades cada una. Sin embargo, de ellas solamente 13 tienen un tamaño, en número de documentos, apreciable, abarcando el 90 % de toda la colección. Un análisis a simple vista de las noticias componentes de cada comunidad significativa muestra la consistencia temática de las mismas; a modo de ejemplo, véanse la Figura 2 y las Tablas I y II.

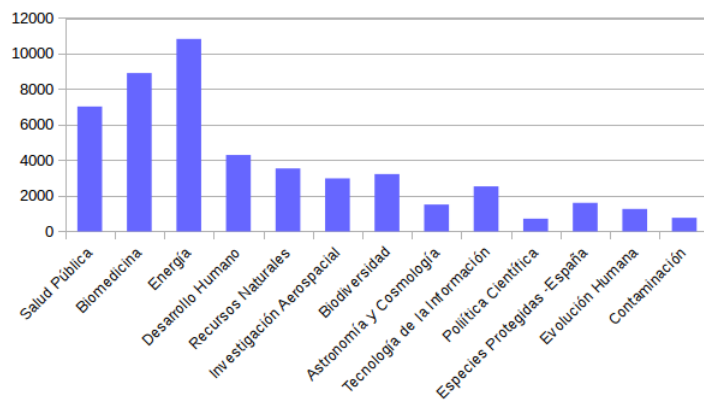


Figura 2. Comunidades etiquetadas y su tamaño

| Comunidad | Titular |
|-----------|---|
| 1.1 | ¿Cambia algo la situación en España tras la primera muerte por la nueva gripe A/H1N1? |
| 1.1 | 'La gripe A no es agresiva. Ha habido suerte' |
| 1.1 | El rebrote de la gripe del pollo en Asia reactiva el riesgo de epidemia mundial |
| 1.1 | China detecta 'gripe del pollo' en cerdos, el paso para que el virus salte a humanos |
| 1.1 | El riesgo de pandemia es inminente |
| 1.1 | Lecciones del pasado |
| 1.1 | La gripe, de vacuna en vacuna |
| 1.1 | Guía informativa de la gripe aviar |
| 1.1 | La OMS espera que el virus de la gripe se propague hasta en 'tres oleadas' distintas |
| 1.1 | ¿Y qué pasaría si todo fuera mal? |
| 1.1 | La primera pandemia del siglo |
| 1.1 | Cerca de 400.000 personas han padecido la gripe A en todo el mundo |
| 1.1 | Sanidad refuerza la campaña de la gripe por el riesgo de una epidemia a gran escala |
| 1.1 | La OMS incluye la cepa de la gripe A en la nueva vacuna de gripe estacional |
| 1.1 | Así será la pandemia de gripe aviar |
| 1.1 | La OMS pronostica una "explosión" de nueva gripe en invierno |
| 1.1 | Zafarrancho contra la gripe aviar - Público.es |
| 1.1 | Alarma en México y EE UU por una rara gripe porcina |
| 1.1 | Hora de armarse contra la pandemia |

Tabla I. Comunidad 1, ejemplo de subcomunidad 1.1

| Comunidad | Titular |
|-----------|--|
| 1.6 | Mosquitos transgénicos contra la malaria |
| 1.6 | Malaria, un recorrido desde la quinina hasta la esperada vacuna |
| 1.6 | Acabar con la malaria costará 4.400 millones durante 50 años - Público.es |
| 1.6 | Males importados pero no exóticos |
| 1.6 | Los más ricos, entre los más pobres |
| 1.6 | Un tratamiento de 0,20 euros al año por niño evita un 20% de la malaria infantil |
| 1.6 | Optimismo ante un tratamiento para prevenir la malaria en niños |
| 1.6 | La malaria regresa a Europa |
| 1.6 | Noventa millones de niños africanos siguen sin dormir bajo una mosquitera |
| 1.6 | La terapia contra la malaria sólo alcanza al 16% de los enfermos |
| 1.6 | La malaria mata cada año a más de un millón de personas en África |
| 1.6 | Una vacuna ejemplar |
| 1.6 | Pedro Alonso: 'La vacuna contra la malaria estará lista en 2011' |
| 1.6 | Bill y Melinda Gates: "No pararemos hasta la erradicación" |
| 1.6 | Las iniciativas antimalaria salvaron a 750.000 niños |

Tabla II. Comunidad 1, ejemplo de subcomunidad 1.6

La Figura 2 muestra las 13 comunidades más importantes (en número de documentos); cada una de ellas ha sido etiquetada manualmente tras una observación aleatoria de muestras de las noticias que las conforman. La Tabla III muestra un listado de las subcomunidades de la comunidad 3 (etiquetada como 'Energía'), junto con las palabras más significativas de cada una de ellas.

| Subc. | Palabras clave | Subc. | Palabras clave |
|-------|--|-------|--|
| 3.1 | energía, nuclear, solar | 3.10 | gas natural |
| 3.2 | cambio climático, emisiones CO2, kyoto | 3.11 | seísmos, volcanes |
| 3.3 | energía nuclear, Fukushima | 3.12 | nuclear, reactor fusión |
| 3.4 | energía nuclear, centrales nucleares en España | 3.13 | medio ambiente, capa de ozono |
| 3.5 | cambio climático, calentamiento global, ecologismo | 3.14 | energía, petróleo |
| 3.6 | seísmos, tsunamis | 3.15 | automóvil, motor, motor y medio ambiente |
| 3.7 | medio ambiente, residuos, contaminación | 3.16 | automóvil, tráfico, seguridad vial |
| 3.8 | medio ambiente, contaminación | 3.18 | arquitectura, rascacielos |
| 3.9 | medio ambiente, cambio climático edificación | 3.19 | casas solares |

Tabla III. Subcomunidades de la comunidad 3

En líneas generales, parece que los grandes temas científicos de los que se ocupan las noticias analizadas tienen que ver con el medio ambiente en sus diversos aspectos (energía, conservación, recursos naturales, biodiversidad, contaminación...); así como los que tienen que ver con la medicina en sentido amplio. Otros temas menos abundantes en noticias, pero bien delimitados, son investigación aeroespacial, astronomía, tecnologías de la información, evolución y política científica.

Una evaluación algo más rigurosa que la simple observación plantea algunas cuestiones, al no disponer de puntos de referencia. En efecto, en la evaluación de la clasificación automática suele distinguirse entre sistemas externos e internos. Los primeros se basan en contrastar los resultados obtenidos en una colección de documentos construida *ad-hoc*, de la cual se conoce previamente su estructura temática. Obviamente éste no es nuestro caso, por lo que debemos recurrir a un sistema interno de evaluación, es decir, medir la similitud entre los componentes de cada comunidad y la diferencia o distancia con los otros clusters.

| Com. | Etiqueta | Silue. |
|------|------------------------------|--------|
| 1 | Salud Pública | 0.62 |
| 2 | Biomedicina | 0.52 |
| 3 | Energía | 0.53 |
| 4 | Desarrollo Humano | 0.56 |
| 5 | Recursos naturales | 0.55 |
| 6 | Investigación Aeroespacial | 0.62 |
| 7 | Biodiversidad | 0.55 |
| 8 | Astronomía y Cosmología | 0.57 |
| 9 | Tecnología de la Información | 0.59 |
| 10 | Política Científica | 0.56 |
| 11 | Especies protegidas-España | 0.66 |
| 12 | Evolución Humana | 0.58 |
| 13 | Contaminación | 0.56 |

Tabla IV. Silueta media por comunidades

Una de las medidas más difundidas es la conocida como silueta (Rousseeuw, 1987; Rendón et al., 2011) de un documento. Éste es un coeficiente que se aplica a cada documento individual e intenta conjugar ambas cuestiones: la similitud con todos los integrantes de su misma comunidad y la separación con los documentos de la comunidad más próxima. Su valor oscila entre -1 y 1; los valores más alejados de -1 son mejores. Además, cuanto más se acercan a 1, indica una buena separación entre comunidades, algo que depende mucho de los campos temáticos y de las características semánticas de los documentos.

Es importante hacer notar que la valoración que ofrece esta medida debe ser tomada con precau-

ción; se basa exclusivamente en la misma valoración de la similitud (o distancia) entre documentos aplicada al efectuar la clasificación de éstos, lo cual introduce sesgos de base en este tipo de evaluación. Es, sin embargo, la medida que tenemos disponible.

La silueta es un valor individual para cada documento, por lo que para una valoración de conjunto tal vez es más útil calcular la silueta media. La Tabla V muestra los valores medios para cada comunidad importante (en número de documentos). Es preciso indicar que la separación entre comunidades se ha medido tomando como referencia la comunidad de primer nivel.

Los valores de silueta son bastante elevados; indican claramente una firme cohesión interna de cada comunidad o *cluster*, y una separación con el resto de los *clusters* bien definida.

La tabla V muestra las subcomunidades más significativas y sus valores de silueta. Dos observaciones son importantes para valorar esa tabla: en primer lugar no figuran todas las subcomunidades, puesto que ello daría lugar a una tabla de excesivo tamaño (más de 700 ítems). En su lugar figuran en ella solamente las subcomunidades de más de 100 documentos.

| Com. | Subcom. | Silueta | Com. | Subcom. | Silueta |
|------|---------|---------|------|---------|---------|
| 1 | 1 | 0,22 | 3 | 11 | 0,44 |
| 1 | 2 | 0,19 | 4 | 1 | 0,10 |
| 1 | 3 | 0,16 | 4 | 2 | 0,10 |
| 1 | 4 | 0,16 | 4 | 3 | -0,02 |
| 1 | 6 | 0,34 | 4 | 5 | 0,37 |
| 1 | 7 | 0,57 | 5 | 1 | 0,05 |
| 2 | 1 | -0,07 | 5 | 2 | 0,11 |
| 2 | 2 | 0,03 | 6 | 1 | 0,09 |
| 2 | 3 | -0,39 | 6 | 2 | 0,38 |
| 2 | 4 | -0,05 | 6 | 6 | 0,33 |
| 2 | 5 | 0,37 | 7 | 1 | -0,11 |
| 2 | 6 | -0,07 | 7 | 2 | 0,36 |
| 2 | 7 | 0,38 | 7 | 3 | 0,32 |
| 2 | 8 | 0,20 | 8 | 1 | 0,34 |
| 2 | 14 | -0,21 | 8 | 2 | 0,24 |
| 2 | 18 | -0,08 | 8 | 3 | 0,51 |
| 3 | 1 | -0,11 | 9 | 1 | 0,01 |
| 3 | 2 | 0,10 | 9 | 2 | 0,00 |
| 3 | 3 | 0,34 | 9 | 3 | 0,44 |
| 3 | 4 | 0,53 | 10 | 1 | 0,02 |
| 3 | 5 | 0,18 | 11 | 1 | 0,42 |
| 3 | 6 | 0,32 | 12 | 1 | 0,04 |
| 3 | 7 | 0,20 | 13 | 1 | 0,15 |
| 3 | 9 | 0,01 | | | |

Tabla V. Silueta de subcomunidades (sólo mayores de 100 documentos)

La tabla completa de todas las subcomunidades se ofrece en un Apéndice independiente, dada su extensión.

El cálculo de la silueta, de otro lado, se ha efectuado teniendo en cuenta solamente los ítems dentro de cada subcomunidad. Por ello los valores de silueta a nivel de subcomunidad son relativamente bajos, incluso por debajo de 0 en algunos casos. La silueta media calculada por subcomunidades (nivel 2) es de 0.21, un valor algo bajo que puede explicarse por la relativamente baja separación entre *clusters*, dado que las subcomunidades más cercanas, que sirven como base para el cálculo de la silueta, forman parte de la misma comunidad y existe, por tanto, afinidad temática aunque de un nivel más amplio. Recordemos, de otro lado, que todas las noticias o documentos tratan sobre Ciencia y/o Tecnología, lo cual introduce ya un elemento de proximidad de contenido entre todas ellas.

6. Conclusiones

La clasificación automática de documentos nos permite conocer la estructura temática y su distribución en una colección de documentos. El Análisis de Redes nos permite modelar una colección de documentos como una red de afinidades temáticas, de manera que la aplicación en esa red de sistemas de detección de comunidades descubre los temas tratados en esos documentos y la estructura temática de la colección. A diferencia de otros sistemas de organización automática, el número de temas es establecido por el propio sistema en función de las características de la colección documental; además, es posible establecer al menos un subnivel de clasificación. Los experimentos efectuados sobre una colección de noticias de prensa muestran una gran precisión en los resultados; como trabajo futuro, se plantea la aplicación de otros sistemas de detección de comunidades y la comparación entre los resultados obtenidos.

Apéndice: valores de silueta por comunidades y subcomunidades

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 2268 | 0,22 | 12 | 107 | 0,63 | 23 | 9 | 0,78 | 34 | 2 | 0,86 |
| 2 | 1681 | 0,19 | 13 | 133 | 0,54 | 24 | 18 | 0,62 | 35 | 2 | 0,92 |
| 3 | 373 | 0,16 | 14 | 51 | 0,69 | 25 | 9 | 0,64 | 36 | 3 | 0,77 |
| 4 | 725 | 0,16 | 15 | 112 | 0,53 | 26 | 6 | 0,82 | 37 | 3 | 0,80 |
| 5 | 161 | 0,46 | 16 | 40 | 0,69 | 27 | 5 | 0,83 | 38 | 3 | 0,56 |
| 6 | 296 | 0,34 | 17 | 32 | 0,62 | 28 | 3 | 0,81 | 39 | 4 | 0,81 |
| 7 | 268 | 0,57 | 18 | 38 | 0,37 | 29 | 4 | 0,82 | 40 | 2 | 0,55 |
| 8 | 138 | 0,58 | 19 | 21 | 0,81 | 30 | 7 | 0,55 | 41 | 2 | 0,72 |
| 9 | 101 | 0,62 | 20 | 29 | 0,65 | 31 | 5 | 0,79 | 42 | 2 | 0,92 |
| 10 | 158 | 0,31 | 21 | 10 | 0,79 | 32 | 2 | 0,80 | | | |
| 11 | 167 | 0,39 | 22 | 21 | 0,76 | 33 | 3 | 0,79 | | | |

Tabla VI. Comunidad 1

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 1415 | -0,07 | 50 | 28 | 0,81 | 99 | 10 | 0,91 | 148 | 6 | 0,74 |
| 2 | 1641 | 0,03 | 51 | 40 | 0,61 | 100 | 6 | 0,83 | 149 | 5 | 0,36 |
| 3 | 1377 | -0,39 | 52 | 27 | 0,45 | 101 | 7 | 0,58 | 150 | 3 | 0,49 |
| 4 | 483 | -0,05 | 53 | 31 | 0,41 | 102 | 5 | 0,65 | 151 | 3 | 0,44 |
| 5 | 311 | 0,37 | 54 | 16 | 0,56 | 103 | 10 | 0,78 | 152 | 3 | 0,79 |
| 6 | 374 | -0,07 | 55 | 38 | 0,39 | 104 | 11 | 0,25 | 153 | 5 | 0,70 |
| 7 | 275 | 0,38 | 56 | 23 | 0,09 | 105 | 9 | 0,37 | 154 | 5 | 0,38 |
| 8 | 266 | 0,20 | 57 | 16 | 0,41 | 106 | 6 | 0,54 | 155 | 4 | 0,85 |
| 9 | 188 | -0,04 | 58 | 45 | 0,12 | 107 | 7 | 0,54 | 156 | 3 | 0,71 |
| 10 | 146 | 0,25 | 59 | 66 | 0,26 | 108 | 7 | 0,63 | 157 | 4 | 0,38 |
| 11 | 97 | 0,22 | 60 | 38 | 0,64 | 109 | 3 | 0,48 | 158 | 3 | 0,71 |
| 12 | 176 | 0,10 | 61 | 33 | 0,52 | 110 | 8 | 0,79 | 159 | 2 | 0,85 |
| 13 | 100 | 0,42 | 62 | 25 | 0,57 | 111 | 9 | 0,62 | 160 | 2 | 0,55 |
| 14 | 361 | -0,21 | 63 | 20 | 0,84 | 112 | 6 | 0,81 | 161 | 3 | 0,76 |
| 15 | 107 | 0,49 | 64 | 23 | 0,76 | 113 | 5 | 0,25 | 162 | 2 | 0,85 |
| 16 | 179 | 0,33 | 65 | 16 | 0,57 | 114 | 4 | 0,59 | 163 | 4 | 0,58 |
| 17 | 94 | 0,15 | 66 | 27 | 0,58 | 115 | 6 | 0,81 | 164 | 2 | 0,53 |

| | | | | | | | | | | | |
|----|-----|-------|----|----|------|-----|---|------|-----|---|------|
| 18 | 211 | -0,08 | 67 | 11 | 0,67 | 116 | 7 | 0,65 | 165 | 3 | 0,58 |
| 19 | 95 | 0,46 | 68 | 24 | 0,58 | 117 | 5 | 0,43 | 166 | 3 | 0,32 |
| 20 | 121 | 0,11 | 69 | 23 | 0,62 | 118 | 8 | 0,54 | 167 | 3 | 0,43 |
| 21 | 110 | 0,34 | 70 | 23 | 0,80 | 119 | 4 | 0,72 | 168 | 3 | 0,48 |
| 22 | 54 | 0,05 | 71 | 18 | 0,52 | 120 | 5 | 0,65 | 169 | 2 | 0,89 |
| 23 | 83 | -0,12 | 72 | 9 | 0,66 | 121 | 5 | 0,73 | 170 | 6 | 0,79 |
| 24 | 75 | 0,09 | 73 | 12 | 0,20 | 122 | 6 | 0,44 | 171 | 2 | 0,69 |
| 25 | 60 | 0,48 | 74 | 13 | 0,53 | 123 | 3 | 0,82 | 172 | 2 | 0,79 |
| 26 | 48 | 0,58 | 75 | 12 | 0,26 | 124 | 5 | 0,53 | 173 | 3 | 0,64 |
| 27 | 84 | 0,21 | 76 | 7 | 0,48 | 125 | 5 | 0,84 | 174 | 2 | 0,88 |
| 28 | 50 | 0,47 | 77 | 14 | 0,76 | 126 | 5 | 0,72 | 175 | 4 | 0,33 |
| 29 | 44 | 0,48 | 78 | 9 | 0,61 | 127 | 4 | 0,77 | 176 | 3 | 0,78 |
| 30 | 53 | 0,47 | 79 | 17 | 0,42 | 128 | 3 | 0,70 | 177 | 2 | 0,80 |
| 31 | 48 | 0,37 | 80 | 11 | 0,40 | 129 | 6 | 0,49 | 178 | 2 | 0,84 |
| 32 | 50 | 0,39 | 81 | 9 | 0,67 | 130 | 4 | 0,68 | 179 | 3 | 0,59 |
| 33 | 30 | 0,55 | 82 | 7 | 0,85 | 131 | 3 | 0,85 | 180 | 5 | 0,77 |
| 34 | 48 | 0,66 | 83 | 11 | 0,61 | 132 | 3 | 0,77 | 181 | 3 | 0,72 |
| 35 | 41 | 0,69 | 84 | 12 | 0,73 | 133 | 6 | 0,26 | 182 | 2 | 0,64 |
| 36 | 42 | 0,27 | 85 | 13 | 0,63 | 134 | 7 | 0,37 | 183 | 2 | 0,81 |
| 37 | 35 | 0,39 | 86 | 11 | 0,33 | 135 | 4 | 0,80 | 184 | 2 | 0,37 |
| 38 | 60 | -0,05 | 87 | 11 | 0,31 | 136 | 7 | 0,87 | 185 | 3 | 0,55 |
| 39 | 78 | 0,38 | 88 | 13 | 0,37 | 137 | 2 | 0,91 | 186 | 2 | 0,62 |
| 40 | 61 | 0,15 | 89 | 11 | 0,55 | 138 | 4 | 0,48 | 187 | 2 | 0,49 |
| 41 | 32 | 0,44 | 90 | 13 | 0,42 | 139 | 3 | 0,61 | 188 | 2 | 0,91 |
| 42 | 49 | 0,46 | 91 | 12 | 0,75 | 140 | 5 | 0,45 | 189 | 2 | 0,57 |
| 43 | 17 | 0,70 | 92 | 22 | 0,66 | 141 | 5 | 0,77 | 190 | 2 | 0,82 |
| 44 | 32 | 0,59 | 93 | 12 | 0,58 | 142 | 6 | 0,52 | 191 | 2 | 0,43 |
| 45 | 36 | 0,66 | 94 | 6 | 0,56 | 143 | 3 | 0,73 | 192 | 2 | 0,72 |
| 46 | 31 | 0,53 | 95 | 15 | 0,52 | 144 | 4 | 0,78 | 193 | 2 | 0,39 |
| 47 | 31 | 0,65 | 96 | 9 | 0,51 | 145 | 2 | 0,91 | 194 | 2 | 0,83 |
| 48 | 30 | 0,58 | 97 | 10 | 0,79 | 146 | 3 | 0,57 | 195 | 2 | 0,91 |
| 49 | 67 | -0,08 | 98 | 12 | 0,60 | 147 | 3 | 0,69 | | | |

Tabla VII. Comunidad 2

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 1861 | -0,11 | 19 | 42 | 0,66 | 37 | 17 | 0,55 | 55 | 5 | 0,86 |
| 2 | 2621 | 0,10 | 20 | 57 | 0,35 | 38 | 13 | 0,64 | 56 | 5 | 0,70 |
| 3 | 710 | 0,34 | 21 | 87 | 0,41 | 39 | 13 | 0,68 | 57 | 4 | 0,67 |
| 4 | 435 | 0,53 | 22 | 44 | 0,39 | 40 | 9 | 0,59 | 58 | 4 | 0,68 |
| 5 | 587 | 0,18 | 23 | 46 | 0,54 | 41 | 16 | 0,63 | 59 | 2 | 0,55 |
| 6 | 225 | 0,32 | 24 | 35 | 0,64 | 42 | 9 | 0,68 | 60 | 5 | 0,90 |
| 7 | 277 | 0,20 | 25 | 37 | 0,64 | 43 | 8 | 0,42 | 61 | 3 | 0,77 |
| 8 | 169 | 0,35 | 26 | 29 | 0,80 | 44 | 6 | 0,78 | 62 | 3 | 0,81 |
| 9 | 222 | 0,01 | 27 | 43 | 0,46 | 45 | 7 | 0,56 | 63 | 3 | 0,46 |
| 10 | 140 | 0,07 | 28 | 39 | 0,49 | 46 | 10 | 0,62 | 64 | 2 | 0,58 |
| 11 | 273 | 0,44 | 29 | 15 | 0,61 | 47 | 2 | 0,60 | 65 | 2 | 0,37 |
| 12 | 62 | 0,74 | 30 | 23 | 0,58 | 48 | 4 | 0,78 | 66 | 2 | 0,85 |
| 13 | 78 | 0,75 | 31 | 16 | 0,81 | 49 | 3 | 0,84 | 67 | 4 | 0,60 |
| 14 | 41 | 0,63 | 32 | 14 | 0,48 | 50 | 4 | 0,29 | 68 | 2 | 0,47 |
| 15 | 145 | 0,25 | 33 | 28 | 0,61 | 51 | 5 | 0,74 | 69 | 3 | 0,62 |
| 16 | 88 | 0,26 | 34 | 23 | 0,60 | 52 | 8 | 0,43 | 70 | 2 | 0,44 |
| 17 | 59 | 0,57 | 35 | 21 | 0,50 | 53 | 5 | 0,61 | 71 | 2 | 0,80 |
| 18 | 94 | 0,32 | 36 | 20 | 0,71 | 54 | 5 | 0,46 | | | |

Tabla VIII. Comunidad 3

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 930 | 0,10 | 12 | 70 | 0,54 | 23 | 12 | 0,67 | 34 | 3 | 0,69 |
| 2 | 563 | 0,10 | 13 | 46 | 0,48 | 24 | 16 | 0,81 | 35 | 2 | 0,44 |
| 3 | 434 | -0,02 | 14 | 49 | 0,41 | 25 | 9 | 0,45 | 36 | 2 | 0,50 |
| 4 | 175 | 0,56 | 15 | 40 | 0,59 | 26 | 11 | 0,56 | 37 | 2 | 0,45 |
| 5 | 213 | 0,37 | 16 | 28 | 0,55 | 27 | 7 | 0,70 | 38 | 2 | 0,82 |
| 6 | 107 | 0,38 | 17 | 14 | 0,52 | 28 | 2 | 0,91 | 39 | 2 | 0,61 |
| 7 | 110 | 0,50 | 18 | 23 | 0,67 | 29 | 5 | 0,71 | 40 | 2 | 0,58 |
| 8 | 68 | 0,71 | 19 | 13 | 0,70 | 30 | 4 | 0,82 | 41 | 2 | 0,79 |
| 9 | 95 | 0,30 | 20 | 13 | 0,48 | 31 | 3 | 0,92 | | | |
| 10 | 51 | 0,56 | 21 | 17 | 0,67 | 32 | 3 | 0,77 | | | |
| 11 | 50 | 0,46 | 22 | 23 | 0,67 | 33 | 3 | 0,85 | | | |

Tabla IX. Comunidad 4

Comunidad 5

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 1496 | 0,05 | 16 | 16 | 0,62 | 31 | 10 | 0,52 | 46 | 6 | 0,47 |
| 2 | 238 | 0,11 | 17 | 18 | 0,58 | 32 | 5 | 0,63 | 47 | 2 | 0,90 |
| 3 | 112 | 0,36 | 18 | 24 | 0,60 | 33 | 11 | 0,62 | 48 | 4 | 0,36 |
| 4 | 58 | 0,37 | 19 | 28 | 0,74 | 34 | 8 | 0,46 | 49 | 4 | 0,70 |
| 5 | 102 | 0,17 | 20 | 21 | 0,56 | 35 | 7 | 0,90 | 50 | 3 | 0,58 |
| 6 | 62 | 0,42 | 21 | 12 | 0,57 | 36 | 6 | 0,59 | 51 | 2 | 0,87 |
| 7 | 86 | 0,12 | 22 | 22 | 0,79 | 37 | 6 | 0,56 | 52 | 4 | 0,84 |
| 8 | 90 | 0,41 | 23 | 12 | 0,49 | 38 | 5 | 0,68 | 53 | 2 | 0,38 |
| 9 | 43 | 0,70 | 24 | 24 | 0,54 | 39 | 6 | 0,60 | 54 | 2 | 0,86 |
| 10 | 76 | 0,22 | 25 | 16 | 0,34 | 40 | 4 | 0,80 | 55 | 2 | 0,67 |
| 11 | 70 | 0,46 | 26 | 15 | 0,74 | 41 | 3 | 0,68 | 56 | 2 | 0,64 |
| 12 | 31 | 0,53 | 27 | 12 | 0,71 | 42 | 4 | 0,61 | | | |
| 13 | 50 | 0,53 | 28 | 8 | 0,63 | 43 | 4 | 0,62 | | | |
| 14 | 37 | 0,63 | 29 | 12 | 0,76 | 44 | 3 | 0,75 | | | |
| 15 | 64 | 0,26 | 30 | 6 | 0,40 | 45 | 4 | 0,53 | | | |

Tabla X. Comunidad 5

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 2232 | 0,09 | 10 | 166 | 0,39 | 19 | 4 | 0,58 | 28 | 3 | 0,79 |
| 2 | 633 | 0,38 | 11 | 61 | 0,76 | 20 | 11 | 0,43 | 29 | 2 | 0,88 |
| 3 | 160 | 0,65 | 12 | 44 | 0,73 | 21 | 4 | 0,53 | 30 | 3 | 0,80 |
| 4 | 127 | 0,56 | 13 | 64 | 0,43 | 22 | 9 | 0,68 | 31 | 2 | 0,84 |
| 5 | 88 | 0,61 | 14 | 7 | 0,45 | 23 | 3 | 0,80 | 32 | 2 | 0,76 |
| 6 | 298 | 0,33 | 15 | 21 | 0,39 | 24 | 4 | 0,87 | 33 | 2 | 0,84 |
| 7 | 63 | 0,71 | 16 | 27 | 0,76 | 25 | 3 | 0,84 | 34 | 3 | 0,80 |
| 8 | 90 | 0,66 | 17 | 21 | 0,54 | 26 | 2 | 0,59 | 35 | 3 | 0,60 |
| 9 | 121 | 0,59 | 18 | 8 | 0,35 | 27 | 2 | 0,91 | 36 | 2 | 0,51 |

Tabla XI. Comunidad 6

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 719 | -0,11 | 15 | 31 | 0,50 | 29 | 16 | 0,63 | 43 | 6 | 0,23 |
| 2 | 304 | 0,36 | 16 | 35 | 0,40 | 30 | 21 | 0,59 | 44 | 3 | 0,86 |
| 3 | 209 | 0,32 | 17 | 34 | 0,18 | 31 | 13 | 0,77 | 45 | 4 | 0,58 |

| | | | | | | | | | | | |
|----|-----|------|----|----|------|----|----|------|----|---|------|
| 4 | 79 | 0,51 | 18 | 21 | 0,64 | 32 | 15 | 0,76 | 46 | 6 | 0,44 |
| 5 | 84 | 0,50 | 19 | 39 | 0,28 | 33 | 13 | 0,60 | 47 | 4 | 0,68 |
| 6 | 93 | 0,71 | 20 | 21 | 0,67 | 34 | 7 | 0,86 | 48 | 4 | 0,75 |
| 7 | 108 | 0,38 | 21 | 22 | 0,53 | 35 | 12 | 0,76 | 49 | 3 | 0,74 |
| 8 | 68 | 0,50 | 22 | 27 | 0,53 | 36 | 10 | 0,58 | 50 | 3 | 0,51 |
| 9 | 68 | 0,39 | 23 | 29 | 0,52 | 37 | 6 | 0,83 | 51 | 4 | 0,46 |
| 10 | 57 | 0,50 | 24 | 43 | 0,55 | 38 | 6 | 0,53 | 52 | 2 | 0,72 |
| 11 | 55 | 0,61 | 25 | 30 | 0,67 | 39 | 6 | 0,83 | 53 | 3 | 0,51 |
| 12 | 33 | 0,39 | 26 | 27 | 0,56 | 40 | 10 | 0,38 | 54 | 4 | 0,66 |
| 13 | 34 | 0,65 | 27 | 12 | 0,45 | 41 | 11 | 0,46 | 55 | 3 | 0,65 |
| 14 | 35 | 0,51 | 28 | 13 | 0,56 | 42 | 4 | 0,79 | 56 | 2 | 0,82 |

