

SCIRE

Representación y organización del conocimiento

SCIRE

Representación y organización del conocimiento

Vol. 28, n.º 1, enero-junio 2022

ISSN 1135-3716

ISSN (e) 2340-7042

Scire:

knowledge representation and organization
Vol. 28, n. 1, January-June 2022

Ibersid:

Red de Investigación
en Sistemas de Información
y Documentación

© 2022 Los autores y autoras conservan sus derechos de autor, aunque ceden a la revista de forma no exclusiva los derechos de explotación (reproducción, distribución, comunicación pública y transformación) y garantizan a esta el derecho de primera publicación de su trabajo, el cual estará simultáneamente sujeto a la licencia CC BY-NC-ND. Los autores aceptan la responsabilidad legal de cumplir plenamente con los códigos éticos y leyes apropiadas, y de obtener todos los permisos de derecho de autor debidos. Se permite y se anima a los autores y autoras a difundir electrónicamente la versión editorial (versión publicada por la editorial) en la página web personal del autor y en el repositorio de la institución a la que pertenece.

ISSN: 1135-3716 = Scire (Zaragoza)

ISSN (e): 2340-7042

Depósito legal: Z. 1.790 — 1995

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

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, ESCI, 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.

Instrucciones para los autores y procedimiento de evaluación

La última versión de las instrucciones para presentación de trabajos y del procedimiento de evaluación editorial están disponibles en: <https://www.iberid.eu/ojs/index.php/scire/about/submissions>

Agradecimientos

Agradecemos el apoyo del Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón con su subvención a grupos de investigación S6520D, al Vicerrectorado de Investigación y a la Facultad de Filosofía y Letras 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: 30 €. Número suelto: 20 €. (IVA inc.)

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, ESCI, 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.

Instructions for authors and evaluation process

The last version of the instructions for authors and assessment process is available at: <https://www.iberid.eu/ojs/index.php/scire/about/submissions>

Acknowledgments

We acknowledge the help of the Department of Science, University and Knowledge Society of the Government of Aragon (grant S6520D to research groups), and of the Research Vice Rectorate and the Faculty of Philosophy and Arts 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 La República, Uruguay | João Batista Ernesto Moraes, Universidade Estadual Paulista, Brasil | Daniel Martínez Ávila, Universidad de León, España |
| José Augusto Chaves Guimarães, Universidade Estadual Paulista, Brasil | Francisco Javier García Marco, Universidad de Zaragoza, España | Francisco Javier Martínez Mendez, Universidad de Murcia, España |
| | | Álvaro Quijano Solís, Colegio de México, México |

Consejo científico / Scientific council

| | | |
|--|---|--|
| Dr. Isidro Aguillo Caño, IPP- CSIC, España | Vicente Pablo Guerrero Bote, Univ. de Extremadura, España | José Luis Otal, Universidad Jaume I de Castellón, España |
| Tomás Baiget, EPI S. A., España | Víctor Herrero Solana, Univ. de Granada, España | Manuel José Pedraza Gracia, Universidad de Zaragoza, España |
| José Luis Bonal Zazo, Univ. de Extremadura, España | José María Izquierdo Arroyo, Universidad de Murcia, España | María Pinto Molina, Universidad de Granada, España |
| Mercedes Caridad Sebastián, Universidad Carlos III de Madrid, España | María Pilar Lasala Calleja, Universidad de Zaragoza, España | Gloria Ponjuán Dante, Universidad de La Habana, Cuba |
| Alberto Carreras Gargallo, Universidad de Zaragoza, España | Alfonso López Yepes, Universidad Complutense de Madrid, España | Blanca Rodríguez Bravo, Universidad de León, España |
| Constança Espelt Busquets, Universidad de Barcelona, España | José López Yepes, Universidad Complutense de Madrid, España | José Vicente Rodríguez Muñoz, Universidad de Murcia, España |
| Juan Carlos Fernández Molina, Univ. de Granada, 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 Urdiciáin, Universidad Complutense de Madrid, España | Purificación Moscoso Castro, Universidad de Alcalá, España | Félix Sagredo Fernández, Universidad Complutense de Madrid, España |
| Isidoro Gil Leiva, Universidad de Murcia, España | Félix Moya Anegón, Universidad de Granada, España | Elías Sanz Casado, Universidad Carlos III de Madrid, España |
| Alan Gilchrist, Cura Consortium, Reino Unido | María del Carmen Negrete Gutiérrez, Universidad Autónoma de México | Carlos Serrano Cinca, Universidad de Zaragoza, España |