Tabla XII. Comunidad 7

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 448 | 0,34 | 5 | 28 | 0,44 | 9 | 9 | 0,54 | 13 | 2 | 0,64 |
| 2 | 708 | 0,24 | 6 | 26 | 0,50 | 10 | 3 | 0,56 | 14 | 2 | 0,79 |
| 3 | 215 | 0,51 | 7 | 6 | 0,84 | 11 | 5 | 0,74 | | | |
| 4 | 52 | 0,62 | 8 | 7 | 0,69 | 12 | 2 | 0,54 | | | |

Tabla XIII. Comunidad 8

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|------|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 2594 | 0,01 | 14 | 15 | 0,68 | 27 | 5 | 0,59 | 40 | 2 | 0,57 |
| 2 | 240 | 0,00 | 15 | 18 | 0,60 | 28 | 4 | 0,75 | 41 | 2 | 0,77 |
| 3 | 241 | 0,44 | 16 | 10 | 0,54 | 29 | 3 | 0,53 | 42 | 2 | 0,69 |
| 4 | 30 | 0,94 | 17 | 31 | 0,16 | 30 | 4 | 0,76 | 43 | 3 | 0,67 |
| 5 | 32 | 0,68 | 18 | 11 | 0,41 | 31 | 4 | 0,72 | 44 | 3 | 0,48 |
| 6 | 49 | 0,45 | 19 | 20 | 0,24 | 32 | 4 | 0,80 | 45 | 2 | 0,82 |
| 7 | 22 | 0,69 | 20 | 9 | 0,63 | 33 | 5 | 0,53 | 46 | 3 | 0,45 |
| 8 | 25 | 0,53 | 21 | 6 | 0,73 | 34 | 3 | 0,77 | 47 | 2 | 0,81 |
| 9 | 18 | 0,43 | 22 | 5 | 0,59 | 35 | 4 | 0,41 | 48 | 2 | 0,86 |
| 10 | 25 | 0,58 | 23 | 3 | 0,91 | 36 | 2 | 0,79 | 49 | 2 | 0,49 |
| 11 | 25 | 0,16 | 24 | 7 | 0,39 | 37 | 3 | 0,75 | 50 | 3 | 0,86 |
| 12 | 10 | 0,80 | 25 | 3 | 0,82 | 38 | 3 | 0,56 | 51 | 2 | 0,72 |
| 13 | 15 | 0,41 | 26 | 3 | 0,61 | 39 | 2 | 0,78 | | | |

Tabla XIV. Comunidad 9

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 802 | 0,02 | 11 | 19 | 0,78 | 21 | 4 | 0,63 | 31 | 2 | 0,46 |
| 2 | 161 | 0,11 | 12 | 15 | 0,76 | 22 | 4 | 0,58 | 32 | 4 | 0,63 |
| 3 | 152 | 0,13 | 13 | 25 | 0,61 | 23 | 3 | 0,67 | 33 | 3 | 0,43 |
| 4 | 105 | 0,20 | 14 | 9 | 0,52 | 24 | 3 | 0,58 | 34 | 2 | 0,62 |
| 5 | 85 | 0,33 | 15 | 10 | 0,48 | 25 | 6 | 0,37 | 35 | 2 | 0,87 |
| 6 | 38 | 0,41 | 16 | 7 | 0,73 | 26 | 3 | 0,85 | 36 | 2 | 0,32 |
| 7 | 26 | 0,76 | 17 | 7 | 0,71 | 27 | 3 | 0,83 | 37 | 2 | 0,59 |
| 8 | 29 | 0,50 | 18 | 7 | 0,76 | 28 | 2 | 0,88 | 38 | 2 | 0,52 |
| 9 | 19 | 0,63 | 19 | 4 | 0,88 | 29 | 4 | 0,48 | 39 | 2 | 0,74 |
| 10 | 23 | 0,48 | 20 | 5 | 0,82 | 30 | 3 | 0,28 | 40 | 2 | 0,64 |

Tabla XV. Comunidad 10

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 369 | 0,42 | 5 | 18 | 0,71 | 9 | 4 | 0,89 | 13 | 2 | 0,71 |
| 2 | 141 | 0,37 | 6 | 6 | 0,74 | 10 | 2 | 0,83 | 14 | 2 | 0,84 |
| 3 | 80 | 0,32 | 7 | 5 | 0,70 | 11 | 2 | 0,87 | | | |
| 4 | 79 | 0,56 | 8 | 5 | 0,67 | 12 | 2 | 0,74 | | | |

Tabla XVI. Comunidad 11

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 509 | 0,04 | 10 | 29 | 0,63 | 19 | 7 | 0,40 | 28 | 3 | 0,35 |
| 2 | 156 | 0,50 | 11 | 26 | 0,34 | 20 | 9 | 0,67 | 29 | 2 | 0,73 |
| 3 | 156 | 0,31 | 12 | 15 | 0,48 | 21 | 4 | 0,52 | 30 | 3 | 0,85 |
| 4 | 41 | 0,63 | 13 | 12 | 0,56 | 22 | 4 | 0,54 | 31 | 2 | 0,63 |
| 5 | 44 | 0,51 | 14 | 8 | 0,52 | 23 | 5 | 0,74 | 32 | 2 | 0,80 |
| 6 | 39 | 0,67 | 15 | 9 | 0,39 | 24 | 3 | 0,73 | 33 | 2 | 0,84 |
| 7 | 27 | 0,62 | 16 | 20 | 0,71 | 25 | 2 | 0,82 | 34 | 2 | 0,88 |
| 8 | 65 | 0,45 | 17 | 6 | 0,70 | 26 | 4 | 0,88 | | | |
| 9 | 36 | 0,27 | 18 | 4 | 0,77 | 27 | 5 | 0,48 | | | |

Tabla XVII. Comunidad 12

| Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta | Subcom. | N.º | Silueta |
|---------|-----|---------|---------|-----|---------|---------|-----|---------|---------|-----|---------|
| 1 | 409 | 0,15 | 6 | 9 | 0,63 | 11 | 5 | 0,65 | 16 | 4 | 0,67 |
| 2 | 139 | 0,46 | 7 | 16 | 0,39 | 12 | 7 | 0,57 | 17 | 3 | 0,70 |
| 3 | 83 | 0,39 | 8 | 11 | 0,58 | 13 | 5 | 0,65 | 18 | 4 | 0,55 |
| 4 | 15 | 0,62 | 9 | 10 | 0,75 | 14 | 3 | 0,83 | | | |
| 5 | 16 | 0,60 | 10 | 22 | 0,50 | 15 | 7 | 0,58 | | | |

Tabla XVIII. Comunidad 13

Referencias

- Aggarwal, C. C. y Zhai, C. (2012). A survey of text clustering algorithms. // Aggarwal y Zhai, eds.: Mining Text Data. Springer US: Boston MA. 77-128
- Ares Brea, M.E.; Parapar López, J.; Barreiro García, A. (2011). Agrupamiento Documental. // Cacheda Seijo, F.; Fernández Luna, J. M.; Huete Guadix, J. F. Eds. (2011). Recuperación de Información: Un enfoque práctico y multidisciplinar. Madrid; Ra-Ma, 2011. 392-416.
- Arun, R.; Suresh, V.; Veni Madhavan, C. E.; Narasimha Murthy, M. N.; Zaki, M. J.; Yu, J. X.; Ravindran, B.; Pudi, V. (2010). On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. // Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010. Hyderabad, India. 391-402. http://dx.doi.org/10.1007/978-3-642-13657-3_43 (2017-01-12).
- Blei, D., Ng, A.; Jordan, M. (2003). Latent dirichlet allocation. // The Journal of Machine Learning Research. 3, 993-1022.
- Baharudin, B.; Lee, L. H.; Khan, K. (2010). A review of machine learning algorithms for text-documents classification. // Journal of Advances in Information Technology. 1:1, 4-20.
- Bohlin, L.; Edler, D.; Lancichinetti, A.; Rosvall, M. (2014). Community detection and visualization of networks with the map equation framework. // Measuring Scholarly Impact. Springer International Publishing. 3-34.
- Campos Ibáñez, L. M.; Romero López, A. E. (2011). Clasificación documental. // Cacheda Seijo, F.; Fernández Luna, J.M.; Huete Guadix, J.F. Eds. (2011). Recuperación de Información: un enfoque práctico y multidisciplinar. Madrid; Ra-Ma, 2011. 359-392.
- Edler, D.; Rosvall, M. (2015). The infomap software package. <http://www.mapequation.org/code.html> (2017-02-16).
- Eyheramendy, S.; Lewis, D. D.; Madigan, D. (2003). On the naive bayes model for text categorization. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.4949> (2017-02-16).
- Figuerola, C. G. (2013). Clasificación automática de documentos: Un caso práctico. <http://grulla.usal.es/figuerola2013clasificacion.pdf> (2017-02-16).
- Figuerola, C. G.; Quintanilla Fisac, M. A.; et al. (2017): Sistema de Indicadores para el SCSC (Spanish Corpus of Scientific Culture). <http://grulla.usal.es/figuerola2017sistema.pdf> (2017-03-28).
- Figuerola, G. C.; García Marco, F. J.; Pinto, M. (2017). Mapping the evolution of library and information science (1978-2014) using topic modeling on LISA. Scientometrics. 11:23, 1507-1535.
- Griffiths, T. L.; Steyvers, M. (2004). Finding scientific topics. // Proceedings of the National Academy of Sciences. 101:1, 5228-5235.
- Groves, T; Figuerola, C. G.; Quintanilla, M. A (2015). Ten years of science news: a longitudinal analysis of scientific culture in the Spanish digital press. Public Understanding of Science. 25:6, 691-705. <https://gredos.usal.es/jspui/handle/10366/127539> (2017-02-16)

- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. // *Pattern recognition letters*. 31:8, 651-666. http://www.ppgia.pucpr.br/~fabricio/ftp/RogesJainClustering_PRL10.pdf (2017-02-18)
- Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. // *European conference on machine learning* (pp. 137-142). Springer Berlin Heidelberg. https://eldorado.tu-dortmund.de/bitstream/2003/2595/1/report23_ps.pdf (2017-02-16).
- Joachims T. (2002) *Learning to Classify Text Using Support Vector Machines – Methods, Theory and Algorithms*. Boston, MA: Kluwer Academic Publishers.
- Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H. (2006). Some effective techniques for naive bayes text classification. // *IEEE transactions on knowledge and data engineering*. 18:11, 1457-1466. <http://ir.kaist.ac.kr/papers/2006/some%20effective%20techniques%20for%20naive%20bayes%20text%20classification.pdf> (2017-02-16).
- Hidayat, E. Y.; Firdausillah, F.; Hastuti, K.; Dewi, I. N.; Azhari, A. (2015). Automatic Text Summarization Using Latent Dirichlet Allocation (LDA) for Document Clustering. // *International Journal of Advances in Intelligent Informatics*, 1:3, 132-139.
- Lancichinetti, A.; Fortunato, S. (2009). Community detection algorithms: A comparative analysis. // *Physical Review E*. 80:5. <http://arxiv.org/pdf/0908.1062v2.pdf> (2017-02-18)
- Langley, P.; Iba, W.; Thompson, K. (1992). An analysis of bayesian classifiers. // *Proceedings of National Conference on Artificial Intelligence*. San Antonio, CA: AAAI Press and MIT Press. 223-228
- Lee, C.; Cunningham, P. (2014) Community detection: Effective on large social networks. // *Journal of Complex Networks*. 2:1, 19-37. <http://comnet.oxfordjournals.org/content/2/1/19.full.pdf+html> (2017-02-18)
- Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. // *Journal of the American Society for Information Science and Technology*. 59:1, 77-85.
- Martin, S.; Brown, M.W.; Klavans, R.; Boyack K.W.(2011). OpenOrd: an open-source toolbox for large graph layout. // *Proc. SPIE 7868, Visualization and Data Analysis 2011*. doi:10.1117/12.871402
- Martin-Pozuelo Campillos, M. P. (1996). *La construcción teórica en archivística: el principio de procedencia*. Madrid: Universidad Carlos III de Madrid.
- McCallum, A.; Nigam, K. (1998) A comparison of event models for naive bayes text classification. // *AAAI-98 workshop on learning for text categorization*. 41-48. <http://www.kamalnigam.com/papers/multinomial-aaaiws98.pdf> (2016-12-14)
- Otte, E.; Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. // *Journal of information Science*. 28:6, 441-453. http://www.academia.edu/download/42254790/Social_Network_Analysis_A_Powerful_Strat20160206-25456-1pc1cl.pdf (2017-02-18)
- Plantié, M. ; Crampes, M. (2013) Survey on social community detection. // *Social media retrieval*, 65-85. <http://hal.archives-ouvertes.fr/docs/00/80/42/34/PDF/Survey-on-Social-Community-Detection-V2.pdf> (2017-02-18)
- Pons, P.; Latapy, M. (2005). Computing communities in large networks using random walks. // *Computer and information sciences (ISCIS)* 284-293. <http://arxiv.org/abs/physics/0512106> (2017-02-18)
- Rendón, E.; Abundez, I.; Arizmendi, A.; Quiroz, E. (2011). Internal versus external cluster validation indexes. // *International Journal of computers and communications*. 5:1, 27-34.
- Rosvall, M.; Axelsson, D.; Bergstrom, C. (2009). The map equation. // *European Physical Journal Special Topics*. 178, 13-23.
- Rousseeuw, P. J. (1987). *Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis*. *Computational and Applied Mathematics*. 20, 53-65. doi:10.1016/0377-0427(87)90125-7.
- Salton, G.; McGill, M.J. (1983) *Introduction to Modern Information Retrieval*. New York, NY: McGraw-Hill.
- Scott, J. (2013). *Social network analysis*. Thousand Oaks, CA, US: Sage Publications, Inc
- Shawn, G.; Milligan, I. (2012). Review of MALLET, produced by Andrew Kachites McCallum. // *Journal of Digital Humanities*, 2:1. <http://journalofdigitalhumanities.org/2-1/review-mallet-by-ian-milligan-and-shawn-graham/> (2017-03-15)
- Yang, Y. (1999). An evaluation of statistical approaches to text categorization. // *Information retrieval*. 1:1-2, 69-90.

Enviado: 2017-04-18. Segunda versión: 2017-06-26.
Aceptado: 2017-09-15.

Análise de domínio em Ciência da Informação: uma análise da produção científica internacional

El análisis de dominio en la Ciencia de la Información: un análisis de la producción científica internacional

Domain Analysis in Information Science: an analysis of the international scientific production

**José Augusto Chaves GUIMARÃES, Daniel MARTÍNEZ-ÁVILA, Amanda MARTINS,
Pedro Henrique Carvalho GOMES**

Departamento de Ciência da Informação, Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Marília, SP (Brasil)
gpfapoi@gmail.com

Resumen

Se analiza la presencia del análisis de dominio en la literatura internacional de ciencia de la Información en Scopus, LISA y Web of Science de 1995 a 2016. Se utiliza el análisis de dominio, en su abordaje bibliométrico (Hjørland, 2002), en aspectos relativos a los autores más citados, vehículos de publicación y períodos más productivos para identificar cómo se constituyen las comunidades epistémicas de ese dominio (Meyer & Molyneux-Hodgson, 2010). Los resultados revelan que este tema viene siendo productivo en el área, creciendo fuertemente creciente en la última década, y publicado en las más importantes revistas científicas del área, más específicamente Knowledge Organization, Scientometrics, JASIST, Journal of Documentation e Information Processing & Management. El análisis de los referentes teóricos reveló una elite de 64 autores, en los que las temáticas de investigación vinculadas a la organización del conocimiento y a los estudios métricos en información destacan, revelando la articulación de dos importantes enfoques de análisis de dominio previstos por Hjørland, autor que recibe el más alto número de citas, en expresiva diferencia con los demás, revelando su centralidad en esa temática, en la que fue precursor en la Ciencia de la Información hace dos décadas.

Palabras clave: Análisis de dominio. Comunidades epistémicas. Organización del conocimiento.

1. Introdução

A análise de domínio vem sendo objeto de abordagem e de investigação pela Ciência da Informação já há cerca de duas décadas, constituindo uma importante abordagem para caracterização e avaliação da ciência pois permite identificar as condições pelas quais o conhecimento científico se constrói e se socializa (Hjørland & Albrechtsen, 1995; Moya-Anegón & Herrero-Solana, 2001, Hjørland, 2002, 2004; Tennis, 2003; Smiraglia, 2011).

Abstract

The presence of domain analysis in the international literature on Information Science in the databases Scopus, LISA and Web of Science is analysed for the period 1995-2016. In order to achieve this, we conduct a domain analysis, and more specifically its bibliometric approach (Hjørland, 2002), of aspects related to authors, scientific venues, most productive periods, as well as the most cited authors, in order to identify how the epistemic communities of this domain are conformed (Meyer & Molyneux-Hodgson, 2010). The results show that this has been a prolific topic in the area, with a considerable growth of publications in the last decade, and published in the most important scientific journals in the area, namely Knowledge Organization, Scientometrics, JASIST, Journal of Documentation, and Information Processing & Management. The analysis of the theoretical references revealed an elite of 64 authors, in which research topics related to knowledge organization and metric studies in information stand out, revealing the articulation of two important approaches to domain analysis by Hjørland. This author is also revealed as the most cited author by far, and a central actor in this subject, as he was also its founder in Information Science two decades ago.

Keywords: Domain analysis. Epistemic communities. Knowledge organization.

Por meio da análise de domínio torna-se possível verificar o que é efetivamente importante ou significativo em um dado campo, de tal modo que aspectos como tendências, padrões, processos, agentes e seus relacionamentos possam ser identificados e analisados (Danuello, 2007). Tem-se, assim, o estudo dos aspectos teóricos de um dado entorno, geralmente representado por uma literatura ou comunidade de pesquisadores, constituindo um meio para a geração de novo conhecimento acerca da interação de uma dada comunidade científica com a informação (Smiraglia, 2011).

Mas foi efetivamente a partir de um artigo seminal de Birger Hjørland e Hanne Albrechtsen que a análise de domínio teve espaço mais efetivo na Ciência da Informação, pautando-se em um “paradigma social, que concebe a CI como uma ciência social, considerando os contextos psicossocial, sociolinguístico e a sociologia do conhecimento e a ciência” (Hjørland & Albrechtsen, 1995, p. 400, traduzido).

A análise de domínio, na Ciência da Informação, e mais especialmente na organização do conhecimento, tem especial importância no estudo dos domínios de coque apresentam certa consonância em termos de pensamento, linguagem e conhecimento (Nascimento & Marteleto, 2004; Danuello, 2007).

O domínio atua como paralelo às disciplinas, contribuindo para a identificação de uma “ecologia do trabalho” caracterizando-se como um grupo com uma ontologia coerente que compartilha uma mesma epistemologia, que lhe confere as fronteiras intelectuais, tem-se um colégio invisível que pressupõe elementos intelectuais comuns e um discurso efetivo que ocorrem em uma unidade socialmente estruturada. Desse modo, torna-se possível identificar, por um lado, correntes teóricas e, por outro, uma rede social do mundo acadêmico (Smiraglia, 2012). No entanto, sua dimensão conceitual apresenta-se ainda um tanto fluida, sendo entendido como “comunidade de pensamento ou comunidade discursiva que integra a divisão social do trabalho” (Hjørland & Albrechtsen, 1995, p.401, traduzido), ou mesmo uma “área de especialidade, um conjunto literário ou um grupo de pessoas trabalhando juntas numa organização” (Mai, 2005, p. 605, traduzido).

Hjørland (2002) enunciou um conjunto de onze “abordagens” que, a seu ver, caracterizariam a análise de domínio, na medida em que é a partir da aplicação de mais de uma delas ao mesmo domínio que se torna possível melhor conhecê-lo enquanto tal. São elas: Produção de obras de referência, Construção de linguagens de indexação, Indexação e recuperação da informação, Estudo de usuários, Estudos bibliométricos, Estudos históricos, Estudos de gêneros/tipologias documentais, Estudos epistemológicos e críticos, Estudos terminológicos, Comunicação científica, Cognição científica, conhecimento especializado e inteligência artificial. A esse grupo soma-se mais uma – a da proveniência arquivística – proposta por Guimarães e Tognoli (2015).

Procurando especificar melhor as proposições de Hjørland, no intuito de dar-lhes maior aplicabilidade, Tennis (2003, p.193, traduzido) propõe dois eixos a partir dos quais a análise de domínio

pode ser abordada: as áreas de modulação, que “fornecem parâmetros para as denominações e os limites do domínio e os graus de especialização, que estabelecem a profundidade – a especificidade - de um domínio”.

A abordagem da análise de domínio vem sendo especialmente importante para a organização do conhecimento na medida em que os processos de tratamento passam a ser abordados a partir do contexto de produção - e de uso - daquele conhecimento, sendo especialmente importante para os estudos comparativos entre a terminologia da literatura científica de um domínio e as linguagens de indexação da área, a análise do universo e das relações entre referentes e correntes teóricas de um domínio ou, ainda, a análise das relações entre temas, referentes teóricos e correntes teóricas em um domínio.

Dentre as abordagens propostas por Hjørland, especial destaque merece, para fins do presente trabalho, a abordagem bibliométrica, uma vez que os estudos bibliométricos (aqui inserindo-se as questões cienciométricas e informétricas), como destacam Hjørland & Albrechtsen (1995, p.450, traduzido), “organizam padrões sociológicos de reconhecimento explícito entre documentos individuais”. Para os referidos autores as análises desse tipo contribuem para evidenciar não apenas a natureza de uma disciplina como também as relações entre disciplinas diversas, em um contexto mais amplo, tais como os padrões sociais na comunicação científica (Hjørland & Albrechtsen, 1995, p. 403; 413).

Desse modo, o presente trabalho visa a realizar uma análise de domínio, em abordagem bibliométrica, da produção científica internacional sobre análise de domínio no campo da Ciência da Informação.

2. Metodologia

Para a consecução do objetivo de investigação, realizaram-se buscas, na primeira semana de janeiro de 2017, a partir do termo “domain analysis” nas bases Scopus, LISA e Web of Science, tendo-se como datas-baliza 1995 a 2016. Justifica-se a escolha do ano de 1995 como início da busca pelo fato de nele haver sido publicado o artigo seminal de Hjørland e Albrechtsen, que motivaram as discussões desse tema na Ciência da Informação.

No caso da base LISA, a busca deu-se pela presença desse termo no título e/ou no resumo (visto não permitir busca por palavras-chaves). Desse modo, obteve-se um total de 105 documentos. No caso da base Scopus, a busca deu-se pela presença desse termo no título e/ou no

resumo e/ou palavra-chave nas áreas temáticas de Ciências Sociais e Artes e Humanidades, obtendo-se um total de 1920 documentos. Vale ressaltar que essa alta incidência deveu-se ao fato de, por serem áreas textuais de linguagem livre, observa-se a possibilidade de uso metafórico do termo domínio, não necessariamente relacionado a esse campo de estudos. A No caso da base Web os Science, a busca deu-se pela presença desse termo no título e/ou no resumo e/ou palavra-chave. Desse modo, obteve-se um total de 60 documentos.

Em seguida, foram analisados os artigos de modo a excluir a repetições entre uma base e outra bem como aqueles que não guardavam relação direta com o campo da Ciência Informação, o que permitiu chegar a um corpus final de 128 documentos diferentes.

Para fins de análise dos dados, utilizou-se da análise de domínio, mais especificamente em sua abordagem bibliométrica (Hjørland, 2002), em aspectos relativos aos autores, veículos de publicação e períodos mais produtivos, bem

como os autores mais citados, de modo a identificar como se constituem as comunidades epistêmicas desse domínio (Meyer & Molyneux-Hodgson, 2010).

3. Apresentação, análise e discussão dos dados

Foram identificados 200 (duzentos) autores de trabalhos, considerando-se, como mais produtivos, aqueles que publicaram dois ou mais trabalhos, totalizando um conjunto de 33 autores (16,5% do total), nomeadamente: Moya-Anegón (13 Artigos); Herrero-Solana (9 Artigos); Smiraglia e Vargas-Quesada (7 Artigos); Chinchilla-Rodríguez (6 Artigos); Chen, Corera-Alvarez, Hjørland, Muñoz-Fernández, Robinson, (4 Artigos); Diaz, Llorens, López-Huertas e Martínez (3 Artigos) e Albrechtsen, Freund, Guerrero-Bote, Guimarães, Hartel, Kim H., Kim M.C., Kim P.J., Lee, Marchese, Marteleto, Miguel, Nascimento, O'Keefe, Paul, Tennis, Wang L., Wang S.E. e White (2 Artigos).

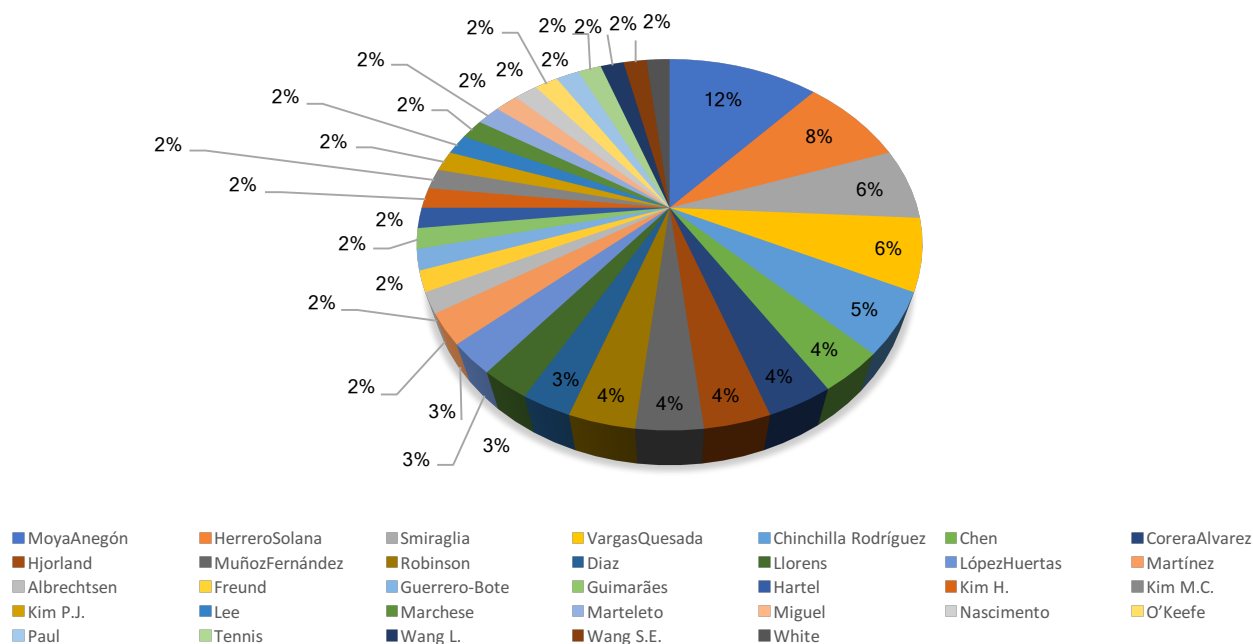


Figura 1

Nesse conjunto, especial destaque merece a Universidad de Granada, relativamente aos autores ligados do grupo Scimago (Moya-Anegón, Herrero-Solana, e Vargas-Quesada), a que se alia a tradição anglo-saxônica com Smiraglia e Hjørland, entre outros. Essa produção foi predominantemente veiculada em revistas científicas (96,1% do total), totalizando 123 artigos publica-

dos em um conjunto de 53 (cinquenta e três) revistas científicas, das quais 17 (dezesete), publicaram dois ou mais artigos, respondendo por 70,7% dos artigos publicados sendo elas, nomeadamente: Knowledge Organization (16 Artigos); Scientometrics (14 Artigos); Journal of the American Society for Information Science and Technology (13 Artigos); Journal of Documenta-

tion (9 Artigos); Information Processing & Management (9 Artigos); Information Research (4 Artigos); Journal of Information Science e Investigacion Bibliotecologica (3 Artigos); Transinformação, Journal of the Association for Information Science and Technology, Information Systems, Information Resources Management Journal, Información, Cultura y Sociedad, Aslib Proceedings, Advances in Classification Research Online, Profesional de la Información e Revista Española de Documentacion Científica (2 Artigos).

É interessante observar que 61 artigos, correspondendo a praticamente metade da produção, encontram-se publicados em 5 revistas, das quais 3 têm tradicionalmente se dedicado a questões teóricas que envolvem a organização do conhecimento, uma diretamente voltada para estudos métricos e outra de viés mais tecnológico. Tais aspectos reforçam as abordagens que vêm predominantemente permeando essa literatura.

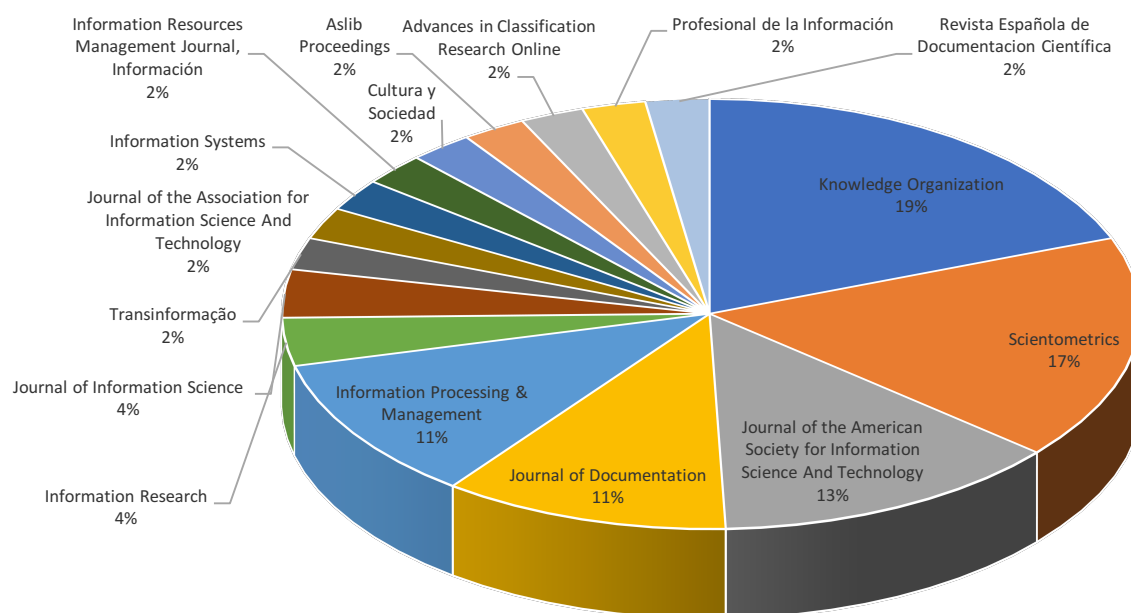


Figura 2. Veículos de publicação em Análise de domínio

Observa-se, a exemplo do que ocorre na área de Ciência da Informação como um todo, uma franca predominância de literatura em língua inglesa, com exceção de um periódico em português (Transinformação, publicado no Brasil) e três em espanhol (Información, Cultura y Sociedad, publicado na Argentina e Profesional de la Información e Revista Española de Documentación Científica, publicados na Espanha). É interessante observar, ainda, que as duas revistas que lideram esse grupo são nomeadamente Knowledge Organization e Scientometrics, o que revela ser a análise de domínio um elemento de conexão entre a organização do conhecimento e os estudos métricos, tal como já explanado por Hjørland (2002). Por outro lado, prestigiosas revistas anglo-saxônicas de cunho mais geral da Ciência da Informação, tais como Journal of the American Society for Information Science and Technology, Journal of Documentation, Information Processing & Management e Information Re-

search igualmente trazem produção sobre análise de domínio o que evidencia a efetiva importância dessa temática para a pesquisa em Ciência da Informação na atualidade. Em termos cronológicos, observa-se que a literatura sobre análise de domínio foi se tornando gradativamente mais frequente na Ciência da Informação, observando-se um impulso mais efetivo a partir de 2008 (64,1% do total) e com forte concentração nos dois últimos anos analisados - 2015 e 2016 - que responderam por 28,1% da produção total. Isso revela que o tema vem se incorporando efetivamente no universo teórico-metodológico da Ciência da Informação e vem recebendo cada vez mais atenção de seus pesquisadores. Uma análise dos autores citados nos trabalhos que compõe o corpus da pesquisa revela um conjunto 3715 autores (excluídas as autocitações) dos quais 75,9% receberam apenas uma citação, o que revela um alto índice de dispersão de fontes, o que pode ser considerado tanto como um aspecto positivo (uma temática ainda

em expansão e aberta aos aportes mais diversificados) quando negativo (uma área que ainda não conseguiu sedimentar de forma mais consistente seu núcleo básico de referentes teóricos). Aplicou-se a Lei de Elitismo de Price (1963) ao conjunto de 3715 autores, obtendo-se a raiz quadrada da ordem de aproximadamente 61, o que leva à análise do conjunto de 64 autores que receberam um mínimo de 8 citações, a saber: Hjørland, B. (151 citações); Albrechtsen, H. (49 citações); McCain, K.W. e White, H.D. (39 citações); Small, H. (36 citações); Leydesdorff, L. (32 citações); Garfield, E. (30 citações); CHEN, C. (26 citações); Vakkari, P. (25 citações); Moya-Anegón, F. (23 citações); Prieto-Díaz, R. (22 citações); Boyack, K.W. (21 citações); Bates, M.J. e Buckland, M. (20 citações); Gladwell, M. (19 citações); Moed, H.F. e Seglen, P.O. (18 citações);

Börner, K. e Griffith, B. (17 citações); Saracevic, T. (16 citações); Ingwersen, K.P., Järvelin, K. e Persson, O. (15 citações); Belkin, N. (14 citações); Hartel, J., Mai, J.E, Salton, G. e Smiraglia R.P. (13 citações); Klavans, R., Lin, X., Van Raan, A.F.J., Vickery, B.C. (12 citações); Cook, D.J., Cronin, B., Holder, L. B. Kostoff, R.N., Llorens, J. Noyons, E.C.M. e Tennis, J.T. (11 citações); Beghtol, C., Callon, M., Herrero-Solana, V., Lykke-Nielsen, M., Mills, J., Rees-Potter, L.K., Talja, S., Wilson, P. (10 citações); Courtail, J. P., Ding, Y., Klein, J.T., Riedl, J., Rousseau, R., Savolainen, R., Schubert A. e Tijssen, R.J.W. (9 citações); Bawden, D., Bazerman, C., Braam, R.R., Capurro, R., Chinchilla-Rodriguez, Z., Narin, F., Price, D.J.S., Swales, J.M. e Tibbo, H.R. (8 citações).

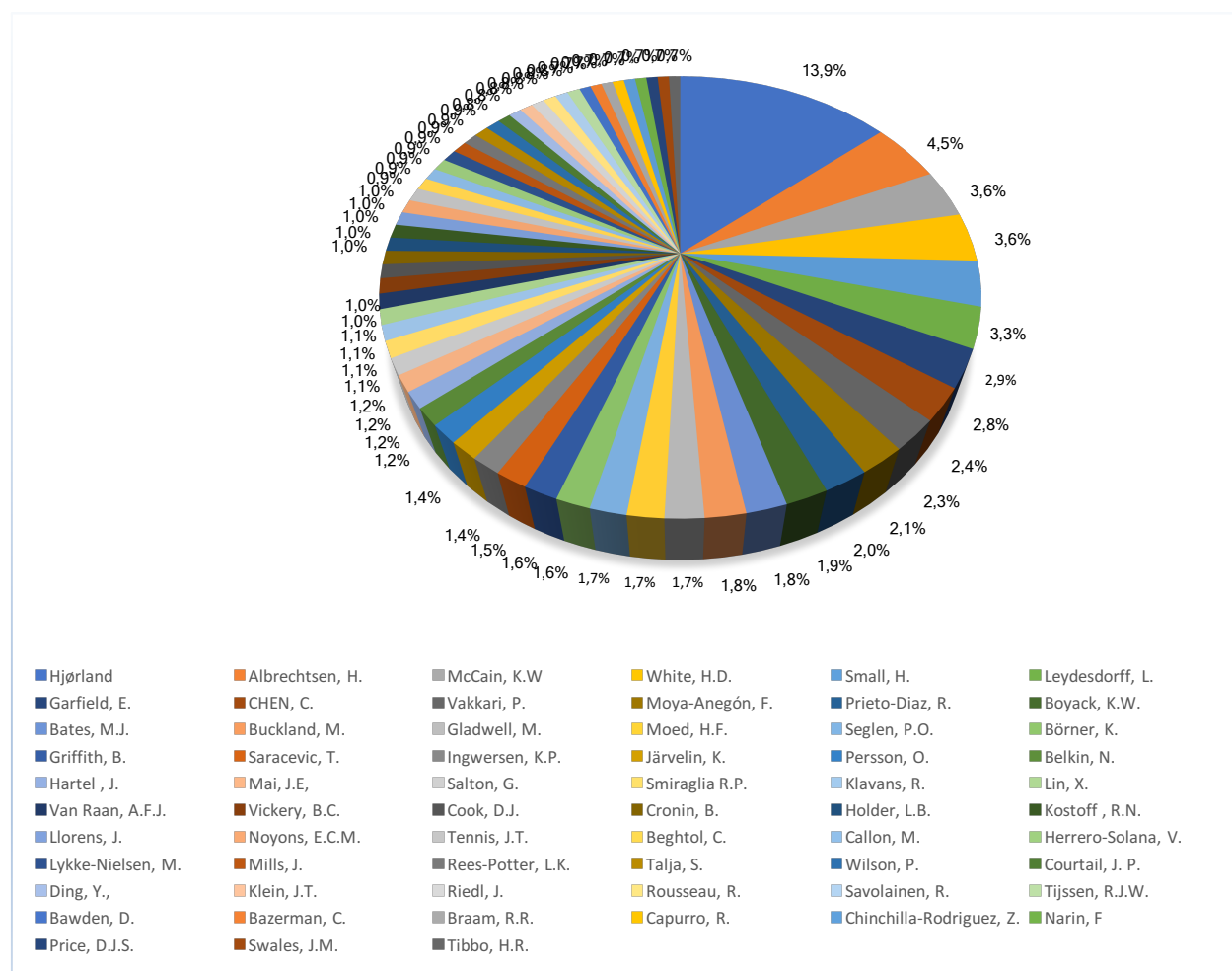


Figura 3

Nesse conjunto, têm-se uma forte predominância de Hjørland, o que reforça seu papel não apenas seminal (a partir de seu artigo com Albrechtsen

há mais de duas décadas) como também sua permanência como um referente teórico basilar nesse campo de estudos. Nesse caso específico,

observa-se que a tônica é muito mais epistemológica – um dos objetos primordiais dos estudos de Hjørland – sendo seguida de abordagens mais aplicadas a partir de autores mais diretamente ligados aos estudos métricos em informação.

Uma análise desse conjunto de autores revelou uma complementaridade de áreas que mais significativamente fornecem fundamentos aos estudos de análise de domínio, como se vê na Figura 2.

Analisando-se, nesse grupo de elite, as especialidades de pesquisa de cada referente teórico, a partir de sua produção bibliográfica geral, obtida em páginas web pessoais e institucionais, observa-se uma maior presença dos Estudos métricos em informação (Bibliometria, Cientometria, Infometria, Webmetria), que respondem por 42,0% das citações, seguida da Organização do Conhecimento, com 30,9% das citações que, juntos respondem por 73% das citações ao que se aliam os Estudos epistemológicos em Ciência da Informação, com 10,8% das citações. É interessante observar a presença de autores ligados à Competência e ao Comportamento Informacional (2,6%). Tal cenário revela que a temática da análise de domínio vem sendo tratada, na literatura da Ciência da Informação, prioritariamente entre três abordagens previstas por Hjørland (2002), a da organização do conhecimento, os estudos métricos e os estudos epistemológicos e históricos, revelando uma interessante triangulação temática.

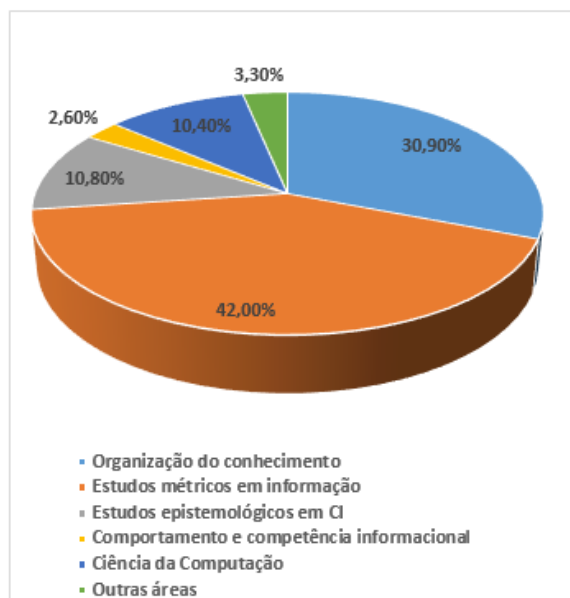


Figura 4. Vinculação temática dos referentes teóricos

Em termos de autores citados, e reiterando estudos anteriores, observa-se um papel nuclear de

Birger Hjørland que, com 151 citações recebidas, responde por 13,9% nesse grupo de elite (e 3,1% do total de citações), o que se explica plenamente dado seu papel seminal nessa temática na Ciência da Informação. Nesse contexto, merece destaque a presença de uma forte e articulada comunidade epistêmica em organização do conhecimento, composta por Albrechtsen, H.; Bates, M.J.; Beghtol, C.; Buckland, M.; Hjørland, B.; Lykke-Nielsen, M.; Mai, J.-E.; Mills, J.; Riedl, J.; Smiraglia R.P.; Swales, J.M.; Tennis, J.T.; e Vickery, B.C. que, em sua grande maioria, tem expressiva ação científica no âmbito da International Society for Knowledge Organization – ISKO e em seus veículos oficiais de divulgação da produção científica: a revista Knowledge Organization e a série Advances in Knowledge Organization. Nessa comunidade de organização do conhecimento, por sua vez, observa-se uma forte predominância de autores com preocupações de natureza epistemológica, o que evidencia o efetivo aporte da análise de domínio para o universo teórico-metodológico da organização do conhecimento.

4. Conclusão

Como se pode observar, a Análise de domínio vem ocupando um espaço cada vez mais significativo na literatura científica internacional da Ciência da Informação. Tal aspecto se revela por meio de uma produção científica em forte crescimento em especial na última década, e visível em publicações de prestígio internacional na área, em sua grande maioria pertencentes ao universo anglo-saxão.

Esse campo revela ainda um momento de sedimentação teórica, com forte diversidade de fontes, mas já permite identificar a presença de comunidades epistêmicas desses referentes, em especial no que se refere aos estudos métricos em informação e à organização do conhecimento, duas abordagens bastantes fortes de análise de domínio dentre as onze propostas por Hjørland (2002), e que efetivam um importante diálogo.

Mais especificamente no que tange à Organização do Conhecimento, observa-se um grupo de onze importantes referentes teórico, em sua maioria em forte dialogicidade como International Society for Knowledge Organization – ISKO.

Por fim, e reafirmando estudos anteriores, destaca-se o papel central de Hjørland, que não apenas foi precursor dessa abordagem na Ciência da Informação como, e principalmente, a vem aperfeiçoando ao longo de duas décadas, cada vez mais incorporando elementos teóricos e metodológicos que a robustecem.

Agradecimento

Os autores agradecem o Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil – CNPq pelo apoio recebido para essa pesquisa.

Referencias

- Danuello, Jane Coelho (2007). Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise do domínio. Dissertação (Mestrado). Marília: UNESP, 2007.
- Guimarães, José Augusto Chaves; Tognoli, Natália Bolfarini (2015). Provenance as a domain analysis approach in archival knowledge organization. // *Knowledge Organization*. 42:8 (2015) 562-569.
- Hjørland, Birger (2002). Domain analysis in information Science: eleven approaches traditional as well as innovative. // *Journal of Documentation*. 58:4 (2002) 422-462.
- Hjørland, Birger (2003). Fundamentals of Knowledge Organization. // *Knowledge Organization*. 30:2 (2003) 87-111.
- Hjørland, Birger (2004). Domain analysis: a socio-cognitive orientation for Information Science research. // *Bulletin of the American Society for Information Science and Technology*. 30:3 (feb./mar. 2004). <http://www.asis.org/Bulletin/Feb-04/hjorland.html> (07/04/17).
- Hjørland, Birger; Albrechtsen, Hanne (1995). Toward a new horizon in information science: domain-analysis. // *Journal of the American Society for Information Science*. 46:6 (1995) 400-425.
- Mai, Jens-Erik (2005). Analysis in indexing: document and domain centered approaches. // *Information processing and management*. 41:3 (May 2005) 599-661.
- Meyer, Morgan; Molyneux-Hodgson, Susan (2010). Introduction: the dynamics of epistemic communities. // *Sociological Research Online*. 15:2 (2010) 14. <http://www.socresonline.org.uk/15/2/14/14.pdf> (07/04/2014).
- Moya-Anegón, Felix de; Herrero-Solana, Víctor (2001). Análisis de dominio de la revista mexicana de investigación bibliotecológica. // *Información, cultura y sociedad*. 5 (2001) 10-28.
- Nascimento, Denise Morado; Marteleto, Regina Maria (2004). A "informação construída" nos meandros dos conceitos da Teoria Social de Pierre Bordieu. // *DataGramaZero*. 5:5 (out./2004). http://www.dgz.org.br/out04/Art_05.htm. (07/04/2014).
- Price, Derek J. de Solla (1963). *Little science, big science*. New York: Columbia University Press, 1963.
- Smiraglia, Richard P. (2011) Domain coherence within Knowledge Organization: People, Interacting Theoretically, Across Geopolitical and Cultural Boundaries. // McKenzie, P.; Johnson, K.; Stevens, S. (ed.). *Exploring interactions of people, places and information. Proceedings of the 39th Annual CAIS/ACSI Conference*, University of New Brunswick, Fredericton, Canada, June 2-4, 2011. <http://www.cais-acsi.ca/ojs/index.php/cais/article/view/310/63>. (07/04/2014).
- Smiraglia, Richard P. (2012). Epistemology of Domain Analysis. // Lee, Hur-Li; Smiraglia, Richard P. (eds.). *Wurzburg: Ergon*, 2012. 111-124
- Tennis, Joseph T. (2003). Two Axes of Domain Analysis. // *Knowledge Organization*. 30:3/4 (2003) 191-195.

Enviado: 2017-04-07. Segunda versão: 2017-09-05.
Aceptado: 2017-09-15.

A não univocidade do conceito de “autor” sob a perspectiva de Wittgenstein e Foucault

La no univocidad del concepto de autor desde la perspectiva de Wittgenstein y Foucault

The non-univocity of the concept of “author” from the perspective of Wittgenstein and Foucault

Eduardo Graziosi SILVA, Daniel MARTINEZ-AVILA, Luciana de Souza GRACIOSO

Departamento de Ciência da Informação, Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Marília, SP (Brasil)
eduardograziosi@marilia.unesp.br, dmartinezavila@marilia.unesp.br, luciana@ufscar.br

Resumen

Debido especialmente a las Tecnologías de la Información y Comunicación (TIC), el concepto de “autor” y las formas de autoría se han ampliado independientemente del área en el que se crea el contenido. En este trabajo se analizan los usos, reutilizaciones y resignificados atribuidos al concepto de autor en los campos del Derecho, Ingeniería y Biblioteconomía bajo la perspectiva del “Juego de lenguaje” de Wittgenstein y el concepto de “autor” de Foucault. Se comprueba la no univocidad del concepto en función de las transformaciones en el tiempo y las reglas y convenciones de cada área de estudio.

Palabras clave: Autor. Autoría. Foucault, Michel. Wittgenstein, Ludwig.

1. Introdução

O conceito de “autor” não diz mais respeito apenas àquele que escreve textos, pinta quadros, compõe músicas ou canções, registra fotografias, mas também àquele que elabora pareceres e análises jurídicas, realiza pesquisas laboratoriais e de campo e provê o acesso à informação de forma equitativa e justa. Se antes era considerado “autor” a pessoa que produzia conteúdo para muitos, atualmente muitos podem produzir conteúdo para muitos, sobretudo devido à ubiquidade das Tecnologias de Informação e Comunicação (TICs). Desta forma, as possibilidades de autoria ampliam-se e, com ela, a própria definição do conceito de “autor”. Surgem, assim, novos “autores”, em distintas áreas, porém, nem sempre em plena concordância sobre o que torna um membro de uma determinada área “autor”.

No momento de produção de um conteúdo, há entendimentos diversos sobre a função de “autor” entre as áreas do conhecimento. No Direito, por exemplo, a Lei de Direitos Autorais brasileira (Brasil, 1998) define “autor” apenas como pessoa física. A Ley de Propiedad Intelectual Española (España, 1996) define autor como pessoa “natural” e também abre a possibilidade de pessoas jurídicas. Na Biblioteconomia, o “autor” foi considerado, a partir do século XIX, no contexto da

Abstract

Due mainly to the Information and Communication Technologies (ICT), the concept of “author” and possibilities of authorship have been broadened, regardless of the area in which the content is created. From the perspective of Wittgenstein’s “language games” and Foucault’s concept of “author,” we analyze the uses, reuses and resignifications attributed to the concept of “author” in the areas of Law, Engineering and, Library Science. We found a non-univocity of the concept based on its transformations and changes over time according to the rules and conventions of each area.

Keywords: Author. Authorship. Foucault, Michel. Wittgenstein, Ludwig.

elaboração de instrumentos bibliográficos, tanto como o responsável pela criação de bens intelectuais como uma entrada para a recuperação de documentos. Na Engenharia, por sua vez, a discussão desse conceito recai sobre diferentes aspectos, como o papel dos pesquisadores nas diversas fases de uma pesquisa e sobre a “contribuição essencial” que permite atribuir a um pesquisador a função de “autor”.

A justificativa do estudo recai na medida em que demonstra como três áreas distintas do conhecimento, Direito, Engenharia e Biblioteconomia, articulam-se na definição de regras e como a participação de suas respectivas comunidades limita a função de “autor” internamente. Demonstra, ainda, como a organização própria de cada área delimita a definição do conceito de “autor”, com destaque para suas práticas e os papéis desempenhados pelas comunidades discursivas. Por meio do apontamento de algumas de suas diferenças, sinaliza que não há um consenso universal sobre esse conceito, pois quando não há regras explícitas para defini-lo, práticas convencionadas entre os próprios membros cumprem essa função normalizadora.

Para compreender as condições que levam a construção de diferentes concepções sobre o conceito de “autor”, serão utilizadas premissas

da filosofia pragmática da linguagem, especificamente a defendida por Ludwig Wittgenstein (1889-1953) em seu segundo momento filosófico caracterizado pela produção de sua obra *Investigações filosóficas* (1953), apropriando-se especificamente da expressão cunhada pelo filósofo de “jogos de linguagem” de Wittgenstein (2009), que prerroga que o uso e o significado das palavras são dinamicamente construídos, reconstruídos e ressignificados pela própria comunidade usuária. Utilizar-se-á, ainda, o estudo de Foucault (1984) sobre o conceito de “autor”, mostrando que esse não possui uma definição única, haja vista que os conceitos estão expostos a diferentes comunidades, cada qual com seu discurso, em espaços e tempos distintos.

A seleção da área do Direito deve-se ao fato do tema Direito Autoral ser de interesse de pesquisa dos autores, bem como devido às discussões acerca da autoria na esfera digital; a Engenharia, por ser uma das áreas em que a questão da autoria é bastante debatida, ao lado da Computação onde, como consequência de sua complexidade, surgiu o movimento de software livre fundado por Richard Stallman, que revolucionou a concepção legal de autoria por meio do Copyleft; e a Biblioteconomia, por ser a área de formação dos autores e também para ampliar o entendimento sobre o conceito de “autor”, especificamente no contexto da Produção e Organização do Conhecimento.

2. Breve introdução a Wittgenstein e Foucault

As potenciais definições de conceitos dão-se entre os membros de uma comunidade, assim como seus significados e a formulação de suas regras de uso. Tal argumento foi demonstrado por Ludwig Wittgenstein, filósofo austríaco, em um momento de sua carreira que ficou conhecido como “a segunda virada linguística”.

Em um primeiro momento, Wittgenstein, durante seus estudos de Engenharia Mecânica, tomou contato com as obras de Gottlob Frege (1848-1925) e Bertrand Russell (1872-1970), despertando seu interesse em Lógica e, assim, seguindo os pressupostos desses estudiosos em busca de uma linguagem ideal. Esse período foi marcado pela obra intitulada “Tratado lógico-filosófico”, publicada em 1921, onde explica a natureza factual e representativa da linguagem, consoante à predominante filosofia analítica, que propunha tornar a linguagem uma forma de pensamento calculável para tornar o filosofar uma construção lógica.

No entanto, ao longo do tempo Wittgenstein reconheceu que uma linguagem ideal não poderia ser plenamente utilizada em todas as áreas, o que o levou a mudar de uma concepção lógica para

uma concepção pragmática da linguagem. Em seu segundo momento filosófico, passou a entender que a forma como as palavras eram utilizadas e suas regras de uso eram definidas no momento e no contexto de seu uso. Isto o levou a formular o conceito de “jogos de linguagem”, cunhado na obra *Investigações filosóficas* (1953), que marca o reposicionamento de seu pensamento.

“Jogos de linguagem” (*Sprachspiel*), segundo o próprio Wittgenstein (2009, p. 19), é “a totalidade formada pela linguagem e pelas atividades com as quais ela vem entrelaçada”. Saldanha e Gracioso (2014) esclarecem que cada termo possui múltiplos usos e sua aplicabilidade não necessita de elementos em comum, mas seus significados estariam relacionados por semelhanças. Portanto, esse conceito é útil para compreender os usos e as oscilações naturais da significação das palavras em comunidades e contextos específicos, além de permitir compreender as regras de formulação e uso estabelecidas.

Isso posto, destaca-se que o conceito de “autor” foi estudado por Foucault (1984) principalmente, em sua obra “O que é um autor”. O filósofo francês, nascido em 1926, cresceu em um ambiente conservador e estudou na Escola Normal Superior de Paris com os filósofos Jean Hypolitte e Louis Althusser. Em 1948, incentivado por Althusser, ingressou no Partido Comunista, desligando-se em 1953 devido à repressão à homossexualidade e ao antissemitismo. Foucault transitou, principalmente, pela Filosofia e Psicologia, apresentando seus estudos em várias obras, dentre elas, “Doença mental e personalidade” (“*Maladie mentale et personnalité*”, 1954), “Loucura e civilização” (“*Folie et Déraison: Histoire de la folie à l’âge classique*”, 1961), “O nascimento da clínica” (“*Naissance de la clinique: une archéologie du regard médical*”, 1963), “As palavras e as coisas” (“*Les mots et les choses: Une archéologie des sciences humaines*”, 1966), “A arqueologia do saber” (“*L’Archéologie du savoir*”, 1969) e “O que é um autor” (“*Qu’est-ce qu’un auteur?*”, 1969). Esta última trata-se, na realidade, de um ensaio em resposta ao texto de Roland Barthes “A morte do autor” (“*La mort de l’auteur*”, 1968), no qual declara que o autor não cria mais textos a partir de sua imaginação, mas sempre são gerados a partir de textos e ideias preexistentes. Foucault, por sua vez, atribui ao autor uma função discursiva, preocupando-se com a circulação do discurso, além de entendê-lo como uma entidade psicológica e como um termo para designar o conjunto de uma obra.

Diante desta breve introdução às teorias de Wittgenstein e Foucault, em especial a partir da compreensão da significação da expressão “jogos de linguagem”, objetiva-se compreender a

definição de “autor” nas áreas de Direito, Engenharia e Biblioteconomia, conforme mencionadas anteriormente. Para tanto, foram elencados exemplos na literatura que trazem definições e regras sobre o conceito de “autor” a fim de compreender, em cada área, o entendimento sobre o mesmo e como seus membros foram nomeados como “autores”. Posteriormente, o estudo de Foucault (1984) sobre o autor é apresentado para explicar a não univocidade acerca desse conceito. Para melhor compreendê-lo, apresenta-se um breve histórico do Direito Autoral e, em seguida, o entendimento do Direito sobre o mesmo.

3. Origens do Direito Autoral e do conceito de “autor” no Direito

Na Antiguidade Clássica, não havia o reconhecimento dos direitos autorais básicos, que só seriam reconhecidos com a invenção da tipografia e da imprensa. Efetivamente, esses direitos passaram a ser resguardados a partir da promulgação do Copyright Act, em 1710, pela rainha Ana da Inglaterra, visando à exploração econômica das obras e contribuindo para a criação de legislações nacionais nos demais países (Carboni, 2010). Posteriormente, durante o Iluminismo francês, foi criado o modelo *droit d'auteur*, que além da exploração econômica, inovou ao acrescentar o direito moral do autor. A legislação brasileira, consoante o modelo francês, prevê os direitos morais do autor, isto é, o direito de personalidade, e o direito patrimonial, que permite explorar economicamente os bens imateriais. A lei autoral atualmente em vigor no Brasil é a 9.610/98, que altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências (Brasil, 1998).

Na legislação brasileira, “autor” é definido como “a pessoa física criadora de obra literária, artística ou científica” (Brasil, 1998). Na legislação espanhola é definido como “la persona natural que crea alguna obra literaria, artística o científica” (Espanha, 1996). Na Lei espanhola também se indica que:

Se presumirá autor, salvo prueba en contrario, a quien aparezca como tal en la obra, mediante su nombre, firma o signo que lo identifique. [...] Cuando la obra se divulgue en forma anónima o bajo seudónimo o signo, el ejercicio de los derechos de propiedad intelectual corresponderá a la persona natural o jurídica que la saque a la luz con el consentimiento del autor, mientras éste no revele su identidad.

Considerando-se suas datas de promulgação originais, observa-se que as mesmas foram criadas em um contexto no qual as TICs ainda eram incipientes e, por isso, havia pouca possibilidade de criação de conteúdo digitais. Na atualidade, porém, aquelas tecnologias possibilitaram às pessoas tornarem-se autoras de conteúdos diversos.

Dentre as “autorias” surgidas neste contexto, Carboni (2010) aponta a meta-autoria em obras criadas por sistemas computacionais e a pluri-autoria nas criações colaborativas. No primeiro caso, tem-se como exemplo a criação do programa de computador AARON pelo pintor abstracionista Harold Cohen em 1968, por meio do qual o usuário apenas definia o tempo de duração da obra, enquanto que o programa, por meio de uma mão protética, cria as obras espontaneamente a partir dos padrões de imagens de seu criador. A pluri-autoria, por sua vez, é exemplificada com o software livre, que é criado por pessoas do mundo todo e distribuído livremente de forma que qualquer usuário possa acessar seu código fonte e aperfeiçoá-lo.