Revisores externos del número / External reviewers in this issue

Agradecemos enormemente la colaboración desinteresada de Manuel Blázquez Ochando, Fernanda Farinelli, Javier Lacasta Miguel, Daniel Martínez-Ávila, María Jesús Martínez Pestaña, Suellen Oliveira Milani, Juan Antonio Pastor Sánchez y Juan Francisco Torregrosa Carmona, así como de otros dos revisores cuyo anonimato se preserva.

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

Tabla de contenidos en español.....9

Tabla de contenidos en inglés.....11

Artículos

*¿De qué se habla cuando se habla
de homeopatía en la web?:
un análisis de contenido*

María-José BAÑOS-MORENO
Eduardo R. FELIPE
Zuriñe PIÑA-LANDABURU,
Maurício ALMEIDA13

*Representación temática del material
audiovisual en repositorios multimedia
de instituciones federales brasileñas:
un estudio de caso con un concepto
paradigmático*

Suelene Santana ASSUNÇÃO
João Vitor Correa DINIZ
Franciele Marques REDIGOLO
Mariângela Spotti Lopes FUJITA.....23

*Voces cultivadas versus tipos de voces
en la CDD: el uso de la metateoría
para la comprensión de la voz cantada*

Fernanda Carolina PEGORARO NOVAES33

*El uso de lenguajes de indización
en las bibliotecas universitarias:
mapeo sistemático de la literatura*

Jessica Beatriz TOLARE
Mariângela Spotti Lopes FUJITA45

*Desarrollo de una taxonomía interoperable
a partir del mapeo de taxonomías empíricas:
un estudio de caso con los sitios web de las
asociaciones de amigos del Camino de Santiago*

Francisco Javier GARCÍA MARCO
Joaquín LÓPEZ DEL RAMO.....55

Índice de autores74

Índice de materias en español.....74

Índice de materias en inglés.....74

Tabla de contenidos en inglés

Table of contents in English

| | |
|---|----|
| <i>Table of contents in Spanish</i> | 9 |
| <i>Table of contents in English</i> | 11 |

Articles

*What do we mean when we speak
about homeopathy on the web?:
a content analysis*

María-José BAÑOS-MORENO
Eduardo R. FELIPE
Zuriñe PIÑA-LANDABURU,
Maurício ALMEIDA 13

*Thematic representation of audiovisual
material in multimedia repositories
of Brazilian federal institutions: a case study
with a paradigmatic concept*

Suelene Santana ASSUNÇÃO
João Vitor Correa DINIZ
Franciele Marques REDIGOLO
Mariângela Spotti Lopes FUJITA 23

*Cultivated voices versus types of voices
in the DDC: the use of metatheory
for the understanding of the sung voice*

Fernanda Carolina PEGORARO NOVAES 33

*The use of indexing languages by university
libraries: systematic literature mapping*

Jessica Beatriz TOLARE
Mariângela Spotti Lopes FUJITA 45

*Developing an interoperable taxonomy
by mapping empiric taxonomies: a case study
on the websites of the associations of friends
of the Way of St James*

Francisco Javier GARCÍA MARCO
Joaquín LÓPEZ DEL RAMO 55

Author index 74

Subject index in Spanish 74

Subject index in English 74

¿De qué se habla cuando se habla de homeopatía en la web?: un análisis de contenido