Ressalta-se, ainda, que embora a lei considere apenas pessoa física como autor, o parágrafo único do artigo 11 da Lei Brasileira discorre que “A proteção concedida ao autor poderá aplicar-se às pessoas jurídicas nos casos previstos nesta Lei” (Brasil, 1998) e na Lei Espanhola, analogamente, “de la protección que esta Ley concede al autor se podrán beneficiar personas jurídicas en los casos expresamente previstos en ella” (Espanha, 1996). Por se tratar de fruto de criação de pessoa física, os direitos relativos aos bens intelectuais podem ser cedidos para pessoas jurídicas, que se tornam titulares dos direitos autorais, conforme a extensão da cessão. No que se refere à Administração Pública Brasileira, ela pode se tornar detentora desses direitos em atividades de fomento à cultura, na contratação de obras intelectuais e na produção de obras intelectuais por meio de seus servidores (Escola Nacional de Administração Pública, 2015a).

Além disso, a lei prevê a existência de coautoria, ou seja, obras que foram criadas conjuntamente por dois ou mais autores. Conforme o conteúdo, as obras em coautoria classificam-se em divisíveis, como uma coletânea de contos ou poemas com textos de dois ou mais autores, e indivisíveis, por exemplo, um livro, no qual não é possível identificar, efetivamente, a contribuição de cada autor. A lei brasileira ressalta, ainda, que o auxílio prestado aos autores na revisão, atualização, fiscalização ou edição não configura coautoria, remetendo à questão da “contribuição efetiva” para a criação de uma obra intelectual, que será discutida posteriormente (Escola Nacional De Administração Pública, 2015b). Na Lei espanhola (Espanha, 1996), a coautoria é regulada no “Artículo 8. Obra colectiva” do Título II “Sujeto, objeto y contenido”, Capítulo I “Sujetos”, que diz:

Se considera obra colectiva la creada por la iniciativa y bajo la coordinación de una persona natural o jurídica que la edita y divulga bajo su nombre y está constituida por la reunión de aportaciones de diferentes autores cuya contribución personal se funde en

una creación única y autónoma, para la cual haya sido concebida sin que sea posible atribuir separadamente a cualquiera de ellos un derecho sobre el conjunto de la obra realizada. Salvo pacto en contrario, los derechos sobre la obra colectiva corresponderán a la persona que la edite y divulgue bajo su nombre.

Assim, na visão do Direito, depreende-se que o conceito de “autor” é regido pela definição da lei. Embora não contemple pessoa jurídica como “autor”, sua proteção está legalmente prevista. Da mesma forma, a menção à coautoria contempla outra modalidade de criação de obras intelectuais. Os limites legais impostos pelas definições visam uniformizar o entendimento sobre o conceito de “autor” para que a comunidade jurídica participe do mesmo “jogo de linguagem”, conforme prevê Wittgenstein (2009), porém, como anteriormente comentado, o conceito encontra-se defasado frente ao contexto tecnológico atual, pois não contempla as diversas formas de autoria possibilitadas pelas TICs, como a meta-autoria e pluri-autoria.

4. O conceito de “autor” na Engenharia

Permeada pela realização de pesquisas, a Engenharia, enquanto área do conhecimento, visa contribuir para a melhoria da qualidade de vida da sociedade por meio do emprego de técnicas que criem soluções para o bem-estar social. Inúmeros são os atores envolvidos e os papéis por eles desempenhados no decorrer de uma pesquisa, o que faz que com o conceito de “autor” torne-se relativamente difuso nessa área, isto é, não há regras explícitas de autoria, mas convenções tacitamente definidas pelos pesquisadores.

A questão da autoria, fortemente relacionada à ética, por vezes é controversa na Engenharia, seja pelo fato dessa área do conhecimento estar cada vez mais desenvolvendo pesquisas conjuntas com outras áreas, portanto, defrontando-se com interesses diversos, seja pelo fato de haver certo ceticismo entre a indústria e os pesquisadores, quando realizam projetos conjuntos (Borenstein, 2011). Em ambas as situações, há limites e entendimentos nebulosos que, somados aos diferentes interesses, implicam em interpretações distintas sobre aspectos relacionados à ética, assim como em relação ao conceito de “autor”.

No que se refere à ética, as interpretações podem recair sobre o desenvolvimento de uma determinada tecnologia ou sobre a melhoria de um processo ou produto existente, que podem culminar em uma patente ou em um modelo de utilidade, por exemplo. Para tanto, os autores devem discriminar precisamente nesses documentos suas criações para justificar a concessão da ex-

ploração respeitando, assim, a propriedade intelectual alheia. Em relação à autoria, Borenstein (2011) comenta que há uma carência de padronização sobre a determinação da ordem dos autores nas publicações de Engenharia. Nos diversos campos há maneiras diferentes de padronização: enquanto algumas publicações indicam como último autor o diretor de um laboratório como o pesquisador sênior, outras indicam o estudante de pós-graduação como o primeiro. Além disso, nem sempre a indicação da ordem de autoria corresponde ao grau de contribuição dos autores, como é o caso de uma ordenação alfabética, pois esta é apenas mais uma convenção utilizada na Engenharia.

A existência de diferentes convenções não é prejudicial, haja vista que reconhecidas publicações científicas e laboratórios devem ter liberdade para definir suas próprias regras, mas também devido essa diferença torna-se difícil relacionar o nível de contribuição dos autores (Borenstein, 2011). O importante é que haja um entendimento comum sobre o significado do conceito e sua aplicabilidade entre seus membros, a fim de que não haja distorções no próprio discurso da comunidade que o utiliza, em consonância com a definição de “jogos de linguagem” de Wittgenstein (2009).

Entre as diferentes orientações para autores nas publicações de Engenharia, há um ponto comum que diz respeito à “contribuição significativa” realizada pelos participantes de uma pesquisa, ainda que não haja concordância na sua definição. Borenstein (2011) também discute se é honesto incluir os indivíduos que coletaram dados e os lançaram em bases de dados como “autores”, isto é, se o desempenho dessas funções permite serem considerados como tais.

Além das questões apontadas, tem sido cada vez mais comum a realização de intercâmbios por estudantes e a consequente realização de pesquisas conjuntas. Embora haja um entendimento internacional comum sobre ética na pesquisa, podem haver diferenças de interpretação entre os países e, conseqüentemente, interesses diversos sobre questões como plágio e reuso de dados de pesquisa. Por isso, é importante que os estudantes observem as regras do periódico do país no qual irão publicar. Além disso, há também a questão da edição de manuscritos por falantes não-nativos, o que pode ocasionar modificações gramaticais, por exemplo, e como há um pressuposto de que os “autores” são aqueles que contribuíram significativamente para o trabalho, o “editor de língua” poderia ser listado na seção de agradecimentos do mesmo, posto que se configura como contribuição intelectual de importância diferenciada (Borenstein, 2011).

Em resumo, percebe-se que a conceito de “autor” na Engenharia por vezes segue regras explícitas, como aquelas nas instruções de publicações da área, e por vezes implícitas, como convenções estabelecidas entre os próprios sujeitos que desenvolveram uma pesquisa. Destaca-se, ainda, a preocupação em se definir um “autor” a partir de sua “contribuição significativa”, isto é, se o papel desempenhado durante a realização de uma pesquisa pode tornar um sujeito “autor”. Por fim, a definição do conceito também recai sobre a realização de intercâmbios e o consequente estabelecimento de parcerias internacionais de publicação e, dessa forma, deve-se atentar para as normas do periódico observando-se o país em que é editado.

5. O conceito de “autor” na Biblioteconomia

Smiraglia, Lee e Olson (2011) concentraram-se em entender o papel da autoria em instrumentos bibliográficos, voltando-se, mas não exclusivamente, para as regras de catalogação. Na Grécia antiga, os autores discorrem sobre os Pinakes redigidos por Calímaco, os quais arrolavam oradores, poetas, filósofos e outros autores e suas obras, enquanto que na China antiga, a preocupação com a autoria é exemplificada com os Sete Epítomes, catálogo criado a partir dos *Separate Résumés*, uma coleção de registros bibliográficos anotados e que distinguiu três tipos de autoria: compositor, transmissor (na forma oral ou escrita) e editor/compilador, funções que se confundiam quanto mais antiga fosse a obra. A análise feita por Smiraglia, Lee e Olson (2011) sugere certa similaridade com a concepção ocidental, isto é, “autor” é aquele que fala ou escreve com autoridade. Além disso, Smiraglia e Lee (2012) comentam que, embora os Sete Epítomes relacionem autores e obras, essa relação não se estende para um princípio organizador, pois autor e título não funcionavam como entradas para recuperar os documentos, mas eram localizados por meio de um sistema de classificação de valores morais dos autores. Por isso, nem sempre as obras eram reunidas na mesma classe, visto que os valores morais determinavam sua classificação.

Posteriormente, a tradição anglo-americana passou a considerar não o aspecto biográfico, mas sim técnico do “autor”. Com a publicação das 91 regras de catalogação para o Museu Britânico em 1841, Panizzi determina que o catálogo seja organizado por ordem alfabética de autor, demonstrando preocupação com o nome e sua respectiva forma (Smiraglia et al., 2011). Foi Panizzi quem também forneceu as primeiras entradas úteis: além do título, considerava informações de

edições, editores, traduções e conteúdo intelectual (Smiraglia e Lee, 2012). Já em finais do mesmo século, Cutter também se preocupava com o nome e sua forma, mas definia o “autor” em um sentido estrito, como aquele que escreve um livro, e em um sentido amplo, como aquele que organiza textos de vários autores. Em meados do século XX, Charles Coffin Jewett apresenta uma nova abordagem sobre a autoria na obra *“On the construction of catalogues of libraries, and their publication by means of separate stereotyped titles: with rules and examples”*, publicada em 1953, pois estende o princípio de autoria e entidades corporativas sob seus nomes, correndo ainda mais a caracterização pessoal do nome do autor. Verifica-se, assim, que nessa tradição, o nome do autor assume uma grande responsabilidade no catálogo, embora não como uma ligação entre o nome próprio e a pessoa que descreve, porém mais como um elemento de designação de uma classificação alfabética de obras. Mais recentemente, o código de catalogação RDA traz uma mudança terminológica do conceito de “autor” em relação ao AACR2, pois propõe o termo “criador” para englobar amplamente pessoas e instituições responsáveis pela concepção de bens intelectuais, enquanto que “autor” fica restrito àquele que cria, principalmente, bens textuais (Smiraglia et al., 2011).

Diante do exposto, vê-se que o conceito de “autor” na Biblioteconomia, especialmente na construção de catálogos, sofreu modificações ao longo do tempo em função das concepções de autoria adotadas pelas diferentes sociedades. Na Grécia Antiga, os autores e suas respectivas obras eram arrolados nos Pinakes, assim como na China Antiga, que ainda utilizava uma classificação baseada nos valores morais, dificultando, relativamente, o agrupamento das obras, mas sem deixar de relacioná-las com o respectivo autor. Já no século XIX, Panizzi e Cutter preocuparam-se com o autor (e outras funções intelectuais) enquanto entrada no catálogo, visando à recuperação dos documentos, ao passo que Jewett estendeu o princípio de autoria e entidades corporativas sob seus nomes no século seguinte e, mais recentemente, o RDA introduziu o conceito de “criador”, distinguindo-o da definição de “autor” adotada pelo AACR2.

Desta forma, compartilha-se a reflexão de Smiraglia e Lee (2012) de que a atribuição é uma forma de classificação cultural por meio da qual o nome de um “autor” torna-se o nome de uma classe de discurso. Portanto, “autor” não representa mais criadores únicos, nem um princípio organizador como entrada para recuperação de documentos, mas estabelece uma associação como identificador de trabalhos relacionados. Por exemplo, o

programa de televisão “The French Chef”, apresentado por Julia Child, deu origem a obras derivadas que tiveram a participação intelectual de várias pessoas, assim, para representá-lo por meio de um sistema culturalmente orientado por meio de identificadores, Smiraglia e Lee (2012) defendem a adoção de “The French Chef” como entrada de um catálogo ao invés de Julia Child, propondo uma revisão do conceito de autoria enquanto princípio ordenador do catálogo. Finalmente, no trabalho de Martínez-Ávila, Smiraglia, Lee e Fox (2015) realizou-se uma análise do discurso foucaultiano dos já citados “Sete Epítomes”, as “Obras de Abelardo” (discutida por Panizzi), e o caso moderno do “The French Chef”, de Julia Child, em ferramentas atuais da web concluindo que a atribuição do autor nos catálogos não concorda com a concepção discursiva de Foucault sobre “autor”.

Para melhor compreender as diferenças identificadas sobre o conceito de “autor”, a próxima seção apresenta os estudos de Foucault (1984) sobre o tema e o tratamento conferido a esse conceito.

6. Foucault e o conceito de “autor”

Olsson (2010) indicava que Foucault era pouco explorado na Biblioteconomia pelo fato da área ter se concentrado no comportamento de busca e nos processos mentais do indivíduo. Na pesquisa e na prática profissional, houve influência do modelo de transferência de informação, segundo o qual os textos eram recebidos pelos leitores passivamente, ao passo que a análise do discurso foucaultiano esclarece que o processo é justamente o inverso: os leitores participam na construção do significado, juntamente com o texto e o contexto social em que estão inseridos. Especialmente na última década, diversos autores e trabalhos aplicaram o trabalho de Foucault e a análise de discurso na Biblioteconomia e na Organização do Conhecimento (Frohmann, 1992, 1994a, 1994b, 2001; Budd e Raber 1996; Budd, 2006; Martínez-Ávila, 2012; Moulaison et al., 2014; Martínez-Ávila e Fox, 2015; Moraes et al., 2016; Castanha et al., 2017; Silva et al., 2017).

Da mesma forma que propõe uma releitura do processo de significação de um texto, Foucault faz o mesmo com o conceito de “autor”, para o qual propõe uma reformulação da relação entre autor, texto e leitor. Em seu artigo “O que é um autor”, o filósofo reduz o “autor” para uma função no discurso, descrevendo-o como “uma categoria ou forma de organização de textos que tem uma história e necessidades para serem desafiadas” (O’Farrell, 2007, tradução nossa), esclarecendo que a entidade física “autor” e “autor” enquanto organização de textos devem ser tratadas

separadamente. Ao discorrer sobre esse conceito, Foucault (1984) considera que as “obras” de um “autor” não se limitam a textos, mas a todo tipo de discurso gerado por ele, sejam rascunhos, obras inacabadas e outros textos.

Diante disso, o filósofo identifica quatro características acerca do “autor”. Os discursos por ele produzidos são objetos de apropriação, cujo surgimento é posterior à atual apropriação penal, pois as obras passaram a ter autores (exceto figuras míticas e sacralizadas) a partir do momento em que seus discursos poderiam ser transgressores. A segunda explica que a função de “autor” não afeta todos os discursos de forma universal e constante: as obras literárias, por exemplo, eram aceitas sem contestação ou atribuição exata de autoria, ao passo que a partir da Idade Média, época em que se inicia o surgimento de obras científicas, o valor de verdade aparece quando o “autor” é explicitamente citado para validar seu discurso como uma declaração verdadeira. A terceira característica expõe que a função de “autor” não se desenvolve espontaneamente como a atribuição de um discurso a um indivíduo, mas é “resultado de operações complexas que constroem uma certa racionalidade que chamamos ‘autor’” (Foucault, 1984, p. 10, tradução nossa). A última, por sua vez, discute que o “autor” é uma fonte de expressão que procura tanto explicar a presença de determinados eventos em suas obras, mas também transformações, distorções e modificações sofridas, além de neutralizar contradições que podem surgir nas mesmas, a fim de organizá-las em torno de uma contradição fundamental ou de origem (Foucault, 1984).

Por meio dessas características, constata-se que o poder de um “autor” se estende para além da criação de bens intelectuais: seu nome não é apenas um elemento no discurso, mas “executa um papel no que se refere ao discurso narrativo, assegurando uma função classificatória” (Foucault, 1984, p. 107, tradução nossa) e possibilita reunir, diferenciar, contrastar e estabelecer relações entre obras.

Como Foucault considera como “obra” toda a produção de um “autor”, isto é, rascunhos, esboços e outros textos, geralmente por estarem dispersas no tempo e no espaço, seu agrupamento com o respectivo “autor” é dificultado, assim como a identificação da responsabilidade intelectual. Diante dessa “ausência autoral”, por vezes motivada por dificuldades inerentes à complexidade da produção intelectual de um conjunto de obras, por outras motivadas pelos interesses próprios das áreas do conhecimento estudadas, o que se percebe ao longo do tempo, mais do que um esforço para resguardar os bens intelectuais gerados pela sociedade, é atribuir às obras, da

melhor forma possível e a partir das possibilidades de cada época, a autoria que lhe é devida.

7. Considerações sobre o conceito de “autor” identificado nas três áreas

A partir da compreensão das concepções de “autor” nas três áreas analisadas (Direito, Engenharia e Biblioteconomia), constata-se o estabelecimento de diferentes “jogos de linguagem” para a construção do significado do conceito, assim como a definição de regras de uso e aplicação próprias de cada uma. No Direito, inicialmente os editores foram protegidos pelo Copyright Act, prevalecendo os direitos patrimoniais, e posteriormente o Iluminismo acrescentou os direitos morais. Atualmente, o conceito é influenciado pelas TICs e inclui a pessoa jurídica, guardadas as devidas proporções, como “autora”. Por isso, entende-se como natural as modificações acerca do conceito de “autor”, haja vista que as mudanças sociais implicam em mudanças legais. Na Engenharia, por sua vez, o conceito de autor não possui a mesma normalização que no Direito, pois constatou-se que essa área se preocupa predominantemente com a técnica, o que leva os pesquisadores a transitarem da autoria de produtos e processos à autoria acadêmica, quando da comunicação dos resultados alcançados. Embora sociedades científicas e periódicos acadêmicos de Engenharia publiquem manuais e regras para guiar os autores no processo de escrita, a definição de “autor” nem sempre segue regras explícitas, seja porque não constam naqueles documentos, seja porque ficam sujeitas ao consenso entre os pesquisadores que desenvolveram o trabalho, por exemplo, o nível de participação na pesquisa ou o cargo que o pesquisador ocupa em um laboratório. Na Biblioteconomia, constatam-se diferentes concepções de “autor” nas sociedades grega e chinesa antigas, que se preocupavam em arrolar autores e obras e listá-los, também, segundo os valores morais, respectivamente; no século XIX, com Panizzi e Cutter, que retiraram o caráter biográfico e atribuíram um caráter técnico ao “autor”, enquanto entrada para recuperação de documentos; e no século XX, com Jewett aplicando o princípio da autoria a entidades corporativas, sendo a concepção mais recente aquela apresentada pelo RDA, que expõe o conceito de “criador”, reservando ao “autor” a responsabilidade pela criação de conteúdos textuais.

Guardadas as devidas diferenças entre as áreas, percebe-se que o conceito de “jogos de linguagem” de Wittgenstein (2009) contribuiu para esclarecer que, de fato, cada comunidade utiliza a linguagem para os fins que atendam suas necessidades, e além disso, nota-se que ela serviu aos seus interesses em tempos específicos, como

exemplificado no Direito e na Biblioteconomia. Desta forma, verifica-se a existência de não univocidade do conceito de “autor”, tanto dentro de uma mesma área como em diferentes momentos de sua evolução.

8. Conclusões

A partir da noção de “jogos de linguagem” de Wittgenstein (2009) evidenciou-se que as regras e acordos estabelecidos pelas áreas visaram o entendimento acerca do conceito de “autor” para atender seus respectivos interesses, mas o que se verifica é que, durante o desenvolvimento de cada uma, o próprio conceito sofreu alterações para atender aos presentes e futuros interesses que surgiriam. Por isso, a partir de Foucault (1984) constata-se que não há univocidade no conceito de “autor”, pois segundo o filósofo não podem ser definidas categorias universais para a definição de conceitos e, portanto, não possuem essência: tratam-se de noções restritivas, isto é, atendem a interesses específicos de um espaço e tempo definidos.

Ressalta-se, entretanto, que essa não univocidade é decorrente das diferenças internas de cada área acerca da formulação do significado do conceito. Isso posto, o que se verificou no Direito foi um entendimento do “autor” pessoa física ou “natural”, com proteção à pessoa jurídica dentro dos limites legais possíveis, além da influência das TICs no que se refere ao surgimento da meta-autoria e pluri-autoria. Na Engenharia, constataram-se entendimentos implícitos sobre a autoria a partir de “contribuições significativas” dos pesquisadores e também das funções que exercem no desenvolvimento de pesquisas. Na Biblioteconomia, por fim, a consolidação do conceito de “autor” passou de um caráter biográfico e subjetivo para um caráter técnico, sendo que desde então serviu como entrada nos catálogos para a recuperação da informação.

Se por um lado as diferenças existentes entre o conceito apontam para concepções particulares da produção de bens intelectuais, por outro apresentam uma preocupação comum, qual seja, a atribuição de crédito aos responsáveis por tais bens. Embora Foucault (1984) demonstre pouca preocupação com quem fala, posto que em sua visão o “autor” é uma função discursiva, há que se considerar que se alguém fala, é por meio de alguma obra, independentemente de sua forma de expressão.

Assim, não se pode falar em “autor” se não se falar em “obra”. As diversas formas de relação entre esses dois elementos propostas pelas áreas estudadas foram bem-sucedidas em suas

respectivas épocas, atendendo às regras dos “jogos de linguagem” dos grupos que as formularam. Os autores de cada momento histórico serviram e continuam servindo aos propósitos de suas áreas do conhecimento, por mais que os critérios para defini-los tenham sofrido modificações ao longo do tempo, ora sendo chamados de editores, ora de compiladores, dentre outras funções intelectuais possíveis. No contexto atual, fortemente influenciado pelas TICs, há que se estar preparado para o surgimento de funções intelectuais vindouras e produtoras de discursos ainda desconhecidos. De fato, Foucault (1984) talvez estivesse à frente do seu tempo ao atribuir ao autor uma função discursiva, preocupando-se com a circulação de seu discurso, à apropriação pelas pessoas e com os lugares em que circula, pois pela forma e velocidade com que são produzidos atualmente, o discurso pode vir a se sobrepor ao autor, deixando-o em plano inferior e, realmente, pouco importando quem fala, mas o que, onde, para quem e como se fala.

Referências

- Borenstein, Jason (2011). Responsible authorship in engineering fields: an overview of current ethical challenges. // *Science and Engineering Ethics*. 17 (2011) 355-364.
- Brasil (1988). Lei nº 9.610/98, de 19 de fevereiro de 1998. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 20 fev. 1998. http://www.planalto.gov.br/CCIVIL_03/leis/L9610.htm (07-09-2015).
- Budd, John (2006). Discourse analysis and the study of communication in LIS // *Library Trends*. 55:1 (2006) 65-82.
- Budd, John; Raber, Douglas (1996). Discourse analysis: method and application in the study of information. // *Information Processing & Management*. 32:1 (1996) 217-226.
- Carboni, Guilherme (2010). Direito autoral e autoria colaborativa na economia da informação em rede. São Paulo: Quartier Latin, 2010.
- Castanha, Renata Cristina Gutierrez; Lima, Larissa de Mello; Martínez-Ávila, Daniel (2017). Análise do discurso sob a perspectiva bibliométrica nos estudos de Ciência da Informação no Brasil. // *Perspectivas em Ciência da Informação*. 22:1 (2017) 17-37.
- Escola Nacional de Administração Pública (2015a). Noções gerais de direitos autorais: módulo 1: obras intelectuais. Brasília: [s.n.], 2015. http://repositorio.enap.gov.br/bitstream/handle/1/1852/M%F3dulo_1_DIREITOS_AUTO-RAIS.pdf?sequence=1 (2016-01-07).
- Escola Nacional de Administração Pública (2015b). Noções gerais de direitos autorais: módulo 2: direitos do autor. Brasília: [s.n.], 2015. <http://repositorio.enap.gov.br/handle/1/1851> (2016-01-07).
- Espanha (1996). Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia. Madrid: Ministerio de Cultura, 1996. <https://www.boe.es/buscar/act.php?id=BOE-A-1996-8930> (2017-03-31).
- Foucault, Michel. (1984). What is an author? // Rabinow, Paul. (Ed.). *The Foucault reader*. New York: Pantheon Books, 1984. 101-120.
- Frohmann, Bernd (1992). The power of images: a discourse analysis of the cognitive viewpoint. // *Journal of Documentation*. 48:4 (1992) 365-386.
- Frohmann, Bernd (1994a). Discourse analysis as a research method in library and information science. // *Library and Information Science Research*. 16:2 (1994) 119-138.
- Frohmann, Bernd (1994b). Communication technologies and the politics of postmodern information science. // *Canadian Journal of Information and Library Science*. 19:2 (1994) 1-22.
- Frohmann, Bernd (2001). Discourse and documentation: some implications for pedagogy and research. // *Journal of Education for Library & Information Science*. 42:1 (2001) 13-26.
- Martínez-Ávila, Daniel (2012). Problems and characteristics of Foucauldian discourse analysis as a research method. // Smiraglia, Richard P.; Lee, Hur-Li. (Eds). *Cultural Frames of Knowledge*. Ergon: Würzburg, 2012. 99-110.
- Martínez-Ávila, Daniel; Fox, Melodie J. (2015). The construction of ontology: a discourse analysis. // Smiraglia, Richard P.; Lee, Hur-Li. (Eds). *Ontology for Knowledge Organization*. Ergon: Würzburg, 2015. 13-37.
- Martínez-Ávila, Daniel; Smiraglia, Richard P.; Lee, Hur-Li; Fox, Melodie (2015). What is an author now? Discourse analysis applied to the idea of an author. // *Journal of Documentation*. 71:5 (2015) 1094-1114.
- Moraes, João Batista Ernesto; Lima, Larissa Mello; Caprioli, Mariana Silva (2016). Análise do discurso e ciência da informação: aportes teóricos para organização e representação da Informação. // *Scire*. 22:2 (jul.-dic. 2016) 75-85.
- Moulaison, Heather Lea; Dykas, Felicity; Budd, John M. (2014). Foucault, the author, and intellectual debt: capturing the author-function through attributes, relationships, and events in knowledge organization systems. // *Knowledge Organization* 41:1 (2014) 30-43.
- O'Farrell, Clare (2007). Key concepts: author. <http://www.michel-foucault.com/concepts/index.html> (2016-07-21).
- Olsson, Michael R. (2010). Michel Foucault: discourse, power/knowledge, and the battle for truth. // Leckie, Gloria J.; Given, Lisa M.; Buschman, John. E. (Eds.). *Critical theory for Library and Information Science: exploring the social from across the discipline*. Santa Barbara: Libraries Unlimited, 2010. 63-74.
- Saldanha, Gustavo S.; Gracioso, Luciana S. (2014). Filosofia da linguagem e Ciência da Informação na América Latina: apontamentos sobre pragmática e linguagem. // *Rendón Rojas, Miguel Ángel (Org.). El Problema del lenguaje en la Bibliotecología/Ciencia de la Información/Documentación: un acercamiento filosófico-teórico*. [S.l.]: UNAM. 1-33.
- Silva, Rafaela Carolina da; Lima, Larissa de Mello; Suenaga, Cynthia Maria Kiyonaga; Martínez-Ávila, Daniel; Moraes, João Batista Ernesto. Hybrid libraries: a discourse analysis of the Central Library at Liverpool. // *Ibersid*. 11:1. (2017).
- Smiraglia, Richard P.; Lee, Hur-Li (2012). Rethinking the authorship principle. // *Library Trends*. 61:1 (2012) 35-48.
- Smiraglia, Richard P.; Lee, Hur-Li; Olson, Hope A. (2011). Epistemic presumptions of authorship. // *iConference 2011, Inspiration, Integrity, and Intrepidity*, ACM, New York, February 8-11. New York: ACM, 2011. 137-143.
- Wittgenstein, Ludwig (2009). *Investigações filosóficas*. 6. ed. Petrópolis: Vozes, 2009.

Enviado: 2017-04-01. Segunda versão: 2017-05-25.
Aceptado: 2017-09-14.

Elementos da teoria semiótica da percepção de Peirce para o desenvolvimento de interfaces de recuperação da informação

Elementos de la teoría semiótica de la percepción de Peirce para el desarrollo de interfaces de recuperación de información

Elements of Peirce's semiotic theory of perception for the development of information retrieval interfaces

Fabricio BAPTISTA (1), Carlos Cândido de ALMEIDA (2), Edberto FERNEDA (2)

Instituto Federal do Paraná, Campus Jacarezinho, Avda. Dr. Tito, s/n, PR, (Brasil), fabriciobri@gmail.com

(2) Departamento de Ciência da Informação, Universidade Estadual Paulista, Av. Hygino Muzzi Filho, 737, Marília, SP (Brasil), carlosalmeida@marilia.unesp.br, ferneda@marilia.unesp.br

Resumen

Se exploran los conceptos básicos de la percepción en Charles Sanders Peirce relacionándolos con los procesos perceptivos que afectan al diseño de interfaces digitales en el campo de la recuperación de información. Se pretenden dar a conocer las bases conceptuales de la semiótica peirceana, junto a otras teorías relacionadas con el proceso perceptivo, como bases teóricas de una Ingeniería Semiótica. Los resultados de la investigación indican que es adecuado considerar los elementos del proceso de percepción recogidos por Peirce y los tipos de mediación (icónica, indicial y simbólica) en la interacción hombre-ordenador. Sintéticamente, se puede observar que la formación del pensamiento comienza por las puertas de la percepción, haciendo relevante su estudio para el desarrollo de interfaces de ordenador, especialmente en el ámbito de la recuperación de información.

Palabras clave: Percepción. Teoría de la percepción. Peirce, Charles Sanders. Interfaces de recuperación de información. Semiótica.

1. Introdução

Uma das formas de se introduzir um tema é a partir de uma definição, porém a palavra percepção, conforme discorreremos durante este artigo, não se refere a um evento isolado. Existem diversas teorias que abordam o tema, como a Empirista, a Intelectualista, a Psicologia da forma (ou teoria da Gestalt) e a Fenomenologia, em que a percepção é descrita através de um processo integrado, e pode ser considerada fundamental para compreender o conhecimento humano. Desta forma, entende-se que a percepção é “a ponte entre o mundo da linguagem - a consciência, o cérebro, a mente - e o mundo lá fora” (Santaella, 2012b, p. 75).

Neste sentido, alguns aspectos da percepção estão relacionados à organização do conhecimento. Para Barros e Café (2012) os processos

Abstract

This article explores the basic concepts of perception in Charles Sanders Peirce relating them to perceptual processes that affect the design of digital interfaces in the field of information retrieval. For this, it was tried to present the conceptual bases of peircean semiotics, along with other theories related to the perceptive process, as theoretical bases of a Semiotic Engineering. The results of the research indicate that it is appropriate to consider the elements of the process of perception collected by Peirce and the types of mediation (iconic, indicial and symbolic) in the human-computer interaction. Synthesizing, it can be observed that the formation of the thought begins at the doors of perception, making relevant its study for the development of computer interfaces, especially in the scope of information retrieval.

Keywords: Perception. Theory of perception. Information retrieval interfaces. Semiotics. Peirce, Charles Sanders.

de significação são importantes aspectos da Ciência da Informação (CI) no que diz respeito à organização e representação do conhecimento, uma vez que a CI desenvolve métodos para interpretar diferentes constituições textuais. Para as autoras esse processo de significação também é vislumbrado pela Semiótica, ciência dos signos produzidos nos âmbitos da natureza e da cultura (Santaella, 1983 apud Nöth, 2003). O processo de percepção pode ser explicado pela arquitetura filosófica de Charles Sanders Peirce (1839-1914). Assim, surgiu a seguinte questão teórica, quais aspectos da semiótica peirceana contribuem para compreensão do processo de percepção humana? Em consequência, quais elementos são importantes para a recuperação da informação ensejada pela Ciência da Informação? Nesse sentido, trabalhou-se com base em dois argumentos. Presume-se que a semiótica peirceana explica o processo perceptivo com

base em uma rede de semiose, contribuindo para o entendimento do processo de percepção. Além disso, outro ponto, é que os fenômenos se apresentam por meio de signos que são percebidos e interpretados, e podem influenciar o processo de percepção no instante da recuperação da informação.

Para dar suporte a essas argumentações delineou-se alguns objetivos que nortearam a estruturação deste trabalho. Como objetivo geral, apresenta-se uma revisão conceitual dos elementos da teoria da percepção segundo a semiótica peirceana que podem contribuir para o entendimento dos fatores intervenientes da recuperação da informação em interfaces digitais. No que tange aos objetivos específicos, procura-se contextualizar a teoria da percepção de Peirce junto a outras teorias que envolvem a percepção, além de examinar as aplicações dessas teorias na área da Ciência da Informação. Conforme explica Santaella (2012b), o entendimento da teoria da percepção de Peirce está intimamente ligada aos conceitos da sua arquitetura filosófica.

2. Semiótica Peirceana

As bases conceituais na elaboração deste trabalho foram a fenomenologia e a semiótica peirceana. Com isso, faz-se essencial a introdução desses temas, ainda que sucinta, para o desenvolvimento do assunto. Dentre os diversos conceitos que compõem essas áreas, objetiva-se realçar os elementos que servem como base para a teoria da percepção de Peirce.

Diferentes autores exploram a semiótica sob diferentes perspectivas. Em a “Semiótica do Século XX”, Nöth (1996) abordou alguns desses teóricos, como: Saussure, Hjelmslev, Jakobson, Barthes, Eco, entre outros. Embora não tenha abordado Peirce junto aos teóricos do século XX, ele reconhece a importância dos paradigmas peirceanos, os quais foram tratados em uma obra anterior do autor. Apesar de existirem diversas abordagens, limita-se à definição de semiótica empregada por Santaella (1983 apud NÖTH, 2003, p. 17), que se alinha com o pensamento peirceano. Neste ponto de vista, a “semiótica é a ciência dos signos e dos processos significativos (semiose) na natureza e na cultura”.

Na arquitetura filosófica de Charles Sanders Peirce há um caminho importante que liga a fenomenologia ao pragmatismo, percorrendo as ciências estética, ética, lógica ou semiótica e metafísica. Conforme descrito por Santaella (2012a), apesar da fenomenologia se desenvolver independente, ela serviu como alicerce para as ciências normativas. Na visão da autora a preocupação fenomenológica constituiu-se na base de toda filosofia de

Peirce. A fenomenologia teria como objetivo postular formas e propriedades universais observando os fenômenos em qualquer área da ciência, a palavra fenômeno deriva de *phaneron*, ou seja, qualquer coisa que aparece de qualquer modo à mente, sendo que, na concepção de Peirce, os fenômenos não possuem nenhuma moldura preestabelecida, e ainda (Santaella, 1995, p. 16):

Fenômeno é qualquer coisa que aparece à mente, seja ela meramente sonhada, imaginada, concebida, vislumbrada, alucinada... Um devaneio, um cheiro, uma ideia geral e abstrata da ciência... Enfim, qualquer coisa.

Nöth (2003, p. 63-64) explicou que na busca por definir categorias que servissem de modelo para conter a multiplicidade de fenômenos do mundo, Peirce desenvolveu uma fenomenologia recorrendo a três categorias, as quais denominou *Firstness*, *Secondness* e *Thirdness*, que podem traduzidas por primeiridade, secundidade e terceiridade:

Primeiridade é a categoria das coisas, sem nenhuma relação com outros fenômenos do mundo [...]. Secundidade começa quando um fenômeno primeiro é relacionado a um segundo fenômeno qualquer [...]. Terceiridade é a categoria que relaciona um fenômeno segundo a um terceiro [...]. A base do signo é, portanto uma relação triádica desses três elementos.

Reforçando esta visão, “Primeiridade é o começo, aquilo que tem frescor, é original, espontâneo, livre. Secundidade é aquilo que é determinado, terminado, final, correlativo, objeto, necessitado, reativo. Terceiridade é o meio, devir, desenvolvimento” (Santaella, 2012b, p.79).

A noção de signo corresponde à terceira das categorias fenomenológicas. Peirce (2005, p. 46) sintetiza que

[...] um signo, ou *representâmen*, é aquilo que, sob certo aspecto ou modo, representa algo para alguém. [...] Ao signo assim criado denomino interpretante do primeiro signo. O signo representa alguma coisa, seu objeto.

Além das categorias fundamentais de Peirce e da noção de signo, outro elemento relevante ao estudo da percepção é a concepção de objeto (objeto dinâmico e imediato). Nöth (2003, p.67) fundamenta o conceito de objeto na obra de Peirce, salientando que ele pode ser “uma coisa material do mundo”, do qual temos um “conhecimento perceptivo”, mas também pode ser entidade meramente mental ou imaginária. Além disso, Peirce descreveu dois tipos de objeto: objeto dinâmico e objeto imediato. Santaella (2015, p. 34) esclarece que objeto imediato denota um objeto dinâmico e, portanto:

[...] o melhor caminho para começar a análise da relação objetual é o do objeto imediato. Afinal, parece não haver outro modo de começar, visto que o objeto dinâmico só se faz presente, imediatamente, via objeto imediato, este interno ao signo.

A divisão dos objetos do signo em dinâmico e imediato mostra que, com o objeto dinâmico, Peirce (1994, CP 5.212) identificou aquilo que está fora da cadeia signica, aquilo que algumas vezes ele chamou de “real” ou “realidade”, mas que pode ser também fictício. Esse esclarecimento será importante para a elucidação dos conceitos de percepto e *percipiuum*, os quais serão abordados adiante.

Nöth (2003, p.76) explica que “Peirce desenvolveu uma tipologia elaborada de signos com base em uma classificação do representâmen, objeto e interpretante, cada uma em três classes denominadas tricotomias”. Almeida (2009, p.246) destaca as três tricotomias do signo mais desenvolvidas:

A primeira tricotomia trata da relação do signo com ele mesmo; a segunda tricotomia propõe signos resultantes da relação do signo com o objeto; e por fim, a terceira tricotomia especifica os signos produzidos no seio da conexão signo e interpretante.

Sabendo-se que o termo tricotomia se refere à divisão dos elementos em três partes, cada uma das tricotomias listadas possuem termos para designar esses elementos, os quais são relacionados às categorias universais (primeiridade, secundidade e terceiridade). Mesmo reconhecendo a importância das três tricotomias, especificamente neste artigo, será evidenciada a segunda tricotomia (relação do signo com o objeto), a qual servirá como apoio à reflexão sobre percepção. Outro fator importante, destacado por Nöth (2003, p. 78), é que Peirce considera a segunda tricotomia como “a divisão mais importante dos signos”. Os elementos que compõem esta divisão são o ícone, o índice e o símbolo. A seguir remete-se a definição de Peirce (2005, p. 73):

Uma progressão regular de um, dois, três pode ser observada nas três ordens de signos, Ícone, Índice e Símbolo. O Ícone não tem conexão dinâmica alguma com o objeto que representa; simplesmente acontece que suas qualidades se assemelham às do objeto e excitam sensações análogas na mente para a qual é uma semelhança. Mas, na verdade, não mantém conexão com elas. O Índice está fisicamente conectado com seu objeto; formam, ambos, um par orgânico, porém a mente interpretante nada tem a ver com essa conexão, exceto o fato de registrá-la, depois de ser estabelecida. O Símbolo está conectado ao seu objeto por força da ideia da mente-que-usa-o-símbolo, sem a qual essa conexão não existiria.

Nesse sentido, por ícone entende-se que ele representa algo através de uma relação de analo-

gia, ou seja, apresenta características que se assemelham com o objeto. Exemplificando, em interfaces digitais de recuperação de informação usualmente utiliza-se imagens como por exemplo a de um teclado de computador para representação do próprio objeto, entretanto ressalta-se que as semelhanças não são necessariamente visuais, um som ou um perfume poderiam ser ícones em potencial. Já o índice, por sua vez, é um indicador de um objeto, é um signo que através de indícios (causa) remete ao que está sendo representado induzindo uma reação a quem o vê. Por fim, o símbolo apresenta uma relação de convenção com o objeto que representa, geralmente é estabelecido de forma arbitrária, como por exemplo uma palavra utilizada para representar um objeto qualquer.

Neste momento é pertinente salientar que a progressividade de atuação dos três elementos (ícone, índice e símbolo) dá suporte a perspectivas no que diz respeito à análise do julgamento perceptivo (terceiro elemento do modelo triádico), entretanto esta relação será apresentada nas próximas seções.

O último componente da arquitetura filosófica de Peirce a ser abordado é a lógica crítica, que também faz relação com o julgamento perceptivo através das inferências abduativas que são derivadas do conceito de abdução. Antes de Peirce, somente a indução e a dedução eram consideradas tipos ou métodos de raciocínio, ele criou o termo abdução. Peirce classificou esses tipos dentro de suas categorias universais, onde a abdução estaria para a primeiridade, como a indução para a secundidade e a dedução em nível de terceiridade. Ibri (2006, p.93) utiliza o próprio Peirce para explicar “a abdução como o processo para formar hipóteses explicativas. É a única operação lógica a introduzir ideias novas; pois que a indução não faz mais que determinar um valor, e a dedução envolve apenas as consequências necessárias de uma pura hipótese” (Peirce, 1994, CP 5.171, p. 1736).

A semiótica peirceana evidencia alguns elementos essenciais da teoria da percepção. A seguir será exposta a teoria da percepção de Peirce e as possíveis relações com o desenvolvimento de interfaces digitais.

3. Teoria da percepção em Peirce e as contribuições para o desenvolvimento de interfaces Web

O objetivo inicial desta seção é apresentar os tópicos centrais da teoria da percepção em Peirce. Em um segundo momento, será relacionada a abordagem da percepção em Peirce com o desenvolvimento de interfaces Web. Contudo, inicia-se

por tratar o assunto de maneira geral. As pesquisas sobre percepção têm colaborado para ampliar a compreensão de cognição, no sentido de analisar como se realiza o fenômeno “conhecer”. Santaella (2012b, p. 1) inicia sua obra sobre percepção, enfatizando o crescente interesse sobre o tema, sobretudo a partir do século XIX, com o impulso das ciências cognitivas. Essas pesquisas estão inseridas em diferentes abordagens que permeiam as diversas áreas de conhecimento.

A maioria das pesquisas sobre percepção apresentou alguma ligação com o sistema visual humano, tais como: percepção visual, visualização de informação e ensino (Correia, 2001; Loomis, 2003; Santos, 2003; Alexandre e Tavares, 2007; Sousa, 2009; Andrade et al., 2012; Jorge; Rezende e Wartha, 2013; Matsunaga, 2015).

Também depara-se com pesquisas que apresentam análises individuais das teorias da percepção de Peirce, Merleau-Ponty e Gödel (Frankenthal, 2004; Nobrega, 2008; Schultz, 2012). Rego, Pessoa e Gala (2015) discutiram as diferenças da teoria da percepção de Bertand Russell, Charles Sanders Peirce e Alberto Caeiro. Moraes, Pereira e Pantaleão (2012) analisaram as correlações da teoria peirceana para a abordagem ecológica através da aproximação entre as perspectivas ecológica e pragmatista para verificar em que medida o pragmatismo peirceano contribuiria para o estudo da percepção. Os autores relataram que a partir dos pressupostos do realismo direto, percepção direta (em certo grau) e abordagem sistêmica é possível a aproximação entre as abordagens classificando como um novo

campo conceitual para a compreensão da percepção segundo a perspectiva direta. A única ressalva é quanto ao julgamento perceptivo (elemento essencial à percepção, segundo Peirce), pois “extrapola o escopo da abordagem ecológica, uma vez que engloba regras e hábitos, que podem ser simultâneos à percepção, que não são explicitados pela abordagem ecológica” (Moraes, Pereira e Pantaleão, 2012, p. 203).

Algumas produções científicas exemplificaram a aplicação da teoria da percepção de Peirce em espaços geográficos (Rocha, 2002; Mucelin e Bellini, 2008). Enfim, estes exemplos proporcionam uma visão mais ampla sobre o conceito de percepção e seus potenciais de aplicação. Além disso, ratificaram a importância dos comentadores e teóricos da semiótica e da percepção.

Nos diversos temas apresentados, invariavelmente, surgiram questões como: “O que é percepção?”, e “Qual o significado de percepção?” uma das primeiras ideias seria a busca do termo em um dicionário da língua. Entretanto, quando se realiza um estudo um pouco mais minucioso, é possível constatar que este tipo de significação é bem restrito, já que sua definição possui diferentes conceitos, a depender da abordagem teórica. As abordagens sobre percepção envolvem uma variedade de compreensões, que nem sempre dirigem-se para um consenso. A seguir, será apresentado um panorama dessas abordagens.

Para prosseguir com a inquirição sobre o conceito de percepção é importante confrontá-lo ao conceito de sensação, as teorias que discutem essa relação são apresentadas no Quadro 1.

| <i>Teorias</i> | <i>Concepções</i> |
|-------------------------------------|---|
| Empirista | A sensação e a percepção dependem das coisas exteriores; A sensação seria pontual, isto é, um ponto do objeto externo toca um de meus órgãos dos sentidos e faz um percurso no interior do meu corpo, indo ao cérebro e voltando às extremidades sensoriais; Cada sensação é independente das outras e cabe à percepção unificá-las e organizá-las numa síntese; A causa do conhecimento sensível é a coisa externa, de modo que a sensação e a percepção são efeitos passivos de uma atividade dos corpos exteriores sobre o nosso corpo; O conhecimento é obtido por soma e associação das sensações na percepção e tal soma e associação dependem da frequência, da repetição e da sucessão dos estímulos externos e de nossos hábitos |
| Intelectualista | A sensação e a percepção dependem do sujeito do conhecimento e a coisa exterior é apenas a ocasião para que tenhamos a sensação ou a percepção; Sentir e perceber são fenômenos que dependem da capacidade do sujeito para decompor um objeto em suas qualidades simples (a sensação) e de recompor o objeto como um todo, dando-lhe organização e interpretação (a percepção); A passagem da sensação para a percepção é, neste caso, um ato realizado pelo intelecto do sujeito do conhecimento, que confere organização e sentido às sensações; Não haveria algo propriamente chamado percepção, mas sensações dispersas ou elementares; sua organização ou síntese seria feita pela inteligência e receberia o nome de percepção |
| Psicologia da forma e fenomenologia | Contra o empirismo, que a sensação não é reflexo pontual ou uma resposta físico-fisiológica a um estímulo externo também pontual; Contra o intelectualismo, que a percepção não é uma atividade sintética feita pelo pensamento sobre as sensações; Contra o empirismo e o intelectualismo, que não há diferença entre sensação e percepção. |

*Quadro 1. Teorias sobre sensação e percepção.
Elaborado pelo autor baseado em Chauí (2000, p. 151-153)*

A partir dessas informações, Chauí (2000, p. 153) conclui que, para os empiristas e intelectuais, a percepção era considerada a atividade que unia as partes numa síntese que seria o objeto percebido. Na psicologia da forma e na fenomenologia a autora salienta que:

[...] não há diferença entre sensação e percepção porque nunca temos sensações parciais, pontuais ou elementares, isto é, sensações separadas de cada qualidade, que depois o espírito juntaria e organizaria como percepção de um único objeto. Sentimos e percebemos formas, isto é, totalidades estruturadas dotadas de sentido ou de significação.

Pode-se constatar que existem várias abordagens em relação a sensação e percepção, mas independentemente da perspectiva, são temas relacionados e, por muitas vezes, estudados em conjunto. A seguir, apresentam-se as abordagens de Merleau-Ponty e Gibson.

Merleau-Ponty, em sua teoria da Fenomenologia da Percepção, apresenta uma análise da percepção por meio da revisão do conceito de sensação, Santaella (2012b) observa que para Merleau-Ponty deve ser rejeitada qualquer tentativa de decomposição da percepção em sensações. Seguindo a mesma linha, na teoria de Gestalt a mente humana configura as informações através dos canais sensoriais, percepção e/ou da memória (pensamento, cognição e resolução de problemas). Na experiência humana do meio ambiente, esta configuração tem um caráter primário sobre os elementos que o compõem, e assim a soma desses elementos por si só não poderia levar-nos à compreensão do todo.

Neste panorama, das teorias que versam sobre percepção, é importante citar "A Ecologia da Percepção", cujo fundador foi James J. Gibson. Seria inviável sintetizá-la neste estudo, entretanto podemos elencar algumas contribuições importantes. Para Santaella (2012b) destacam-se algumas características: designação do conceito de *Affordance*: tudo aquilo que o ambiente oferece ao observador; esta teoria sofreu influência da teoria da Gestalt, mas também a refutou; as teses gibsonianas mais refutadas e submetidas a discussões são a percepção direta (sem a mediação de representações mentais internas) e imediata e a da negação de qualquer papel desempenhado pela mente na percepção.

Completando este cenário das teorias, posiciona-se aqui a Teoria da Percepção de Peirce, a qual é um dos objetos deste trabalho.

3.1. O processo de percepção em Peirce

Pretende-se abordar, neste tópico, as conexões entre a semiótica peirceana e o contexto em que essa teoria se desenvolveu, com o objetivo de

fornecer uma visão clara daquele que é o tema central deste trabalho. Algumas dessas conexões já foram descritas ao longo do trabalho, mas reaparecem aqui de forma mais ordenada.

De acordo com Santaella (2012b, p.89), foram poucos estudiosos da obra peirceana que se empenharam a investigar seus escritos sobre percepção. Ela realizou uma análise dos comentadores de Peirce que se dedicaram ao tema, e também aos ensaios peirceanos publicados. Como consequência desta exploração, destacaram-se diversos elementos importantes, os quais serviram de suporte ao desenvolvimento deste assunto.

No livro, *A percepção: uma teoria semiótica*, Santaella (1998, p. 16) afirma que, "em termos lógicos, ontológicos e epistemológicos", nenhuma outra teoria sobre percepção é tão completa quanto a peirceana, fundamentando-se em dois argumentos apresentados por Peirce: 1) "não há, nem pode haver, separação entre percepção e conhecimento"; 2) "todo pensamento lógico, toda cognição, entra pela porta da percepção e sai pela porta da ação deliberada."

Nessa perspectiva, as afirmações atestam a relevância do processo perceptivo na construção do conhecimento, sendo a percepção o elo entre a mente e o mundo exterior. Com o objetivo de esclarecer a composição da teoria da percepção, apresenta-se a seguir o modelo triádico formulado por Peirce, o qual está intimamente conectado aos conceitos já apresentados.

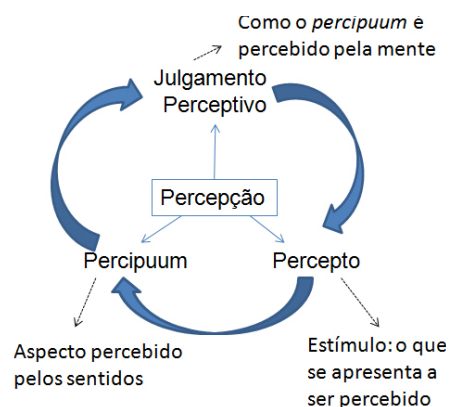


Figura 1. Constituintes peirceanos da percepção. Elaborado pelo autor baseado em Mucelin e Bellini (2008, p.124)

Ao iniciar a explanação sobre o modelo triádico da percepção peirceana é importante salientar que os elementos a serem descritos a seguir são associados aos conceitos iniciais do seu projeto filosófico. Na explicação de Santaella (2012b, p. 75), a percepção desempenha o papel de ponte entre o mundo da linguagem - a consciência, o cérebro,

a mente – e o mundo lá fora. A percepção é, para Peirce, o objeto de estudo da semiótica que ocorre segundo um modelo triádico e está constituído de: Percepto, *Percipuum* e Julgamento Perceptivo. A autora salienta que Peirce produziu a única teoria triádica da percepção, e explica que a percepção está sob a dominância da secundidade. A figura 1 procura mostrar a relação entre os elementos peirceanos da percepção.

A partir da proposta triádica ilustrada pela figura 1, Mucelin e Bellini (2008, p.125) consideram os constituintes da percepção como “entes interdependentes e indecomponíveis que permitem que se analise e caracterize isoladamente cada um deles”. Explica-se esses elementos da seguinte forma: percepto ou objeto, verdadeira coisa em si mesma, que independe daquilo que dele se possa pensar, devido à sua existência e insistência sobre os sentidos sem nada manifestar; o *percipuum* ou o modo como o percepto, captado pelos órgãos sensoriais, é imediatamente interpretado no julgamento de percepção; e o julgamento perceptivo, que corresponde a uma espécie de proposição a nos informar sobre aquilo que está sendo percebido. Ainda sobre essa relação triádica, deve-se destacar uma passagem do próprio Peirce (, 1994, p. 39, CP 7.643):

Não sabemos nada sobre o percepto a não ser pelo testemunho do juízo perceptual, exceto que sentimos o seu golpe, a reação dele contra nós, e vemos os conteúdos dele arranjados num objeto, na sua totalidade – excetuando-se também, é claro, o que os psicólogos são capazes de extrair inferencialmente. Mas, no momento em que fixamos nossa mente sobre ele e pensamos a menor coisa sobre o percepto, é o juízo perceptual que nos diz o que nós assim ‘percebemos’. Por esta e outras razões, proponho considerar o percepto tal como ele é imediatamente interpretado no juízo perceptual, sob o nome de *percipuum*.

Peirce nos apresenta neste trecho a indissociabilidade dos elementos no processo perceptual. Nesta concepção considera-se o *percipuum* a imediata interpretação do percepto através do juízo perceptual, ou seja, o *percipuum* apresenta-se como um elemento derivado do próprio percepto que já é resultado de uma elaboração mental que utiliza-se da organização que regulam os nossos julgamentos. Santaella (2012b, p. 95) também discorre sobre a tríade perceptiva, da seguinte maneira:

A percepção é determinada pelo percepto, mas este só pode ser conhecido através da mediação do signo, que é o julgamento da percepção. Para que esse conhecimento se dê, o percepto deve, de algum modo, estar representado no signo. Aquilo que representa o percepto, dentro do julgamento per-

ceptivo, é o *percipuum*, meio mental de ligação entre o que está fora e o juízo perceptivo, que já é fruto de uma elaboração mental.

Hausman (2006, p. 242) comenta que Peirce discute a função do percepto no início do processo de interpretação. Nesta visão, são sugeridos dois novos elementos o “antecepto” e o “ponecepto”. Um percepto isolado é apenas uma possibilidade perceptiva. O percepto sempre ocorre em sequências seriais que envolvem memórias recentes e antecipações futuras, chamadas respectivamente de poneceptos e anteceptos. Romanini (2006, p.68) complementa:

Perceptos similares chegam e coalescem na mente para formar o universo perceptivo, uma espécie de imagem mental, uma “ideia” que é o *ground* do julgamento perceptivo. Essa ideia é uma metáfora. O *percipuum*, portanto, é o percepto representado no julgamento perceptivo. É uma ficção, uma hipótese, assim como também o são o ponecipuum e o antecepium.

De acordo com Santaella (2012b), ao analisar a teoria peirceana da percepção, no âmbito da semiótica, o percepto funciona como objeto dinâmico enquanto o *percipuum* seria o objeto imediato. Nesta mesma análise é possível incluir o terceiro elemento da tríade, o julgamento de percepção. Santaella(2012b) referencia a obra de R. Bernstein (1964), em que o julgamento de percepção é correlacionado com as inferências abduativas sendo o elemento que desempenha o papel cognitivo na percepção. Nesse aspecto, a autora exemplifica afirmando que “são os juízos perceptivos que nos dizem, por exemplo, que o cheiro que estamos sentindo é de brócolis cozido, que aquilo que estamos vendo é uma lua cheia solitariamente iluminando o céu” (Santaella, 2012b, p.95).

A tríade da percepção permite vislumbrar análises de diferentes situações do cotidiano. Para reforçar o entendimento pode-se exemplificar através da seguinte situação: um indivíduo observa um obra de arte. Nesta breve circunstância pode-se ilustrar o funcionamento da percepção identificando os três elementos já apresentados. Primeiramente, a obra de arte em si seria o percepto, já o *percipuum* seria o conjunto de aspectos emanados da obra, e por fim, o aspecto percebido pelo indivíduo que denomina-se julgamento perceptivo, o qual é produzido pela mente através de um processo inferencial da percepção.

Desta forma, relaciona-se os elementos da tríade de percepção formulada por Peirce, a alguns conceitos fundamentais de sua arquitetura filosófica. Em síntese relaciona-se o percepto ao objeto dinâmico, o *percipuum* ao objeto imediato e o julgamento perceptivo às inferências abduativas.

A seguir apresenta-se a relação do julgamento perceptivo com os elementos da segunda tricotomia dos signos, o ícone, o índice e o símbolo. Ao agir como signo, o julgamento perceptivo apresenta três perspectivas equivalentes aos tipos sígnicos de Peirce: a icônica, a indicial e a simbólica. Santaella (2012b, p. 128) comenta que “nessa medida, o ingrediente icônico é justamente aquilo que dá suporte ao processo perceptivo, funcionando como substrato da ilusão, subjacente a toda percepção, de que o objeto, tal como percebido, é o próprio objeto”. Na visão de Granja (2005), a mediação icônica nos faz reconhecer tanto a forma geral dos objetos (a forma circular, a forma de uma cadeira), como a forma particular do objeto imediato que se apresenta à percepção. A junção entre o geral da semelhança icônica ao particular do objeto percebido corresponde justamente à dimensão indicial do julgamento perceptivo. Na sua dimensão simbólica ou de generalidade, o juízo de percepção é falível, e pode ser reavaliado a partir de experiências futuras.

Em geral, a representação de um objeto é complexa, que consiste em várias funções icônicas, indexais e simbólicas, as quais atuam na formação do julgamento perceptivo. Os ingredientes icônicos contribuem ao apresentar características gerais que permitem a identificação de traços essenciais à formação mental do *percipuum*. Em outro aspecto uma conexão indicial também é importante para fixar a referência, através dos indícios que se aproximam numa relação efetiva de seu objeto, afinal o índice é um indicador do seu objeto. A dimensão simbólica opera no terceiro nível do interpretante, onde através das características e indícios relacionados é formulado pensamento na mente, onde este pensamento é um símbolo. Dessa forma o julgamento perceptivo produz o que é percebido mentalmente.

Conforme destacado por Rego, Pessoa e Gala (2015, p. 105), Peirce não desenvolveu um texto que sistematizasse sua teoria da percepção, entretanto alguns intérpretes de sua obra organizaram essas ideias apresentando uma teoria muito rica. Desta forma, os elementos desta teoria estão calcados em sua arquitetura filosófica, onde os conceitos do processo perceptivo estão intimamente relacionados a todo trabalho desenvolvido no decorrer de suas pesquisas.

3.2. A percepção e o desenvolvimento de interfaces Web

O objetivo desta seção é encontrar elementos da teoria da percepção de Peirce que possam beneficiar o desenvolvimento de interfaces Web, tendo em vista o processo de recuperação da in-

formação. A percepção humana é um fator essencial a ser considerado no projeto de interfaces digitais, Andersen (2001) salienta a importância de estudos focados na interação entre as máquinas e os seres humanos, ele destaca o núcleo da semiótica na realização de estudos que discutam as representações e também a forma como os usuários interpretam essas representações. A seguir são elencadas algumas análises que consideram a percepção e a semiótica em diversas vertentes, e que colaboram para o desenvolvimento de interfaces digitais.