What do we mean when we speak about homeopathy on the web?: a content analysis

**María-José BAÑOS-MORENO (1), Eduardo R. FELIPE (2),
Zuriñe PIÑA-LANDABURU (3), Mauricio ALMEIDA (4)**

(1) Facultad de Comunicación y Documentación, Universidad de Murcia, Campus de Espinardo, s/n 30100 Murcia, mbm4963@um.es (2) Instituto de Engenharia de Sistemas e Tecnologia da Informação, Universidade Federal de Itajubá, Rua Irmã Ivone Drumond Distrito Industrial II, 35903087 - Itabira, MG – Brasil, eduardo.felipe@unifei.edu.br (3) Facultad de Ciencias de la Documentación, Universidad Complutense de Madrid, C. de la Santísima Trinidad, 28010 Madrid, zuri-ne.p.l@gmail.com (4) Escola de Ciência da Informação, Universidade Federal de Minas Gerais (UFMG), Av. Antônio Carlos, 6627, Pampulha, 31270-010 - Belo Horizonte, MG – Brasil, mba@eci.ufmg.br

Resumen

Se realiza un análisis del contenido de las páginas web que tratan sobre la homeopatía con el objetivo de determinar su tendencia, tipología, dominios más destacados y términos más utilizados. Para ello, se han identificado los términos más representativos del ámbito y se han seleccionado un grupo de semillas. Ambos elementos son el punto con que arranca “Crawler by domain”, una aplicación desarrollada para la recolección de páginas del dominio. Los resultados muestran que la inmensa mayoría de páginas del sector tienen una visión positiva de la homeopatía. Algo que es lógico teniendo en cuenta que gran parte tiene como fin la venta de productos y/o servicios de homeopatía o se tratan de portales especializados en ésta. Como conclusión general, la tendencia de las fuentes a ofrecer un contenido con sesgo positivo y fácilmente comprensible por el usuario medio, junto con la relativa escasez de páginas con un sentido crítico o incluso sin tendencia puede ser un factor que impulse a los usuarios a inclinarse por la utilización de esta pseudoterapia, ya que es posible que sea interpretada como prueba efectiva de sus beneficios.

Palabras clave: Homeopatía. Análisis de contenido. Web. Crawler by domain. Crawlers.

1. Introducción

Desde hace un tiempo, se da la paradoja de que se genera más información que nunca y, al mismo tiempo, crecen las dificultades para acceder a fuentes de información fiables. Prueba de ello es el auge de las redes sociales como puerta de acceso de información o la popularización de los llamados contenidos virales que, a menudo, pueden dar acceso a contenidos poco fiables, manipulados o directamente falsos. En muchas ocasiones, el usuario no atiende a la calidad de sus fuentes. Esto es clave para el éxito de las llamadas noticias falsas o “fake news”, que son un elemento fundamental para influir en la opinión

Abstract

An analysis of the content of the web pages that deal with homeopathy is carried out. The aim is to determine their trends, typology, most prominent domains, and most used terms. To do this, the most representative terms in the field have been identified and a group of seeds has been selected. Both elements are the starting point for “Crawler by domain”, an application developed to collect web pages. The results show that many pages in the sector have a positive view of homeopathy. That is logical considering that a large part has as its purpose the sale of homeopathy products and/or services or they are specialized portals in it. As a general conclusion, the tendency of these sources to offer content with a positive bias and easily understood by the average user, together with the relative scarcity of pages with a critical sense or even without bias, may be a factor that encourages users to lean towards for the use of this pseudotherapy, since it is possible that it is interpreted as effective proof of its benefits.

Keywords: Homeopathy. Web. Content analysis. Crawler by domain. Crawlers.

pública e incluso explican el triunfo de ciertos movimientos políticos (Alcott y Gentzklow, 2017).

La información poco contrastada no es un fenómeno nuevo. Es parte de la cultura popular, siendo el origen de historias basadas en la literatura oral. Desde el chisme o la historia que antes corría de boca en boca, hasta los actuales contenidos “virales” que llegan a millones de personas en un breve periodo de tiempo, estamos ante algo que, simplemente, ha aprovechado las ventajas de las Tecnologías de la Información para expandirse. Como un organismo vivo, la información poco contrastada ha evolucionado adaptándose al entorno.