A relação entre a percepção humana e a organização do conhecimento foi destacada por Barat (2005), nessa pesquisa apresenta-se alguns aspectos referentes à teoria da percepção e à teoria linguística. O autor discute as vantagens das teorias, na teoria perceptual ele elenca três itens: a representação homogênea da primeira percepção, a visualização clara e lógica e os padrões de alto nível de processamento de dados e recuperação de informação. Na teoria linguística, é diagnosticado que ela é comum, conhecida e os símbolos são compreensíveis, além de descrever que os símbolos se comunicam diretamente e instantaneamente. A partir desses conceitos são listados também alguns pontos sobre percepção visual e recuperação de informação. Em suas conclusões reafirma a importância da compreensão da percepção no processo de gestão da informação. A apresentação do estudo fortalece a ideia de analisar a percepção como fator essencial no desenvolvimento de elementos que visam a melhor compreensão humana.

No que diz respeito aos estudos sobre interfaces digitais relacionadas à semiótica, a área de Engenharia Semiótica (ES) é apresentada por Souza (2005) como a teoria semiótica da área de IHC (Interação Humano Computador). Neste sentido a autora discute em seu artigo a ES como suporte teórico no desenvolvimento de interfaces com artefatos interativos que realmente possam trazer novas perspectivas e possibilidades de design em IHC. A ES trata a IHC como um tipo específico de comunicação mediada por computador, considerando a troca de mensagens dos usuários com os designers de sistemas computacionais, essas mensagens são o *representamen* (ou seja, o sinal perceptível que representa a mensagem) em que o próprio sistema informa aos usuários como se comunicar com ele. Em favor da engenharia semiótica defende-se que ela pode ajudar os designers a compartilhar responsabilidades com os usuários no processo comunicativo.

Também seguindo a vertente da engenharia semiótica, Vandí (2007) avaliou a forma como a semiótica possibilita analisar a ação prática dos

usuários reavaliando o uso de metáforas em interfaces gráficas, o autor considera metáforas os itens destinados a dar ao usuário um contexto familiar para agir: um escritório com lixeiras, pastas e arquivos. No artigo foram exemplificadas algumas interfaces computacionais que contém tais itens, sendo diagnosticado que a tarefa de um bom projetista é usar esses elementos para estimular a aprendizagem, e a fácil criação de hábitos de ação. Os hábitos de ação iniciam-se através da percepção, para Peirce (1994, CP 5.212) “os elementos de todo conceito entram no pensamento lógico pelos portões da percepção”, ou seja, toda informação deve entrar pelas portas da percepção, apresentando-se em forma de uma ideia ou hábito mental compartilhado numa comunidade.

Islam (2013) realizou uma revisão de literatura que abordou a percepção semiótica em interfaces de usuários no que diz respeito ao projeto e avaliação de usabilidade, foram selecionados 65 artigos que identificaram alguns desafios da área. No artigo, afirma-se a necessidade de realização de mais pesquisas ligadas ao tema com o intuito de melhorar a qualidade das publicações. O autor revela que apenas 18% dos artigos apresentaram o design de interface ou princípios de avaliação de usabilidade, cerca de 51% estavam relacionados a percepção semiótica, apresentando teorias, métodos e estruturas. Após uma verificação criteriosa nos estudos, o autor conclui que a revisão mostrou a importância da semiótica nas pesquisas interface do usuário, além disso sugere-se direcionar novos estudos à algumas lacunas de pesquisa identificadas, entre elas a percepção semiótica na avaliação da usabilidade.

Embora não tenha tratado especificamente sobre interfaces digitais, Gambarato (2013) discutiu a semiótica de Peirce no processo de design, em especial o conceito de abdução. A autora apresentou o potencial comunicativo dos signos com uma breve análise da teoria da percepção explicando a importância do conceito de abdução criado por Peirce. No artigo, é destacada a capacidade de representação por meio dos signos, e por fim conclui que o processo criativo no design tem o objetivo de associar signos e gerar novos interpretantes. Esse estudo apresentou-se interessante no que diz respeito à visão da teoria da percepção em processos comunicativos, visto que as interfaces também podem ser consideradas parte do processo comunicacional.

Magnani e Bardone (2006) discorreram sobre a relação do conceito de abdução e o projeto para interfaces web. No artigo apresentou-se a importância de três dimensões para as inferências ab-

duativas: a visual, a espacial e a emocional. A dimensão visual é exemplificada através do uso de cores, tamanhos e formatos. Já a dimensão espacial diz respeito a disposição dos elementos na interface, proporcionando a realização de tarefas específicas. No caso da dimensão emocional consideram-se aspectos de alertas aos usuários a partir de caixas de diálogo, sons e cores com significados de advertência e sucesso por exemplo. Os autores consideram a interface como fundamental na compreensão da interação humano-computador, e que esta interação pode ser melhor compreendida através do modelo inferencial. Por conta da possibilidade da interface fornecer pistas a partir das quais os usuários possam ter uma melhor percepção do ambiente, os autores consideram o processo como genuinamente abduutivo.

No que diz respeito à navegação em interfaces digitais web, Santaella (2004) identifica perfis relacionados às três categorias gerais de Peirce. Os tipos de raciocínio, abduutivo, indutivo e dedutivo, determinam diferentes percursos de leitura para três tipos de leitores (ou internautas): o errante, o detetive e o previdente. Neste cenário, o internauta errante é considerado o leitor novato, que por meio do raciocínio abduutivo navega aleatoriamente através das possibilidades à ele apresentadas. A indução é própria do internauta detetive, que está em processo de aprendizado, já que navega seguindo indícios (pistas) em busca da informação que necessita. Por sua vez, o internauta previdente é capaz de prever o resultado de suas ações na pesquisa, pois já passou pelo processo de aprendizagem e aplica o raciocínio dedutivo para avançar nos ambientes informacionais. Embora esses estilos sejam diferentes, existem pontos em comum, pois independente do itinerário os internautas recebem estímulos que despertam a percepção e a cognição. Dessa forma é identificado o papel fundamental da percepção no processo de navegação, onde os fatores sensoriais provocam imediatas reações perceptivas em sincronia com operações mentais.

Os estudos elencados acima, apresentaram diversos aspectos da perspectiva semiótica, englobando a teoria da percepção de Peirce, sugerindo que sua contribuição está voltada ao desenvolvimento de interfaces digitais. Contudo, verifica-se a necessidade de aprofundamento nos estudos experimentais relativos à percepção, sendo a semiótica de Peirce certamente um caminho teórico a ser percorrido.

4. Considerações finais

Estudar a percepção e decifrar como ela age na formação do pensamento, é um ganho impor-

tante para o entendimento da cognição em qualquer área disciplinar. Retornando a questão que norteou este trabalho, quais aspectos da semiótica peirceana contribuem para compreensão do processo de percepção humana? Foi possível relacionar os diversos fatores da arquitetura filosófica peirceana que se destacam no processo perceptivo. A proposta triádica, composta pelo percepto, o *percipiuum* e o julgamento perceptivo, se conectou com as categorias universais, aos objetos do signo e ao conceito de abdução.

A partir da apresentação das teorias observou-se as diversas abordagens sobre o tema, além de elencar algumas aplicações na área da Ciência da Informação que utilizam essas teorias como base para suas pesquisas. Com a revisão conceitual sobre a teoria da percepção de Peirce concluiu-se que seus conceitos fornecem subsídios para análise de diversos domínios do conhecimento, e proporciona um melhor entendimento do processo de percepção, substituindo uma relação diádica cartesiana de sujeito-objeto, para uma relação triádica do processo de percepção.

Nas diferentes áreas observadas foi possível identificar os componentes da tríade perceptual, isto é, entende-se que existem elementos a serem percebidos (percepto), como esses elementos se apresentam ao julgamento de percepção (*percipiuum*), e como tudo isto é imediatamente interpretado, o julgamento de percepção. Além dessa apresentação dos aspectos dos signos, outra vertente destacada diz respeito ao perfil dos internautas que utilizam interfaces Web, os aspectos perceptivos e cognitivos são influenciados por diferentes itinerários de navegação que podem ser distinguidos pelos tipos de raciocínio, abdução, indutivo e dedutivo. Dessa forma, uma sugestão de trabalhos futuros seria aperfeiçoar os processos cognitivos através do estudo da percepção, isto é, agir sobre cada elemento identificado no processo. Conforme destacado por Jorge, Rezende e Wartha (2013), a Teoria da Percepção poderia sustentar eficientemente o desenvolvimento da habilidade viso espacial, ou a questão da carga cognitiva sobre conceitos que demandam altas habilidades viso espaciais por meio do uso de ferramentas de visualização.

O desenvolvimento de interfaces digitais deveria levar em conta as fases do processo de percepção elencados por Peirce, além, é claro, dos tipos de mediação (icônica, indicial e simbólica) que se apresentam na interação homem-computador. Sem estes conhecimentos, as interfaces seriam apenas um mecanismo artificial e pouco eficiente de possibilitar a interação humana, o que, ao fim e ao cabo, interfere negativamente na aquisição de conhecimento e na fluência no uso do sistema.

Referencias

- Alexandre, D. S.; Tavares, J. M. R. S. (2007). Factores da percepção visual humana na visualização de dados. // APMTAC. <http://hdl.handle.net/10216/357>. (2015-11-10).
- Almeida, C. C. (2009). Peirce e a organização da informação: contribuições teóricas da semiótica e do pragmatismo. Marília: Universidade Estadual Paulista (UNESP), 2009. Tese de Doutorado.
- Andersen, P. B. (2001). What semiotics can and cannot do for HCI. // Knowledge-Based Systems. 14:8, 419-424.
- Andrade, S. M. et al (2012). Construção e evidências psicométricas de uma escala de avaliação da percepção visual. // Psicologia: Reflexão e Crítica. 25, 21-29.
- Barat, A. H. (2005). The relationship between human perception and knowledge organization. // La dimensió humana de l'organització del coneixement. Facultat de Biblioteconomia i Documentació. p. 286-295.
- Barros, C. M. de; CAFÉ, L. M. A. (2012). Estudos da semiótica na Ciência da Informação: relatos de interdisciplinaridades. // Perspectivas em Ciência da Informação. 17:3, 18-33.
- Chauí, M. (2000). Convite à filosofia. 7. ed. São Paulo: Ática, 2000.
- Correia, M. F. P. V. (2001). Técnicas computacionais na percepção visual do movimento. Porto: Universidade Fernando Pessoa, 2001. Tese de Doutorado.
- Frankenthal, E. S. (2004). O pragmatismo na abdução e na percepção. // Cognition-Estudos: Revista Eletrônica de Filosofia. 1, 1-9.
- Gambarato, R. R. (2004). Peircean semiotics in the context of design praxis: Abduction and perception in dialogue. // Sign Systems Studies. 41:4, 424-432.
- Granja, C. E. S. C. (2005). Música, conhecimento e educação: harmonizando os saberes na escola. São Paulo: Universidade de São Paulo (USP), 2005. Dissertação de Mestrado.
- Hausman, C. R. (2006). A Semiótica de Peirce aplicada à percepção: O papel dos objetos dinâmicos e dos perceptos na interpretação perceptiva. // Cognition. 7:2, 231-246.
- Ibri, I. A. (2006). The heuristic exclusivity of abduction in Peirce's philosophy. // Leo, R. F.; Marietti, S. (eds.). Semiotics and Philosophy in C. S. Peirce. Cambridge: Cambridge Scholars Press.
- Islam, M. N. (2013). A systematic literature review of semiotics perception in user interfaces. // Journal of Systems and Information Technology. 15, 45-77.
- Jorge, A. M. G. ; Rezende, D. B.; Wartha, E. J. (2013). Visualização, semiótica e teoria da percepção. // Tríade. 1:1, 149-166.
- Loomis, J. M. (2003). Percepção visual do espaço: fenomenologia e função. // Arquivos Brasileiros de Oftalmologia. 66:5, 26-29.
- Magnani, L. Bardone, E. (2006). Abduction and Web Interface Design. // Mueller, G. C. (org.). Encyclopedia of Human Computer Interaction. Liverpool: IGI.
- Matsunaga, C. (2015). Entre ver e interpretar: o papel fundamental do design. // Anais 7º Congresso Internacional de Design da Informação. 797-808. <http://www.proceedings.blucher.com.br/article-details/entre-ver-e-interpretar-o-papel-fundamental-do-design-20260> (2016-07-20).
- Moraes, J. A.; Pereira, P. H. A. O.; Pantaleão, N. C. A. (2012). Percepção direta: contribuições do pragmatismo peirciano para a abordagem ecológica. // Revista Eletrônica de Filosofia. 9, 195-204.
- Mucelin, C. A.; Bellini, M. (2008). Estudo da percepção em ecossistema urbano: uma contribuição para a educação,

- planejamento e gestão ambiental. // Revista Brasileira de Estudos Pedagógicos. 89:221, 119-144.
- Nóbrega, T. P. (2008). Corpo, percepção e conhecimento em Merleau-Ponty. // Estudos de Psicologia. 13:2, 141-148.
- Nöth, W. (2003). Panorama da semiótica de Platão a Peirce. 4. ed. São Paulo: Annablume, 2003.
- Peirce, C. S. (2005). Semiótica. 4. ed. São Paulo: Perspectiva, 2005.
- Peirce, C. S. (1994) The electronic edition of the collected papers of Charles Sanders Peirce. Vols. I-VI. // Hartshorne, C.; Weiss, P. (eds.). Cambridge: Harvard University Press.
- Rego, J. M. R.; Pessoa, G. H. R.; Gala, A. C. O. S. (2015). A Percepção em três registros: Russell, Peirce e Caeiro. // Revista Eletrônica de Filosofia. 12, 98-116.
- Rocha, L. B. (2002). Fenomenologia, semiótica e geografia da percepção: alternativas para analisar o espaço geográfico. // Revista da Casa da Geografia de Sobral. 4:1, 67-79.
- Romanini, A. V. (2006). Semiótica minuta: especulações sobre a gramática dos signos e da comunicação a partir da obra de Charles S. Peirce. São Paulo: Escola de Comunicações e Artes da Universidade de São Paulo (USP), 2006. Tese de Doutorado.
- Santaella, L. (2015). Semiótica aplicada. São Paulo: Cengage Learning, 2015.
- Santaella, L. (2012a). O que é semiótica. São Paulo: Editora Brasiliense, 2012a.
- Santaella, L. (2012b). Percepção: fenomenologia, ecologia, semiótica. São Paulo: Cengage Learning, 2012b.
- Santaella, L. (2004) Navegar no ciberespaço: o perfil cognitivo do leitor imersivo. São Paulo: Editora Paulus, 2004.
- Santaella, L. (1998) A percepção: uma teoria semiótica. 2. ed. São Paulo: Experimento, 1998.
- Santaella, Lucia. A teoria geral dos signos: semiose e auto-geração. São Paulo: Ática, 1995.
- Santos, N. A. (2003). Tópicos em percepção e processamento visual da forma: acuidade visual versus sensibilidade ao contraste. // Estudos e Pesquisas em Psicologia. 3:1, 61-71.
- Schultz, S. (2012). A análise gödeliana do conceito de percepção. // Dois pontos. 9:2, 135-150.
- Sousa, H. M. R. C. (2009). Percepção humana na visualização de informação crítica. Porto: Universidade do Porto, 2009. Dissertação de Mestrado.
- Souza, C. S. (2005). Semiotic engineering: bringing designers and users together at interaction time. // Interacting with Computers. 17:3, 317-341.
- Vandi, C. (2007). Beyond metaphors and icons: towards a perception-action model for graphic user interfaces. // Anais 9th World Congress of IASS/AIS. http://www2.uiah.fi/sefun/DSIU_papers/DSIU_Vandi_Beyond%20metaphors%20and%20icons.pdf (2016-07-20).

Enviado: 2017-03-31. Segunda versão: 2017-09-06.

Aceptado: 2017-09-15.

Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias

Evaluación de las características de TemaTres y Multites para el control de autoridades en bibliotecas universitarias

An assessment of TemaTres and Multites characteristics for authority control in academic libraries

**Mariângela Spotti Lopes FUJITA, Luciana Beatriz Piovezan dos SANTOS,
Maria Carolina Andrade CRUZ, Walter MOREIRA**

UNESP, Campus de Marília, Av. Hygino Muzzi Filho, 737, Marília, SP, 17525-300, fujita@marilia.unesp.br, nanipiovezan@gmail.com, maria.andradeacruz@gmail.com, Walter.moreira@marilia.unesp.br

Resumen

Se analizaron dos softwares para la construcción y mantenimiento de tesauros, TemaTres y MultiTes, de cara a facilitar el control de autoridades en bibliotecas universitarias. Los resultados apuntan que ambos softwares cumplen con la mayoría de los requisitos, y que TemaTres se emplea ya en sistemas de bibliotecas universitarias en la recuperación por materia de registros bibliográficos del catálogo en línea. Sin embargo, se recomienda la aplicación experimental de los dos softwares para el análisis de su comportamiento operativo tanto en la construcción del tesauro como en su uso con registros de autoridad y en la recuperación por materias.

Palabras clave: Tesauros. Software de gestión de tesauros. Catálogos en línea. Control de autoridades. Bibliotecas.

1. Introdução

Diante ao grande fluxo de produções tecnológicas e científicas, os centros de informação, como as bibliotecas, buscam aprimoramento no armazenamento e recuperação da informação. O método para contribuir com essa situação resultou na criação de recursos próprios para o controle de vocabulário no tratamento da informação: as linguagens de indexação.

De acordo com Baritè (2011, p. 127, tradução nossa) o controle de vocabulário “contribui com a seleção, depuração, formalização e normalização da terminologia, incluindo, o estabelecimento recíproco entre os termos, considerando critérios semânticos, linguísticos e disciplinares”. Com essas características, compreende-se que o controle de vocabulário em um sistema de informação, favorece tanto o sistema de organização da informação quanto o sistema de recuperação.

A linguagem de indexação é uma ferramenta de controle de vocabulário utilizada pelo indexador na etapa de tradução dos conceitos do processo

Abstract

Two thesaurus management software applications, TemaTres and MultiTes, were assessed regarding their suitability for controlling authorities in university libraries. The results indicate that both software applications comply with most of the requirements, and that TemaTres is already used in university library systems for retrieving bibliographic records from online catalogues. However, an experimental assessment of their operational behavior both in the construction of thesauri, in its use with authority records and in subject retrieval is recommended.

Keywords: Thesauri. Thesauri management software. Authority control. Online catalogues. Libraries.

da indexação e na estratégia de busca pelo usuário. Compreende a mediação entre a linguagem natural extraída dos documentos e a do usuário na estratégia de busca, buscando assim contribuir para a busca e recuperação do documento (Martins, 2013).

De acordo com a norma norte americana ANSI/NISO Z39.19 as linguagens de indexação configuram-se em (American National Standards Institute/ National Information Standards Organization 2005, p. 11, tradução nossa)

[...] vocabulário controlado ou um sistema de classificação com regras para sua aplicação. Uma linguagem de indexação é utilizada para a representação dos conceitos tratados nos documentos [objeto de conteúdo] e para a recuperação de tais documentos [objetos de conteúdo] de um armazenamento de informações e de sistema de recuperação

Ainda de acordo com a norma norte americana, as linguagens de indexação configuram-se nos seguintes formatos (segundo a lógica de complexidade): as listas, os anéis de sinônimos, as

taxonomias e os tesouros. Neste artigo, pretende-se abordar *softwares* para a gestão especificamente dos tesouros.

Compreende-se a importância dos tesouros nas palavras de Cervantes (2004, p. 35) ao dizer que o tesouro é considerado como o mais adequado para o tratamento e recuperação da informação, pois “além de estabelecer relações entre os termos utilizados, denominados de preferidos e não preferidos, este instrumento permite uma visão sistêmica da área - objeto de representação”. Os avanços tecnológicos vêm aprimorando de modo significativo as ferramentas destinadas à representação e recuperação da informação. Em relação aos sistemas de informação, percebe-se a necessidade de investir em novas tecnologias para garantia do bom desempenho de seus sistemas de recuperação.

Os tesouros, assim como outras linguagens mais complexas, necessitam dos *softwares* para a sua construção e manutenção. Em vista das necessidades de recuperação por assunto em catálogos online de bibliotecas, cujos padrões descritivos e temáticos utilizam metadados para a construção de registros bibliográficos e registros de autoridade, torna-se fundamental a construção e manutenção de linguagens de indexação com uso de *softwares* acessíveis e que suportem grande quantidade de termos e seus relacionamentos, tendo em vista que as bibliotecas, em sua maioria, possuem coleções bibliográficas em todas as áreas de conhecimento. Cabe destacar também, como requisito mínimo desejável ao *software*, as condições que este oferece para a construção e manutenção de tesouros que abriguem aspectos de interoperabilidade em ambientes multilíngues.

Para a identificação e acesso à *softwares* de gestão de tesouros, destacam-se os trabalhos de reunir em uma lista os recursos existentes com descrições técnicas e operacionais, tais como o de Leonard D. Will no site “*TaxoBank: Access, deposit, save, share, and discuss taxonomy resources*” (1) e o da Biblioteca da Universität Basel associada à International Society of Knowledge Organization (ISKO) “*Basel Register of Thesauri, Ontologies & Classifications*” (2) que disponibilizam informações sobre os *softwares* de gestão de tesouros.

Consideramos de fundamental importância para o desenvolvimento de avaliação de *softwares* de gestão de tesouros os resultados dos trabalhos de Campos et al (2006), de Santos et al. (2013) e Martínez González e Alvite Díez (2014). O primeiro trabalho “Estudo comparativo de *softwares* de construção de tesouros” realizou comparação com seis *softwares* de construção de tesouros, entre eles o TemaTres e o MultiTes. O segundo

trabalho “Elaboração de vocabulário controlado em formato SKOS usando Tematres: implicações metodológicas e web-semântica” incluiu experiência de construção de um tesouro sobre “cerveja” utilizando os *softwares* TemaTres e o MultiTes. O terceiro trabalho apresenta uma análise sistemática de seis *softwares* para a gestão de tesouros, entre eles o TemaTres e o MultiTes, mediante aplicação de um conjunto de critérios estabelecidos a partir de sistematização de literatura que considera o propósito e funcionalidade dos aplicativos, integração com outros sistemas de informação ou interoperabilidade da informação e a capacidade de combinar diferentes fontes de tesouros. A proposta é analisar os dois *softwares* para construção e manutenção de tesouros citados nos trabalhos de Campos et al (2006), de Santos et al. (2013) e Martínez González e Alvite Díez (2014) o TemaTres e o MultiTes.

2. O controle de registros de autoridades em catálogo online de biblioteca: funções e finalidades

Partindo da idéia de Jewett (1853) de catalogação cooperativa e uso de recursos tecnológicos, a *Library of Congress* (LC) criou a maior Central de Catalogação cooperativa. A introdução da tecnologia computacional na década de 1960 e o desenvolvimento do formato MARC para padronização de registros bibliográficos e formatos de intercâmbio refletiu-se até os dias atuais nas bibliotecas e centros de informação.

O desenvolvimento de tecnologias de *software* e hardware juntamente com a Internet proporcionou às centrais de catalogação cooperativa disponibilizarem seus catálogos online, acelerando a troca de informações. Para propiciar o intercâmbio de informações, foram criados, junto ao formato MARC, os protocolos Z39.50 (3) e ISO-2709 (4) que orientam sobre a estrutura de um registro e a transferência de dados bibliográficos.

Quando um registro bibliográfico é criado, o catalogador faz a descrição bibliográfica do item documentário nos campos devidos e quando se depara com os campos de assunto, recorre à Base de Autoridades Assuntos que mantém os denominados Registros de autoridade que formam, atualmente, o controle de vocabulário em formato MARC21.

Os registros de autoridade destinam-se a conter informação autorizada para padronizar nomes e assuntos a serem usados como pontos de acesso em registros bibliográficos, bem como permitir os relacionamentos entre eles. Compõem uma base de dados que é consultada, tanto pelo catalogador ao atribuir os termos que representam assuntos, quanto automaticamente pelo próprio sistema para correção de termos e

inclusão de remissivas dentro dos registros bibliográficos. Isso significa que os registros de autoridade têm o objetivo de assegurar a constante padronização da linguagem de indexação de forma automática e podem ser transferidos de uma base de dados para outra na formação dos catálogos de bibliotecas e rede de bibliotecas.

Em seguida, apresenta-se exemplo de registro de autoridade de assunto de termo tópico da ferramenta *Classification Web* da *Library of Congress Subject Headings* (LCSH) em formato de ficha terminológica simples e registro MARC.

Yoga

[\[B132.Y6 \(Philosophy\)\]](#)
[\[BL1238.52-.56 \(Hinduism\)\]](#)
 UF Yoga—Hinduism
 Yoga exercises
 BT [Exercise](#)
[Philosophy. Indic](#)
 NT [Astānga yoga](#)
[Chakras](#)
[Christianity and yoga](#)
[Hatha yoga](#)
[Music for voga](#)
[Naam yoga](#)
[Siddha yoga \(Service mark\)](#)
[Yin yoga](#)
[Yogis](#)

Figura 1. Registro de autoridade em ficha completa (Fonte: *Classification Web* – LCSH)

| Subject Record [Yoga] | | | |
|------------------------------------|---|---------------------|---------------|
| ID: | sh 85149174 | Entered: | 050328 |
| 008:06 Geo Subd: | -No attempt | 008:11 SH System: | a-LCSH |
| 008:07 Roman: | -No attempt | 008:15 Subj Use: | a-Appropriate |
| 008:09 Kind Rec: | a-Estab lhdg | 008:17 Type Subd: | n-Not applic |
| | | Replaced: | 20120312 |
| | | 008:29 Ref Eval: | a-Eval |
| | | 008:31 Rec Upd: | a-Can be used |
| | | 008:33 Level Estab: | a-Fully |
| 010 | sa sh 85149174 | | |
| 035 | sa (DLC)sh 85149174 | | |
| 035 | sa (DLC)6498834 | | |
| 035 | sa (DLC)sp 85149174 | | |
| 035 | sa (DLC)327038 | | |
| 040 | sa DLC \$c DLC \$d DLC | | |
| 053 | 0 sa B132.Y6 \$c Philosophy | | |
| 053 | 0 sa BL1238.52 \$b BL1238.56 \$c Hinduism | | |
| 150 | sa Yoga | | |
| 450 | sa Yoga \$x Hinduism | | |
| 450 | sa Yoga exercises | | |
| 550 | sw g sa Exercise | | |
| 550 | sw g sa Philosophy. Indic | | |
| Record: 271494 | | | |
| Added: Mon Mar 28 00:00:00 2005 | | | |
| Modified: Mon Mar 12 12:34:34 2012 | | | |

Figura 2 Registro de autoridade em formato MARC21 (Fonte: *Classification Web* – LCSH)

Ressalte-se que a Library of Congress Subject Headings – LCSH, evoluiu em suas últimas edições com a introdução e uso da nomenclatura de relações semânticas adotadas por tesouro, incluindo uma estrutura lógica de relações hierárquicas anteriormente atrofiada na linguagem. Essa evolução da LCSH, os formatos de intercâmbio de registros bibliográficos e os de autoridades e

os catálogos online demonstram que são importantes para a indexação e recuperação por assuntos em catálogos online de bibliotecas.

O controle de registros de autoridades tem sua importância porque estão ligados à própria necessidade de construção de vocabulários controlados legíveis por computador que contribuam para o desenvolvimento da web semântica, sobretudo porque torna possível sua reutilização por outros sistemas de informação (Martins, 2013).

A esse respeito é preciso refletir sobre a importância de registros de autoridade com o necessário controle de vocabulário aliado à construção de linguagens de indexação. Nesse sentido, o controle de autoridades tem duas funções importantes a de manter a coerência da forma usada na representação dos pontos de acesso e das relações entre conceitos. Os registros de autoridade têm esse nome porque foram criados para o controle dos cabeçalhos de autores, mas como os cabeçalhos de assuntos necessitam também manter o controle da forma de termos e de seus relacionamentos com outros termos e conceitos, o registro de autoridade passou a ter função imprescindível na construção de linguagens de indexação.

A importação de registros de autoridade de assunto por catalogadores sem a definição de diretrizes de construção e manutenção de linguagens de indexação contraria a dupla função do controle de autoridades e não propicia a adequação do uso de linguagens de indexação para o controle de vocabulário durante a indexação e durante a estratégia de busca para a recuperação por assuntos em catálogos online. Portanto, é necessário que o *software* de gestão da construção e manutenção de tesouros a ser utilizado na indexação e recuperação por assuntos em catálogos online tenha condições de fazer a importação dos registros de autoridade para que se mantenha a integridade da linguagem de indexação.

3. A construção e manutenção de tesouros na perspectiva da normalização para análise de softwares

No âmbito da Ciência da Informação, a palavra tesouro nomeia o instrumento que reúne um conjunto controlado de termos de um determinado domínio, demonstrando as relações conceituais e terminológicas ou relações semânticas e de equivalência entre estes termos. Tal instrumento tem a função de auxiliar nos processos de representação e de recuperação da informação documentária em domínios específicos, respeitando sua terminologia própria.

Em obra na qual apresenta aspectos essenciais da construção de tesouros, Broughton (2006, p.4, tradução nossa), define tesouro como:

[...] uma ferramenta usada para a indexação do assunto de documentos. Consiste em uma lista estruturada de termos (geralmente em um campo de assunto específico) que um indexador ou gerenciador de registros pode usar para descrever documentos de modo que os usuários finais possam recuperar itens relevantes ao pesquisar material sobre um determinado assunto.

Na definição apresentada na norma ISO 25964-1 (2011) enfatiza-se o aspecto estrutural do tesouro. Nesta norma, define-se tesouro como:

Vocabulário controlado e estruturado em que os conceitos são representados por termos, organizados de modo que as relações entre conceitos sejam explicitadas e os termos preferidos sejam acompanhados por termos de entrada para sinônimos ou quasi-sinônimos.

Em nota explicativa para a definição apresentada a referida norma caracteriza o aspecto funcional do tesouro. Considerando-se este aspecto (International Organization for Standardization 25964-1, 2011, tradução nossa),

Nota: O propósito de um tesouro é orientar tanto o indexador quanto o pesquisador para a seleção do mesmo termo preferido ou de uma combinação de termos preferidos para representar um determinado assunto. Por esta razão, um tesouro é otimizado para a navegabilidade humana e a cobertura terminológica de um domínio.

Em contexto histórico, o aparecimento dos tesouros é relativamente recente e foi determinado pela conjuntura de 1950 e 1960 que apresentou progressos no campo tecnológico e descobertas científicas em diversas áreas do conhecimento, gerando o aumento da produção científica. Com isso, a atenção para os sistemas de organização e recuperação da informação se intensificaram, pois, compreendia a necessidade de buscar ferramentas para o processamento dos documentos, como as linguagens próprias. Nesse período, teve o início do desenvolvimento dos primeiros tesouros (Cintra et al, 2002).

O aparecimento dos primeiros tesouros, da Dupont em 1959, do *Defense Technical Information Center* (DTIC) (anteriormente *Armed Services Technical Information Agency* - ASTIA) em 1960, *Chemical Engineering Thesaurus* em 1961, *Thesaurus of Engineering Terms*, publicado pelo *Engineering Joint Council* (EJC) em 1964, influenciou a elaboração das primeiras diretrizes sobre a construção de tesouros, publicadas pelo *Committee on Scientific and Technical Information* (COSATI) em 1967, utilizadas para a construção do *Thesaurus of Engineering and Scientific Terms* (TEST) aem 1967 que serviram de

base para a elaboração das primeiras normas sobre construção de tesouros: *Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri* da UNESCO em 1970 e a ANSI Z39.19 de 1974 da *American National Standards Institute* (Cervantes e Fujita, 2012; Mendes, Reis e Maculan, 2015).

Em um estudo sobre as origens do processo de construção de tesouros, Cervantes e Fujita (2012) e Campos et al. (2006), citam a figura de Lancaster (1995, p.50) que demonstra graficamente a evolução dos tesouros, considerada por Campos et al. (2006) com duas vertentes, a alfabética, representativa da linha americana, e a sistemática, da linha europeia. Cervantes e Fujita (2012, p. 124), em complementação, consideram que essas “origens podem ser ou parcialmente teóricas, iniciadas pela linha teórica ranghanatiana (classificações bibliográficas), ou parcialmente empíricas, com embasamento aplicado às práticas da indexação alfabética de assunto iniciadas pelas regras de Cutter”.

Na evolução atual das normas estão presentes a Norma Z39.19 - *Guidelines for the construction, format, and management of monolingual controlled vocabularies* (American National Standards Institute/ National Information Standards Organization, 2005) – proposta desenvolvida pela organização norte-americana *National Information Standards Organization* (NISO) e a Norma ISO 25964 (2011) “Information and documentation: part 1: thesauri and interoperability with other vocabularies” e mais recentemente a Norma ISO 25964 (2013) “Information and documentation – thesauri and interoperability with other vocabularies - part 2: Interoperability with other vocabularies”.

Clarke e Zeng (2012, p.25) esclarecem que a Parte 1 da Norma ISO 25964 “atualiza, revisa e substitui a ISO 2788 e a ISO 5964, bem como algumas partes da BS 8723”. Isso significa que, mais uma vez, a atualização abrange várias normas e segue influências de vertentes metodológicas importantes. As normas atuais, como se constata, evoluíram consideravelmente a partir da última edição da ISO 2788:1986 – “Guidelines for the establishment and development of monolingual thesauri” (1986), demonstrada ao final da figura de Lancaster, cujas influências são provenientes tanto da indexação alfabética de assuntos quanto da classificação bibliográfica advindas das duas vertentes metodológicas. Anterior a esta última edição é a norma ISO 5964:1985 “Guidelines for the establishment and development of multilingual thesauri” cuja principal contribuição foi o tratamento de várias línguas em um mesmo recurso.

Conforme Martins (2013) a norma ANSI/NISO Z39.19 (2005) é a quarta edição da norma ANSI Z39.19-1974 cujo fator determinante de mudança foi a aplicação de tecnologias de informação para desenvolvimento de tesouros eletrônicos em função dos avanços da informação eletrônica e da internet.

O lançamento da norma britânica BS 8723 nos anos de 2005-2008 adianta-se em relação a esses avanços da norma ANSI Z39.19 (2005) e traz diretrizes acerca da interoperabilidade linguística, análise de facetas e interoperabilidade entre linguagens de indexação, incorporadas, na evolução da Norma ISO 25964-1 (2011). Dessa forma, a norma ISO 25964 incorpora em suas diretrizes vários aspectos evolutivos das normas que a precedem: os aspectos de mono e multilinguismo, interoperabilidade linguística, interoperabilidade entre vocabulários controlados, aplicação da análise de facetas na construção do tesouro, definição de protocolos e formatos necessários para o intercâmbio de dados e a proposta de modelização de dados.

De acordo com divulgação das normas na página da NISO (5), a parte 1 da ISO 25964 de 2011 “fornece recomendações para o desenvolvimento e manutenção de tesouros destinados a aplicações de recuperação de informação, seja monolíngue ou multilíngue” com destaque para seu item 14 que é dedicado especialmente a orientações quanto aos requisitos de *software* para a gestão de tesouros. A parte 2 da ISO 25964-2: 2013 “dá recomendações para o estabelecimento e manutenção de mapeamentos entre vários tesouros, ou entre tesouros e outros tipos de vocabulários”. Nesta segunda parte da norma, há interfaces relativas à interoperabilidade com os esquemas de classificação, as taxonomias, os cabeçalhos de assunto, as ontologias, as terminologias, as listas de autoridade de nomes e os anéis de sinônimos.

O avanço existente na norma ISO 25964, em relação às demais normas anteriores, refere-se fundamentalmente à diferenciação entre conceito e termo explicitados por modelos de dados para que as máquinas possam manipular e interpretar dados e realizar processo de inferência (Alvite Diez e Martínez Gonzalez, 2015, Santos, et al., 2013). Isso significa que os conceitos estão representados por termos e que é preciso identificar e descrever tanto as relações que ocorrem entre os conceitos (relações conceituais) quanto as que ocorrem entre os termos (relações terminológicas). Tal proposta nos remete à ideia do tesouro conceitual de Dahlberg (1978) como foi explicado por Campos et al. (2006, p.71):

Este é um tesouro com base em conceitos, onde na sua construção são necessários princípios para o

estabelecimento do termo/conceito e das relações entre eles. Neste tipo de tesouro o sentido linguístico não é mais determinante na criação e organização dos termos e sim o seu conteúdo conceitual.

Na avaliação das normas atuais, além das concepções conceituais que as fundamentam, observa-se que dedicam itens específicos para a gestão e manutenção de tesouros com uso de *softwares*. A evolução das tecnologias de informação na sociedade humana é notoriamente um fator de mudança e nas bibliotecas não foi diferente mesmo porque, o fluxo de produção científica tem se intensificado cada vez mais e a busca por tecnologias que contribuam para a recuperação da informação criada eletronicamente é imprescindível.

Tal necessidade é constatada por Boccato (2011, p. 5) em sua afirmação de que

Com o crescimento da indústria da informação e, conseqüentemente, com o advento das bases de dados bibliográficas, houve a necessidade de diretrizes normativas na construção de tesouros disponíveis não só no formato impresso, como também no formato eletrônico, via web.

Os *softwares* para a construção dos tesouros são, de fato, uma necessidade para elaboração da ferramenta em ambiente digital. Os *softwares* são necessários pelas “[...] facilidades de armazenamento, manipulação e apresentação dos termos definidos e suas relações” (Campos et al 2006, p. 69).

Na década de 1990, Milstead (1991) apontou a carência de *softwares* voltados especificamente para contemplar os requisitos básicos para garantir o bom desempenho no sistema de recuperação na base de dados de um tesouro. Hoje, o cenário se modifica e apresenta *softwares* capacitados para atender as demandas dos sistemas de informação que investem na esfera da recuperação da informação.

O precursor na criação de tesouros assistidos por *software* no Brasil foi o IBICT. Conforme histórico descrito na página do IBICT (6) *software* TECER foi desenvolvido no final dos anos 1980 e amplamente distribuído em âmbito nacional. Pode se perceber a necessidade da ferramenta diante do interesse e uso pela comunidade. Apesar disso, por um período de tempo, o *software* saiu do mercado por falta de atualização. Atualmente, o IBICT decidiu atualizar o *software* e disponibilizá-lo para download. Em análise da nova versão do TECER, Mendes, Reis e Maculan (2015, p.62) observam que, “apesar de os estudos anteriores (conforme demonstrado no Quadro 2) já terem identificado diversificada gama de tipos de relacionamentos, podemos notar que o *software* não foi capaz de atender a essa demanda.” O Quadro

2 do artigo citado, refere-se aos relacionamentos (equivalência, hierárquicas, associativas) nos tesouros segundo marcos teóricos de revisão de literatura.

Os *softwares* destinados à construção de tesouros precisam oferecer recursos para atender as necessidades da instituição que o adquirir. Em suas características operacionais, é essencial que o *software* seja flexível e dinâmico, permitindo manutenção em sua estrutura, como por exemplo, a inclusão ou exclusão de termos da estrutura do tesouro. Isso acontece devido aos avanços tecnológicos e a premissa que as ciências não são engessadas, na verdade se modificam a todo o momento gerando novos saberes e descobertas. Moura et al. (2005) ressaltam que o tesouro é uma ferramenta em permanente construção e isso gera a sua longevidade.

É imprescindível que a instituição invista inicialmente na elaboração de um planejamento, pontuando suas características e necessidades para compra de um *software* ou mesmo desenvolvimento de um próprio (Gomes; Campos, 2004). A partir do conhecimento desses aspectos há maiores condições para melhores investimentos nos *softwares*. Referente a esses aspectos, Gomes e Campos (2004, p. 2) ressaltam que é alcançada uma maior segurança para a manutenção da estrutura, evitando a criação de termos anômalos e acrescenta-se que “O que se deve almejar são regras genéricas que contemplem os casos particulares. Isto é: regras prescritivas”.

Em seus estudos, Milstead (1991, p.174) relaciona os critérios básicos para um *software* ser utilizado na construção de tesouros:

Operação em uma plataforma de computador disponível; Integração com *software* de banco de dados existente; Uma interface amigável; Geração automática de recíprocos; Prevenção de relações conflitantes; Validação de termos em relacionamentos; Suporte de todos os relacionamentos padrão ANSI/NISO; Disponibilidade para adicionar campos de texto; Confirmação de alterações; Uma grande variedade de relatórios e exibições.

A norma utilizada pela autora é ANSI/NISO “*Guidelines for Thesaurus Structure, Construction and Use*” é datada em 1980, já a norma ANSI/NISO Z39.19 de 2005 “*Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*” aponta os seguintes requisitos para um *software* de gerenciamento de vocabulários controlados, ou seja, as linguagens de indexação, tal qual os tesouros:

1. oferecer suporte, no mínimo, as relações básicas: USE/UP, TG/TE e TR; 2. oferecer suporte para o uso de notas de escopo e de notas históricas; 3. for-

necer tanto exibições hierárquicas quanto alfabéticas; 4. ser, preferencialmente, não-proprietário; 5. não ser limitado a um único usuário; 6. ser independente de hardware e operar em sistemas abertos ou em ambiente padrão do sistema operacional; 7. requerer pouco ou nenhum treinamento do usuário; 8. fornecer documentação detalhada ao usuário (American National Standards Institute/ National Information Standards Organization Z39.19, 2005, p. 99).

É interessante observar que a norma Z39.19 (1980) possui precauções ainda necessárias a um *software* de construção e manutenção de linguagem de indexação, porém em alguns pontos pressupõe-se que os *softwares* já estão atualizados nesses quesitos, como operar em uma plataforma de computador disponível, integração com *software* de banco de dados existente, possuir uma variedade de relatórios e exibições ou mesmo uma interface amigável. Estes são elementos importantes, mas na avaliação dos *softwares* e diante da atualizada norma Z39.19 (2005), é visível a atenção voltada às relações para a estruturação do tesouro e a inserção dos usuários, diferente da norma Z39.19 (1980) que pontuava elementos mais ao modo operacional do *software*.

Os problemas de compatibilidade no uso de linguagens de indexação, tais como o tesouro, dizem respeito à como o usuário poderá familiarizar-se intuitivamente e sem ajuda. Sobre isso a ISO 25964-1 (2011, p. vi) considera que “[...] no passado os tesouros foram planejados para profissionais de informação treinados em indexação e busca, hoje existe uma demanda para vocabulários que usuários sem treino irão considerar intuitivos e para vocabulários que permitam inferências por máquinas”.

A Norma ISO 25964-1 (2011) recomenda o uso de *software* de gestão de tesouros para sua construção e manutenção a fim de se evitar erros de digitação e outros erros mais comuns. Para isso, indica em seu item 14 diretrizes específicas para gestão de *softwares* de tesouro com a orientação de que devem ser capazes de gerenciar, importar e exportar dados. Para a escolha do *software* indica requisitos a serem seguidos, tais como tamanho e limitações de caractere; relações entre termos e entre conceitos; notas aplicáveis a termos ou conceitos; códigos e notação; etiquetas de vértices ou nó entre termos; status de idiomas; importação e exportação de dados; navegação editorial e apoio; salvaguardas editoriais; e ferramentas de organização.

É preciso destacar dois aspectos importantes na construção e manutenção de tesouros que influenciam na escolha do *software* de gestão de tesouros: o incremento das relações de equivalên-

cia e associativas e a construção de tesouros navegáveis. No primeiro aspecto, Boccatto (2011, p. 20) ressalta que “o incremento das relações de equivalência e associativas no repertório terminológico possibilita maior aproximação entre o tesouro e a linguagem de busca do usuário, condição essa de suma importância para o adequado desempenho do instrumento terminológico em sistemas automatizados de informação”. Com relação ao segundo aspecto, Boccatto (2011, p. 21) ressalta que tesouros navegáveis permitem “a migração de tesouros tradicionais a formatos inteligíveis por máquinas, que propicia o intercâmbio e a interoperabilidade entre sistemas de organização do conhecimento”.

4. Softwares de construção e manutenção de tesouros: identificação e análise para discussão de resultados

Os *softwares* TemaTres e MultiTes foram identificados nos trabalhos de Campos et al (2006), de Santos, et al. (2013) e Martínez González e Alvite Díez (2014) e selecionados para análise e discussão quanto à proposta de avaliação para construção e manutenção de linguagem de indexação com e para uso no controle de registros de autoridades de catálogos online de bibliotecas.

A análise realizada por Campos et al. (2006) comparou seis *softwares*, entre eles o TemaTres e o MultiTes, quanto a requisitos divididos em oito grupos: a) Características gerais, b) Tratamento de relações, c) Tratamento de dados, d) Interface/manipulação dos dados, e) relatórios, f) Características de implementação, g) Apoio metodológico, e h) Apoio ao uso. Em Santos et al. (2013) foi realizada a análise dos dois *softwares*, TemaTres e MultiTes, no desenvolvimento experimental de um micro-tesouro sobre “Cervejas” que avaliou o desempenho metodológico na construção, uso e conversão em vocabulário estruturado para a web semântica. O estudo de avaliação de Martínez González e Alvite Díez (2014) aplicou em seis *softwares* sete grupos de critérios de avaliação: a) Objetivo do aplicativo, b) Requisitos do sistema, c) Funcionalidades, d) estruturas conceituais e terminológicas compatíveis, e) Gestão da Integridade, f) interoperabilidade da informação, g) Interoperabilidade e integridade do *Software*, utilizando para os testes de avaliação os tesouros multilíngues: EuroVoc - Tesouro multilíngue da União Europeia, UKAT - UK Archival Thesaurus, NAL - National Agricultural Library thesaurus - e um microtesouro elaborado *ad hoc* a partir de EuroVoc.

Os dois *softwares* serão descritos quanto aos seus requisitos técnicos e funcionalidades opera-

cionais e sua aplicabilidade será analisada a partir dos trabalhos de Campos et al (2006), Santos et al. (2013) e Martínez González e Alvite Díez (2014) acompanhados de literatura adicional.

4.1. TemaTres

O TemaTres é uma aplicação web utilizada para a gestão de representações linguísticas formais do conhecimento, incluindo vocabulários controlados, taxonomias, tesouros e listas de cabeçalhos de assunto (Gonzales Aguilar, Ramírez Posada, Ferreyra, 2012; Ferreyra, 2016).

É um *software* livre (7) desenvolvido por Diego Ferreyra, na Argentina em 2004 (Rodríguez Gairín; Russo Gallo; Sulé Duesa, 2008), e seu sistema também permite desenvolver estruturas de navegação *web* e funcionar como complemento de um sistema de gerenciamento de bibliotecas, para a gestão da linguagem.

O *software* TemaTres está disponível para aquisição gratuitamente em sua página web (8) onde também é possível encontrar um tesouro de demonstração construído com o *software*. A interface de uso do TemaTres é multilíngue e está atualmente em sua versão 2.0, disponibilizada em espanhol, alemão, francês, inglês, italiano e português.

De acordo com seu manual de utilização (Tematres, 2015), para instalação, o *software* requer um servidor web que suporte PHP; uma instalação de PHP em funcionamento com versão 4.3.0 ou superior; e um servidor de bases de dados.

A instalação do *software* em si, se dá em três passos (Tematres, 2015): 1) realizar o download do arquivo e descomprimi-lo; 2) configurar a conexão com a base de dados MySQL no arquivo /include/db.tematres.php; e 3) seguir as instruções de instalação do arquivo.

O *software* TemaTres permite a criação de linguagens dotadas de controle de vocabulário, entre suas funções permite a criação de relações entre termos, em número ilimitado por termo: relações de equivalência (USE - UP), relações hierárquicas (TG - TE), relações associativas (TR), e permite criar novos tipos de relacionamento entre termos

Possui ferramenta de consistência que permite estabelecer termos equivalentes, não equivalentes e parcialmente equivalentes; atribuição de múltiplas notas aos termos (Notas de aplicação, escopo, explicativa, histórica, bibliográfica etc.); apresenta fluxos de situação termos (termo candidato, termo aceito, termo rejeitado) e possui suporte para vocabulários poli-hierárquicos. Per-

mite também a navegação entre termos, com sugestão de termos de busca; navegação sistemática e hierárquica (Tematres, 2017).

É possível a exportação completa do vocabulário em formatos que visam à interoperabilidade, tais como, em RDF SKOS-Core, Dublin Core (ISO 15836-2003), BS 8723 (*Structured Vocabularies for Information Retrieval*), entre outros formatos, para TopicMaps (ISO/IEC 13250:2003), em XML Zthes, MADS (Metadata Authority Description Schema), em formato .txt (ordem alfabética e hierárquica) e SiteMap 0.8 (Tematres, 2017).

O *software* permite interligar dados entre vocabulários (gerenciar relações com termos de outros vocabulários); auditoria de termos (busca por termos livres e repetidos); administração de usuários e editores e oferece relatórios estatísticos: média de palavras por termo, termos por número de termos genéricos, termos por número de termos específicos, termos por nível de profundidade (Tematres, 2017).

A aplicação web do TemaTres permite visualizar as relações determinadas entre os termos em uma mesma tela.

É importante ressaltar que o *software* TemaTres possui ferramenta para definição de categorias, com o uso do recurso de Meta-Termos, porém não é possível estabelecer facetas e não permite a indicação de categorias ou macroestruturas, como observado por Santos et al. (2013) que na experiência de aplicação experimental necessitou desenvolver o microtesauro sobre “Cervejas” no MultiTes porque a instalação do TemaTres era mais complexa. Nesse sentido o trabalho aponta que o TemaTres em versão mais atualizada foi escolhido “por conter recursos para geração de listas em formatos como SKOS, XML/BS8723 e XML/DC.

Também, como relatado por Pérez-León e Martínez-González (2010), embora o TemaTres ofereça diversas opções para exportação do tesauro, não há opção para importar tesauros prontos para serem editados no *software*.

Na análise realizada por Campos et al. (2006) com o TemaTres na versão de 2005, ressaltou que era o *software* livre mais completo dentre os demais dessa categoria e que, de maneira geral atendeu de maneira satisfatória a maioria dos requisitos. A análise mais específica de cada requisito apontou algumas limitações tais como: não possui capacidade para referenciar a fonte do termo; não permite a criação de relações definidas pelo usuário que, na análise dos autores, pode ser um problema para a caracterização de diferentes tipos de relações partitivas; não permite a importação nem controle de atualização

dos dados; tem problemas na reorganização de hierarquias; permite apenas a emissão de relatório alfabético.

Martínez González e Alvite Díez (2014) descrevem o *software* TemaTres como uma ferramenta dirigida a desenvolvedores tradicionais de tesauros que torna possível a edição de vocabulários controlados por meio de uma interface web simples e que dispõe de versão web livre para navegar e fazer buscas no vocabulário, além de criar e editar tesauro próprio por meio de mecanismos do *software* que deverá ser instalado em servidor web local. A avaliação de funcionalidades tais como criação e edição, busca e recuperação, navegação, fusão de tesauros, importação e exportação e colaboração de usuários demonstraram que somente a fusão de tesauros não é possível e que só permite a busca de descritores ou termos preferidos. Por outro lado, dentre os seis tesauros é o único que permite a edição colaborativa de tesauros. Em estrutura conceitual e terminológica apresenta estrutura clássica de tesauro. No que se refere à Integrabilidade destacam como vantagem a simplicidade do *software* e o enfoque de serviço web que motivam o usuário.

Considerando-se as limitações apontadas pelos autores, observa-se pelas informações atualizadas de Martínez González e Alvite Díez (2014) e do TemaTres em 2017 uma evolução técnica e operacional no TemaTres de instalação e uso, porém, as normas ISO 25693 (2011) e ANSI/NISO Z39.19 (2005) não são citadas. Apesar disso, citam a BS 8723 que fundamentou avanços significativos atualmente existentes nessas duas normas mais atualizadas.

4.2. MultiTes

MultiTes Pro é um *software* proprietário (9), desenvolvido em 1983 para gestão de tesauros, taxonomias e outros vocabulários controlados. A aplicação possui atualmente, uma versão online acessível por um navegador web, além da versão convencional, que deve ser instalada em um computador local. Possui também três serviços complementares que permitem a publicação na web do tesauro, a criação de interfaces sob demanda para pesquisas no tesauro e um conjunto de ferramentas para a publicação do tesauro no servidor da organização para manutenção local. A aplicação está disponível para uso em ambientes Windows 7 ou superior (Multites, 2017) mas também pode ser usado em ambientes Linux (Mandal, 2015).

Entre suas funções, o MultiTes permite a criação de relacionamentos padronizados em conformidade com a norma norte-americana ANSI/NISO

Z39.19, (USE, UP, BT, NT, RT) com geração automática de relações recíprocas, bem como a criação de notas de escopo e de relações entre termos definidas pelo usuário (Multites, 2017). De acordo com Mandal (2015), MultiTes também realiza validação automática de relações conflituosas entre termos.

O *software* aplica automaticamente aos *top terms* e termos não-preferidos formatação padrão (negrito e itálico, respectivamente) para melhor identificação. Permite a criação de poli-hierarquias e tesouros multilíngues. Permite a exportação do tesouro em formatos tais como texto simples, RDF/SKOS, XML, RTF, CSV, HTML e em formato configurável pelo usuário.

Na análise realizada por Campos et al. (2006) o *software* MultiTes de 2005 é um *software* proprietário que atende de maneira mais geral a todos os requisitos analisados e de forma específica observou algumas limitações tais como: o recurso de interface web não é atendido e dificulta a disponibilização em acesso remoto; não permite a geração de um relatório sistemático único com hierarquia de gênero-espécie e parte-todo.

A aplicabilidade experimental do microtesouro sobre “Cerveja” destacou que foi possível criá-lo no MultiTes e convertê-lo para o TemaTres, porém, no processo de migração de dados observou que o arquivo de exportação do Multites não traz elementos suficientes para o reconhecimento do TemaTres e, por isso, as relações entre associativas e de equivalência e notas de escopo precisaram ser manualmente indicadas no arquivo antes da importação.

Martínez González e Alvite Díez (2014) descrevem o *software* MultiTes como ferramenta comercial disponível pela web e um editor de tesouros com objetivo de construção e manutenção. A avaliação de funcionalidades, acima citadas para o TemaTres, demonstraram que somente a funcionalidade de fusão de tesouros não é possível. Nos critérios de: estrutura conceitual e terminológica apresenta estrutura clássica de tesouro, gestão da integridade, superou todos os testes realizados e de integrabilidade oferece serviços web.

Martínez González e Alvite Díez (2014) concluem que os serviços web com livre acesso oferecidos pelo *software* TemaTres e outros *softwares* são úteis para consulta, mas as funcionalidades analisadas não resolvem os problemas de criação e edição de tesouros. Recomendam que seja elaborado um padrão geral de API (10) para a gestão de tesouros.

5. Considerações finais

Considerando a importância do controle de vocabulário para as atividades de organização e representação do conhecimento, é certo que a qualidade de seus instrumentos deve ser sempre perseguida. A construção e manutenção de linguagens de indexação tais como o tesouro com uso de *softwares* adequados é essencial para a garantia de sua permanente atualização e consistência.

Nesse sentido, realizou-se análise dos *softwares* de construção e manutenção de tesouros TemaTres e MultiTes, que foram citados em avaliações anteriores realizadas por Campos et al. (2006), Santos et al. (2013) e Martínez González e Alvite Díez (2014).

Diante das características identificadas nos dois *softwares* abordados, tanto por suas informações oficiais quanto na literatura analisada sobre avaliação de *softwares* de gestão de tesouros, compreende-se que ambos têm as características desejadas para a função a qual se propõem. Porém, em razão de termos deparado com informações conflituosas sobre estas funções, bem como, em razão de que tais *softwares* disponibilizam atualizações em suas plataformas, podendo no momento presente mostrarem-se com funções mais desenvolvidas ou alteradas, consideramos necessária a realização de aplicação experimental destes *softwares* para realizar análises mais aprofundadas.

A análise referencial demonstrou que é possível identificar se os *softwares* possuem os requisitos estabelecidos na norma ANSI/NISO z39.19 (2005) quanto aos critérios de escolha de *softwares* deste tipo, porém para os requisitos do item 14 da norma ISO 25693 (2011) esta análise não pode ser feita, visto que requer a instalação e teste dos *softwares*.

Também se coloca como relevante a aplicação experimental para identificar se é possível a realização de edição de tesouros mediante importação de arquivos de autoridades, contendo registros de autoridades em formatos utilizados em sistemas de gestão de catálogos de bibliotecas, posto que tal informação não ficou clara na análise realizada e é de essencial importância para o uso em sistemas de bibliotecas.

Ainda em relação às características desejáveis de um *software* destinado à construção e manutenção de tesouros, destaca-se de modo enfático a necessidade de investigar, como se pretende fazer em trabalhos posteriores, as funcionalidades relativas à interoperabilidade, como, *e.g.*, a possibilidade de exportação do tesouro em formato SKOS, XML, RDF, entre outros. A adoção

de um *software* que não contemple a inserção do tesouro em ambientes de *web* semântica, além de anacrônica contribui para ações de isolamento entre as diversas bases de dados e unidades de informação. O resultado dessa postura já é conhecido mesmo fora do ambiente *web*: inoperância do tesouro em função de aspectos que restringem sua aplicação a contextos geográficos e linguísticos mais delimitados.

Quanto ao uso destes *softwares* em sistemas de bibliotecas, identificou-se que atualmente o TemaTres está sendo utilizado para gestão do Vocabulário Controlado da USP e consta em sua interface de busca por assunto (11). Já quanto ao MultiTes não foi possível identificar experiência de uso em ambientes semelhantes.

Reconocimientos

Investigação desenvolvida com apoio da FAPESP e CNPq

Notas

- (1) www.taxobank.org
- (2) <https://bartoc.org/>
- (3) "Z39.50 é um protocolo de comunicação entre computadores desenhado para permitir pesquisa e recuperação de informação - documentos com textos completos, dados bibliográficos, imagens, multimeios - em redes de computadores distribuídos" (Rosseto, 1997, p.2).
- (4) ISO 2709-1981, Documentation - Format for bibliographic information interchange on magnetic tape, criada para o intercâmbio de dados em fitas magnéticas na década de 1960, ainda é utilizada (Assumpção e Santos, 2015).
- (5) <http://www.niso.org/schemas/iso25964/>
- (6) <http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/software-para-gerenciamento-de-tesauros-tecer/historico>
- (7) *Software* livre é uma denominação que indica que depois de adquirido o *software* pode ser usado, copiado, estudado, modificado e redistribuído livremente.
- (8) <http://www.r020.com.ar/tematres/wiki/doku.php?id=tematres:install>
- (9) *Software* proprietário é uma denominação que indica que a aplicação está sob proteção legal de direitos autorais, de modo que não é permitido ao usuário copiar, ter acesso ao código-fonte ou redistribuir a aplicação.
- (10) Application program interface (API) is a set of routines, protocols, and tools for building software applications. An API specifies how software components should interact. Additionally, APIs are used when programming graphical user interface (GUI) components. Fonte: <http://www.wikipedia.com/TERM/A/API.html>
- (11) Cf. <http://vocab.sibi.usp.br/pt-br/index.php>

Referencias

Alvite Diez, M. L., Martínez González, M. M. (2015). Armonización de ISO 25964 y SKOS em *software* de gestión de tesauros: propuesta de análisis. // XII Congreso ISKO España y II Congreso ISKO España-Portugal, 19-20 de no-

vembro, 2015, Organización del conocimiento para sistemas de información abiertos. Murcia: Universidad de Murcia, 2015. 266-278. http://www.iskoiberico.org/wp-content/uploads/2015/11/62_Alvite.pdf (2017-03-25).

Assumpção, F. S.; Santos, P. L. V. A. da C. (2015). Representação no domínio bibliográfico: um olhar sobre os Formatos MARC 21. // *Perspectivas em Ciência da Informação*, 20:1 (2015) 54-74. <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2054> (2017-03-29).

American National Standards Institute/ National Information Standards Organization Z39.19-2005. (2005) Guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Maryland: NISO Press, 2005. <http://www.niso.org/standards/resources/Z39-19-2005.pdf> (2017-03-25).

Baritê, M. G. (2011) Sistemas de Organización del Conocimiento: uma tipologia atualizada. *Informação & Informação*. 16:3 (2011) 122-139.

Bocatto, V. R. C. (2011) Os sistemas de organização do conhecimento nas perspectivas atuais das normas de construção. InCID: R. Ci. Inf. e Doc., 2(1),165-192, jan./jun. <http://www.revistas.usp.br/incid/article/view/42340/4601> (2017-03-25).

Broughton, V. (2006) *Essential thesaurus construction*. London: Facet, 2006.

Campos, M. L. A.; Gomes, H. E. (2006). Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. // *Perspectivas em Ciência da Informação*. 11:3, 348-358. <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/273> (2017-03-25).

Campos, M. L. de A. et al. (2006) Estudo comparativo de *softwares* de construção de tesauros. *Perspectivas em Ciência da Informação*. 11:1 (2006) 68-87. http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1413-99362006000100006&lng=p&tlng=pt (2017-03-25).

Cervantes, B. M. N. (2004). Contribuição para a terminologia do processo de inteligência competitiva: estudo teórico e metodológico. Universidade Estadual Paulista, Marília, 2004. 183 f. Dissertação de Mestrado em Ciência da Informação.

Cervantes, B. M. N.; Fujita, M. S. L. (2012). Uma perspectiva diacrônica de parâmetros e modelos de construção de tesauros. // Cervantes, B. M. N. (Org.). *Horizontes da organização da informação e do conhecimento*. Londrina: EDUEL, 2012. v. 1, p. 1-18.

Cintra, A. M. M.; et al. (2002) *Para entender as linguagens documentárias*. 2.ed.rev.ampl. São Paulo: Polis/APB, 2002. 96p.

Dahlberg, I. (1978) Teoria do conceito. // *Ciência da Informação*. 7:2 (1978) 101-07.

Dextre Clarke, S. G.; Zeng, M. L. (2012). From ISO 2788 to ISO 25964: the Evolution of thesaurus standard towards interoperability and data modeling. // *Information Standards Quarterly*. 24:1 (Winter 2012), 20-26. <http://www.niso.org/publications/isq/2012/v24no1/clarke/> (2017-03-25).

Gomes, H. E.; Campos, M. L. A. (2004) Tesauro e normalização terminológica: o termo como base para intercâmbio de informações. // *DataGramaZero*, 5:6 (2004). <http://basessibi.c3sl.ufpr.br/brapci/v/a/2052> (2017-03-25).

Gonzales Aguilar, Audilio; Ramírez Posada, María; Ferreyra, Diego (2012). *Tematres: software para gestionar tesauros*. // *El profesional de la información*. 21:3 (Mayo-Junio) 319-325.

International Organization for Standardization (2011). ISO 25964-1:2011 Information and documentation - Thesauri

- and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization, 2011.
- International Organization for Standardization (2013). ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies. Geneva: International Organization for Standardization, 2013.
- International Organization for Standardization 2788:1986 (1986) Documentation – guidelines for the establishment and development of monolingual thesauri. Geneva: ISO.
- Jewett, C. (1853). Smithsonian Catalogue System. Washington, D.C.: Smithsonian Institution. // Mason, Moya K. (2017) Historical Development of Ideas Concerning Library Catalogues: Their Purpose and Organization. <http://www.moyak.com/papers/history-library-catalogues.html>
- Lancaster, F. W. (1995) El control del vocabulario en la recuperación de información. Tradução de Alejandro de la Cueva Martín. València: Universitat de València, 1995. (Educaió. Materials, 12).
- Mandal, S. (2015). MultiTes: a knowledge organization thesaurus construction tool for college libraries under the University of Burdwan. // International Research Journal of Interdisciplinary & Multidisciplinary Studies (IRJIMS). 1:8 (September 2015), 63-79. <http://irjims.com/files/Sukumar-Mandal.pdf> (2017-03-24).
- Martínez González, M. M.; Alvite Díez, M. L. Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. // Anales de Documentación, 17:1. <http://dx.doi.org/10.6018/analesdoc.17.1.186271> (2017-03-24).
- Martins, M. C. B. (2013) Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema no Instituto Politécnico de Portalegre. 391p. Departamento de Biblioteconomia y Documentación. Universidad de Salamanca. Salamanca. Tese de Doutorado. <http://comum.rcaap.pt/handle/10400.26/7171> (2017-03-25).
- Mendes, P. R.; Reis, R. M. dos; Maculan, B. C. M. dos S. (2015). Tesaurus no acesso à informação: uma retrospectiva. // Revista ACB: Biblioteconomia em Santa Catarina, 20(1), 49-66. http://revista.acbsc.org.br/racb/article/view/980/pdf_110 (2017-03-25).
- Milstead, Jessica L. (1991) Specifications for thesaurus software. // Information Processing & Management, 27:Z/3 (1991), 165-175.
- Moura, M. A.; et al. (2005) Linguagens de indexação em contextos cinematográficos: a experiência de elaboração do tesouro eletrônico do cinema brasileiro. // Perspectivas em Ciência da Informação. 10:1, 54-69. <http://www.eci.ufmg.br/pcionline/> (2008-04-08).
- MULTITES. (2017). MultiTes Pro. // MultiTes: Thesaurus and Taxonomy Authoring Tools. <http://www.multites.com/productsPRO.htm> (2017-03-25).
- Pérez-León, B., & Martínez-González, M. M. (2010). A comparative study of thesauri tools: a perspective from integrability in information systems. // Filipe, Joaquim y Cordeiro, José (eds.). Proceedings of the 6th International Conference on Web Information Systems and Technologies, WEBIST 2010. Valencia: INS-TICC, 2010. 203-206. <https://goo.gl/bTYD7q> (2017-03-26).
- Rodríguez Gairín, Josep Manuel; Russo Gallo, Patricia; Sulé Duesa, Andreu (2008). A virtual exhibition of open source software for libraries. // 16th BOBCATSSS Symposium 2008, Zadar (Croatia), 28-30 January 2008. <http://hdl.handle.net/10760/11151> (2015-12-13).
- Rosetto, M. (1997). Uso do protocolo z39. 50 para recuperação de informação em redes eletrônicas. // Ciência da Informação, 26:2 (1997).
- Santos, C. A. C. M. dos et al. (2013) Elaboração de vocabulário controlado em formato SKOS usando Tematres: implicações metodológicas e web-semântica. // Dodebei, V.; Guimarães, J. A. C. (Org.). Complexidade e organização do conhecimento: desafios do nosso século. Rio de Janeiro, Marília: ISKO-Brasil, 2013. FUNDEPE. v. 2, p. 205-210. http://isko-brasil.org.br/?page_id=42 (2017-03-04).
- Scornik, C.; Sosa, O. (2007) Análisis de software para la gestión de lenguajes documentales: caso de estudio: software TemaTres. Corrientes: Universidad Nacional del Nordeste, 2007. <http://www.unne.edu.ar/unnevieja/investigacion/com2008/E-020.pdf> (2017-03-25)
- TemaTres. (2017). TemaTres: gestión de vocabularios controlados. <http://r020.com.ar/tematres/wiki/doku.php?id=inicio> (2017-03-24).

Enviado: 2017-03-31. Segunda versión: 2017-09-08.
Aceptado: 2017-09-15.

Procedimiento de evaluación

Evaluation process

1. Introducción

El procedimiento de evaluación en Scire es el de revisión por pares mediante el sistema de doble ciego —los revisores no conocen el nombre de los autores, ni viceversa— semejante al de otras revistas científicas, aunque tiene algunas peculiaridades que permiten ajustarlo a las condiciones específicas de su campo científico y de sus lectores.

2. Objetivos

El objetivo del procedimiento de evaluación es que se puedan admitir con rapidez los trabajos buenos, mejorar los que pueden ser susceptibles de aceptación —esto es, que, siendo interesantes, contienen errores o lagunas que deben ser corregidas— y rechazar los que no son adecuados para la revista por su tema o calidad, redirigiéndolos hacia otras publicaciones adecuadas o haciendo sugerencias al autor para la posible reutilización y reorientación de su trabajo.

El procedimiento de revisión tiene que tener en cuenta a los autores de los artículos que revisa, con los que el revisor debe intentar ser justo, pero también debe responder ante los lectores y los demás autores que compiten por el espacio de la revista, así como, de forma más general, con el proyecto editorial, la propia comunidad científica y el conjunto de la sociedad.

3. Procedimiento

Cada artículo requiere al menos la opinión favorable de al menos dos revisores y cada nota breve de al menos uno. Si el director de la revista estima que el artículo entra dentro de los objetivos de Scire —que se encuentran públicamente disponibles en cada número y en la página web de la revista—, lo envía a dos revisores —o, si es necesario, a más— elegidos por el director. Los evaluadores pueden ser miembros del consejo científico, revisores propuestos por los miembros del consejo científico con anterioridad o con motivo del artículo en cuestión, o científicos de reconocido prestigio propuestos por el autor o autores del artículo. La elección se hace según su conocimiento del tema y el número de trabajos que han evaluado hasta el momento (para equilibrar la carga de trabajo entre ellos).

Si el director estima sin lugar a dudas que el tema del artículo no se adecua al de la revista, se lo comunica al autor con una explicación suficiente, y, si le es posible, le sugiere otra revista que pueda ser más adecuada. De esta forma, se ahorra tiempo del autor y de los revisores.

La agilidad del procedimiento de revisión es un aspecto fundamental, por lo que el director solicita a los revisores que envíen su valoración en dos semanas, y que, en caso contrario, renuncien expresamente al proceso de revisión. Sin embargo, si el revisor no realiza ninguna de estas acciones, el proceso se retrasa sin remedio. En este sentido, el autor debe tener en cuenta que el proceso de revisión tiene un carácter voluntario. La decisión de aceptar el trabajo depende, salvo en casos excepcionales, de la opinión de los revisores. En los casos en que no están de acuerdo, la decisión es tomada por el director de la revista, que, excepcionalmente, puede decidir enviar el artículo a un nuevo revisor.

4. Criterios de evaluación

Se solicita a cada revisor que indique su valoración sobre los siguientes aspectos en una escala del 1 al 5 (1 muy cuestionable, 2 marginal, 3 suficiente, 4 sólido, 5 excelente):

1. *Tipo de contribución*: Se indica si se trata de una contribución de tipo teórico o metodológico, un estado de la cuestión, una comunicación de resultados de investigación, una experiencia específica, una exposición con valor introductorio o docente sobre el tema o una noticia de interés. Si es una mezcla de varios tipos, es posible marcar más de uno.

2. *Adecuación del tema de la revista*: Aunque al enviar el artículo al revisor el director de la revista ya ha realizado un juicio sobre su adecuación al enfoque de la revista Scire, el revisor puede también valorar este aspecto, comparando el tema del artículo con la declaración de objetivos y alcance de la revista, disponible en cada número y en la página web.

3. *Interés e importancia del tema*: Grado en el que el artículo aborda una cuestión de gran interés para el área científica de Scire, ya sea tradicionalmente, en el momento actual o en el futuro previsible.

4. *Actualidad del tema*: Grado en el que el tema se sitúa en los frentes actuales de investigación. No es un aspecto imprescindible, pero sí se considera informativo.

5. *Novedad y originalidad teórica de la aportación*: Grado en el que el artículo aporta una novedad teórica a su campo de investigación; esto es, un nuevo problema, un nuevo enfoque, un nuevo modelo, una nueva relación o ley, una nueva hipótesis, un nuevo concepto o una interpretación o modificación original de otros anteriores. Debe tratarse lógicamente de una aportación relevante y fundamentada.

5. *Novedad y originalidad metodológica de la aportación*: Grado en el que el artículo aporta una novedad metodológica a su campo de investigación; esto es, un nuevo procedimiento de estudio, de análisis o de prueba, aunque sea a un problema conocido.

7. *Aplicabilidad práctica*: Grado en el que las ideas del artículo pueden servir para la mejora de procedimientos y actividades prácticas y, en definitiva, para el desarrollo y la innovación.

8. *Rigor metodológico y validez de los resultados*: Grado de coherencia, exactitud, precisión y cuidado con el que se aborda el tema de estudio; validez y actualidad de la metodología utilizada; rigor en la selección de las muestras y en la aplicación de la metodología a las mismas; grado de replicabilidad del estudio.

9. *Rigor en el estado de la cuestión y el reconocimiento del trabajo previo*: Hasta qué punto se ha realizado un estado de la cuestión suficiente y actualizado y se han manejado las referencias pertinentes para el problema abordado.

10. *Claridad expresiva, orden y facilidad de lectura*: El artículo debe estar redactado de forma clara y bien ordenada, sin redundancias ni ideas fuera de contexto; y debe ser fácilmente entendible por cualquier lector potencial de Scire. No debe faltar información necesaria ni deben existir lagunas sobre el problema abordado, el método, los resultados, la discusión, las conclusiones o las recomendaciones. Si el artículo contiene aparato matemático, debe ser comprensible por un lector no especialista.

11. *Adecuación de gráficos, ilustraciones, tablas y apéndices*: Los gráficos, ilustraciones, tablas y apéndices deben ser necesarios y relevantes, estar bien resueltos técnicamente y disponer de una leyenda suficientemente explicativa.

13. *Conformidad de las referencias a las normas de la revista*: Grado en el cuál las referencias del artículo son suficientes y se conforman a las normas de la revista Scire.

5. Familiaridad del revisor con el tema

Aunque el director de la revista envía los artículos para su revisión intentando que coincidan con el campo de especialidad del revisor, pueden producirse desajustes. El revisor puede rechazar la revisión sugiriendo o no un revisor alternativo, o puede efectuar la revisión indicando de 1 (mínimo) a 5 (máximo) su grado de familiaridad con el tema del artículo evaluado. Si no rellena esta información, se entenderá que su grado de familiaridad es alto (4) o muy alto (5).

Esta información sirve para que el director de la revista pueda estimar la necesidad de enviar el artículo a un nuevo revisor en caso de conflicto de pareceres entre los revisores.

6. Valoración final

El revisor puede realizar cinco tipos de propuestas finales:

1. Aceptar el artículo para su publicación sin restricciones ni cambios, porque es un trabajo de alta calidad.

2. Aceptar el artículo para su publicación una vez realizadas las correcciones indicadas, bien sea incondicionalmente, al ser un trabajo de calidad, o bien de forma condicionada a que haya espacio disponible en el número, por ser un trabajo de calidad marginal. Las correcciones solicitadas se indican en las instrucciones para el autor, son de carácter menor y su comprobación puede ser delegada en el comité editorial.

3. Someter a reevaluación el artículo una vez realizadas las correcciones indicadas en las instrucciones para el autor, porque son de importancia y amplitud, y deben ser comprobadas por el revisor original u otro revisor.

4. Rechazar el artículo por las razones indicadas en las instrucciones para el autor, donde el revisor especificará con claridad y amabilidad las razones por las que su trabajo debe ser rechazado y, si le resulta posible, realizará sugerencias sobre su mejora o lugares alternativos de publicación.

5. El revisor rechaza evaluar el artículo por su falta de familiaridad con el tema o por razones que especifica en los Comentarios confidenciales para el equipo directivo.

La labor de los revisores es fundamental para que una publicación científica realice su labor de manera adecuada. Scire agradece de forma muy especial el tiempo y esfuerzo que dedican, y reconoce su trabajo publicando periódicamente sus nombres.

Instrucciones para la presentación de trabajos

Instructions for authors

1. Cuestiones generales

Scire admite artículos con estados de la cuestión, discusiones de carácter teórico y metodológico, e informes de investigación sobre proyectos terminados, así como noticias sobre proyectos en curso, experiencias, notas y reseñas, que traten de la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Las lenguas de publicación de Scire son el español, el inglés y el portugués.

Los artículos son sometidos a proceso de revisión por pares según el sistema de doble ciego, aunque los autores pueden proponer revisores libremente. El procedimiento es público, y puede ser leído en el apartado anterior a las instrucciones.

Los autores se comprometen a presentar y discutir públicamente y presencialmente sus trabajos en el congreso Scire correspondiente al año en curso.

Los trabajos deben ser originales y tendrán una extensión máxima de 20.000 caracteres en el caso de los artículos; 12.000 caracteres para los estudios de caso; y 6.000 caracteres para las noticias, proyectos en curso, experiencias, notas y reseñas. Excepcionalmente, estas cifras podrán ajustarse en virtud del interés del trabajo, aunque siempre con el acuerdo previo del editor.

Los autores deben presentar en una hoja aparte sus datos personales completos: nombre, cargo, dirección, teléfono, fax y correo electrónico.

Los artículos han de presentarse en un documento Word o RTF debidamente maquetados según la plantilla disponible en www.ibersid.org con sus gráficos incluidos. Irán precedidos de sus títulos, resúmenes informativos y palabras clave, todos ellos en inglés y español. Se incluirán en ficheros aparte los gráficos e imágenes en su archivo fuente para su eventual tratamiento, y, en cualquier caso, con resolución de impresión.

Todos los apartados, incluidos las notas, referencias y apéndices irán numerados secuencialmente. Los subapartados —cuyos títulos serán claros y sintéticos— deberán tener la forma “1.1. ”, “1.2. “, “1.2.1. “ y así sucesivamente. No numere los resúmenes en español y en inglés, ni

los apartados de notas, agradecimientos y referencias. Evite un rango de encabezamientos superior a tres.

No se deben relatar datos analíticos repetitivos ni elencos de recursos en el texto principal. Por favor, preséntelos en forma de tablas o, si fuera imposible, en forma de apéndice.

El editor podrá realizar cambios de estilo para adecuar el trabajo a los requisitos de la revista.

Se recuerda al autor que la violación de derechos intelectuales y códigos éticos es un comportamiento inaceptable y puede constituir un grave delito.

2. Formatos de texto

No utilice subrayados, negritas ni versales.

Utilice mayúsculas sólo para acrónimos. No ponga puntos y espacios entre las letras que los forman.

Las cursivas se deben utilizar tan sólo para resaltar términos en otras lenguas, títulos de obras, palabras clave muy importantes que podrían servir a modo de títulos de un párrafo o conjunto de párrafos, y, con parquedad, para enfatizar los demás términos y frases a gusto del autor.

No utilice listas con salto de línea o de párrafo. Redacte párrafos completos según las reglas y los recursos gramaticales disponibles en su lengua de redacción. Puede indicar orden en las enumeraciones por medio de números o letras entre paréntesis, como, por ejemplo, (1), (2)...

3. Figuras y gráficos

Las figuras deben presentarse también en ficheros aparte en formato EPS o PDF con resolución de impresión de alta calidad (600 ppp) y, además, en el formato de la aplicación original con la que se creó. Tenga cuidado si genera imágenes a partir de la pantalla de su ordenador. Si recurre a imprimir pantalla en Windows obtendrá solo una resolución de 72 ppp. En ese caso, utilice una pantalla lo más grande posible, amplíe la ventana al máximo, copie con la tecla “impr pant”, pegue en el programa “Paint” y guarde el fichero; compruebe luego los resultados. Existen programas que permiten obtener imágenes de pantalla

a una resolución aceptable. En el cuerpo del artículo las imágenes portarán epígrafes indicativos de su contenido precedidos de la palabra “Figura” y un número correlativo (por ejemplo, “Figura 1. “).

4. Tablas

Deben realizarse con el mismo procesador de textos que el artículo. En texto portarán epígrafes indicativos de su contenido precedidos de la palabra “Tabla” y un numeral romano correlativo (por ejemplo, “Tabla I. “).

5. Notas

Las notas explicativas deben usarse excepcionalmente. No utilice el sistema automático de su procesador de textos. Refiéralas en texto tecleando su número secuencial entre paréntesis, e inclúyalas al final del trabajo en un apartado titulado “Notas” situado inmediatamente antes del dedicado a las “Referencias”, en orden secuencial y precedida cada una de ellas de su correspondiente número entre paréntesis seguido de tabulador. Las citas bibliográficas se realizarán como en el resto del texto; y la referencia completa se incluirá en el apartado de “Referencias”. Las páginas web se consideran referencias bibliográficas y deben tratarse como tales, evitando su cita como nota.

6. Citas bibliográficas en texto

Las referencias bibliográficas en texto se denotarán por el apellido del autor, el año del trabajo y, si es necesario, el número de página. Si conviene, el nombre y el apellido del autor podrán quedar fuera del paréntesis. Si dos autores de la lista de referencias poseen el mismo primer apellido, se referirán por sus dos apellidos y, si tienen ambos apellidos iguales, por el nombre también. Si dos o más trabajos del mismo autor son del mismo año, se distinguirán mediante letras del alfabeto secuenciales en minúsculas pospuestas a la fecha sin espacio. Si hay varias referencias dentro del paréntesis, se separarán por punto y coma si son de diferentes autores, y coma si son del mismo.

Así, por ejemplo, “Es una teoría sustentada por varios autores (Gallego, 1975; Fernández y Alonso, 1993) [...]”. “Según Jaime Pérez (1993, p. 24) se establecen [...]”. “El mismo autor (Pérez, p. 27) hace constar la conveniencia de [...]”. “Se han detectado diferencias en la replicación de la experiencia (Menéndez, 1994a, 1994b; Menéndez y Alonso, 1997)”.

7. Formato de las referencias bibliográficas

Las referencias bibliográficas se incluirán por orden alfabético al final del artículo bajo el título “Referencias”. Deberán seguir la norma UNE 50104:1996/ISO 690:1987, y, en cuanto a la puntuación, aproximarse lo más posible a las ISBD abreviadas, aunque sin espacio delante de punto, coma y punto y coma. Las referencias a páginas web deben incluirse también en este apartado, no en el texto ni en nota. Refiera adecuadamente en el texto del artículo toda la bibliografía utilizada; no incluya en el apartado referencias documentos que no haya incluido en el texto. Se presenta a continuación el formato para los principales tipos de materiales, con los elementos que pueden darse o no en cursivas, y seguidos de un ejemplo.

7.1. Monografías, congresos, informes, normas y tesis

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Tesis doctoral.

Delclaux, Isidoro; Seoane, Julio (1982). *Psicología cognitiva y procesamiento de la información: teoría, investigación y aplicaciones*. Madrid: Ediciones Pirámide.

7.2. Artículos de publicaciones periódicas

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Título de la publicación periódica. ISSN numero. Volumen:número, primera página-última página.

Ellis, David (1992a). The physical and cognitive paradigms in Information Retrieval Research. // *Journal of Documentation*. 48:1, 45-46.

7.3. Capítulos de una monografía, de un informe o contribuciones a un congreso

Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. // Apellido, Nombre; ApellidoN, NombreN (función) (Año). Título: subtítulo. Edición. Lugar de publicación: editor; lugar de publicación: editor. ISBN número. Primera página-última página.

Markey, Karen (1990). Keyword searching in an online catalog enhanced with a library classification. // Bengtson, Betty G.; Hill, Janet Swan (eds.). *Classification of library materials: current and future potential for providing access*. New York: Neal-Shuman Publishers. 99-125.

Smith, Ph. J.; Beghtol, C.; Fidel, R.; Kwasnik, B. H. (eds.) (1993). *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop: Columbus, OH, Oct.24, 1993*. Silver Spring, MD.: American Society for Information Science.

7.4. Páginas web y recursos en línea

Se tratarán como monografías, artículos o capítulos según corresponda y se añadirá al final detrás de punto el URL y la fecha de consulta entre paréntesis:

Referencia. URL (Fecha de consulta).

Sagredo Fernández, Félix; Espinosa Temiño, María Blanca (2000). *Del libro, al libro electrónico-digital. // Cuadernos de Documentación Multimedia*. 9 (2000). <http://www.ucm.es/info/multi doc/multidoc/revista/num9/cine/sagredo.htm> (2001-01-22).

8. Apéndice I. El resumen

Un resumen proporciona información sobre objetivos, alcance, metodología, resultados, conclusiones y recomendaciones. Por alcance se entiende la determinación de hasta qué punto se han tratado los objetivos. Los resultados, conclusiones y recomendaciones se distinguen unos de otros así: los resultados son las informaciones o datos específicos y concretos que se obtienen de la aplicación de la metodología; las conclusiones se derivan de los resultados bien al interpretarlos en el marco de conocimientos teóricos previos o bien al crear un marco de interpretación nuevo a partir de ellos a través de un proceso de abstracción; finalmente, las recomendaciones se refieren a los nuevos problemas y vías de estudio que plantean a la vista de los resultados y conclusiones.

El resumen se debe redactar de forma clara y concisa, sin redundancias ni perífrasis ni información difícil de interpretar, como neologismos o abreviaturas que no se desarrollan.

Es muy conveniente que su primera frase contenga en primer lugar la mención del tipo de documento. Por ejemplo, "Estado de la cuestión sobre la aplicación del modelo europeo de calidad a las bibliotecas universitarias", "Proyecto de aplicación del MARC21 a la catalogación de páginas web institucionales", etc.

9. Apéndice II. La redacción del artículo científico: algunos aspectos obvios frecuentemente descuidados

9.1. Sistemática

El artículo científico debe redactarse de forma muy organizada, con una estructura textual firme. En los informes de resultados de proyectos, se suele utilizar el esquema objetivos-metodología-resultados- conclusiones- recomendaciones. En los estados de la cuestión se aborda el asunto en orden cronológico o bien sistemático, según la ontología propia del área científica en cuestión; sigue la discusión, la síntesis, las conclusiones y las recomendaciones. Los artículos críticos suelen seguir el esquema tesis-síntesis-antítesis. Otros tipos documentales pueden exigir otras estructuras textuales, pero, en cualquier caso, dicha estructura debe ser explícita y aparente.

9.2. Claridad y concisión

El lenguaje del artículo científico debe ser claro y conciso. Excepcionalmente, en los trabajos de tipo ensayístico las personas con un alto dominio del lenguaje escrito pueden permitirse la utilización de recursos literarios.

9.3. Pertinencia y parsimonia

Debe evitarse proporcionar información ya conocida, salvo si constituye el punto central de la discusión. A ella cabe referirse por medio de las citas o, si estas no son necesarias, suponerla patrimonio de la comunidad profesional y científica. Sistematizar conocimiento es una función que se realiza solo en un tipo de artículo científico —la revisión o estado de la cuestión—; en los demás casos es contraproducente, aunque se deben citar las fuentes y las obras clave en la introducción del trabajo.

9.4. Civildad

Es necesario citar a otros autores que han trabajado sobre el tema, demostrando que se conoce la literatura previa y que el trabajo se inserta civilizadamente en un determinado contexto científico.

Índice de autores

Author index

Almeida, Carlos Cândido de, 53
Alonso Berrocal, José Luis, 25
Baptista, Fabricio, 53
Cruz, Maria Carolina Andrade, 63
Ferneda, Edberto, 53
Figuerola, Carlos G., 25
Fujita, Mariângela Spotti Lopes, 63
Gomes, Pedro Henrique Carvalho, 37
Gracioso, Luciana de Souza, 45
Guimarães, José Augusto Chaves, 37

Marcos Recio, Juan Carlos, 13
Martínez-Ávila, Daniel, 37, 45
Martins, Amanda, 37
Moreira, Walter, 63
Olivera Zaldua, María, 13
Sánchez Vigil, Juan Miguel, 13
Santos, Luciana Beatriz Piovezan dos, 63
Silva, Eduardo Graziosi, 45
Zazo Rodríguez, Ángel, 25

Índice de materias en español

Subject index in Spanish

Análisis de dominio, 37
Autor, 45
Autoría, 45
Catálogos en línea, 63
Clasificación automática, 25
Clustering, 25
Comunidades epistémicas, 37
Control de información, 13
Facebook, 13
Foucault, Michel, 45
Interfaces de recuperación de información, 53
Medios de comunicación, 13

Mentira, 13
Organización del conocimiento, 37
Peirce, Charles Sanders, 53
Percepción, 53
Postverdad, 13
Redes sociales, 13, 25
Semiótica, 53
Software de gestión de tesauros, 63
Teoría de la percepción, 53
Tesauros, 63
Wittgenstein, Ludwig, 45

Índice de materias en inglés

Subject index in English

Author, 45
Authority control, 63
Authorship, 45
Automatic classification, 25
Clustering, 25
Domain analysis, 37
Epistemic communities, 37
Facebook, 13
Foucault, Michel, 45
Information control, 13
Information retrieval interfaces, 53
Knowledge organization, 37
Libraries, 63

Lies, 13
Media, 13
Online catalogues, 63
Peirce, Charles Sanders, 53
Perception, 53
Post truth, 13
Semiotics, 53
Social Network Analysis, 25
Social networks, 13
Theory of perception, 53
Thesauri, 63
Thesauri management software, 63
Wittgenstein, Ludwig, 45