Se trata de un fenómeno que, en muchas ocasiones, no tiene mayor relevancia al margen de su interés sociológico o cultural. Sin embargo, esto cambia cuando comienzan a extenderse ideas y/o prácticas que pueden resultar de riesgo tanto para uno mismo como para terceros o, sin llegar a ese punto, cuando el ciudadano tiene acceso a información que, al no tener una base científica real, puede llevarlo a abandonar terapias con evidencia médica en favor de otras basadas en pseudociencias. Uno de los ejemplos más claros es el auge de las medicinas alternativas, como la homeopatía, del griego *homoiós* (igual) y *pathos* (sufrimiento), en la que se centra este trabajo como objeto de estudio.

La homeopatía surge en 1796, de manos del doctor alemán Samuel Hahnemann, basándose en la idea de que, si una sustancia causa en personas sanas los síntomas de una enfermedad, esa sustancia curaría estos síntomas en personas enfermas (Nayernouri, 2017). Es lo que se conoce como la Ley de los similares. El otro principio en el que se asienta esta pseudociencia es la Ley de los infinitesimales, según la cual, un remedio se vuelve más efectivo al diluirse (Young, 2014). La práctica de la homeopatía se extendió a lo largo del siglo XIX en Europa y cayó en desuso a principios del siglo XX. Hasta que, en los años 70, volvió de forma significativa (Nayernouri, 2017) como terapia alternativa, casi cien años después.

La creciente concienciación del consumidor y la presión ejercida por los sectores críticos con las pseudociencias han hecho que la Real Academia de la Lengua Española haya enmendado la acepción de este concepto, que ha pasado de ser definido como un sistema curativo a una mera práctica que “consiste en administrar a alguien, en dosis mínimas, las mismas sustancias que, en mayores cantidades, producirían supuestamente en la persona sana síntomas iguales o parecidos a los que se trata de combatir” (Redacción Médica, 2019). No es la única institución, el Ministerio de Sanidad continúa revisando la categorización de otras pseudoterapias por carecer de evidencia científica, tales como la magnetoterapia estática o la dieta macrobiótica (Europa Press, 2021).

En esta investigación se combina el uso de técnicas propias de las Ciencias de la Documentación, como la indización y el análisis de contenido, y de la Informática, mediante el diseño e implementación de un rastreador que analiza el contenido de páginas web relacionadas con la homeopatía.

El crawler es un tipo de agente computacional (Chowdhury, 2004) que se comporta como un usuario (buscador) y visita y descarga el contenido de páginas web para su procesamiento posterior (Chakrabarti, 2003). Este tipo de herramienta ya

ha sido empleada antes para la recuperación de información filtrada para el análisis o la creación de bases de conocimiento específicas. Además, existen diversas técnicas para la realización de estos programas, como Chakrabarti et. al (1999), que empleó técnicas de aprendizaje automático y taxonomías. Por su parte, Safran et. al (2012) hizo uso de términos próximos a las URLs relevantes, marcadas inicialmente como semillas, para construir un conjunto de entrenamiento para el aprendizaje automático. Hegade et. al (2021) trabaja con un modelo de inferencia contextual, basado en similitudes y reglas de inferencia.

Por otro lado, De Groc (2011) propone un modelo de categorización de páginas web a partir de un léxico basado en filtros temáticos. Aggarwal (2019) hace uso del aprendizaje automático basado en un conjunto de premisas que pueden incluir URLs, páginas padre, textos de anclaje y textos próximos a los elementos relevantes. Dahiwale (2014) adopta un modelo de análisis de contenido basado en determinadas etiquetas HTML, con el fin de computar la relevancia de la página. Bedi et al. (2012) trabaja con un mecanismo multihilo asociado a las ontologías de dominio. Farag et al. (2018) estudia el concepto de eventos para filtrar datos específicos. Du et al. (2015) elabora un camino de similitud a partir de un modelo vectorial.

2. Justificación y objetivos

El objetivo general de este trabajo es realizar un análisis de contenido de las páginas web relacionadas con la homeopatía con el propósito de conocer de qué tratan las páginas web relacionadas con la homeopatía y su intencionalidad. De manera más específica, se pretende realizar el seguimiento de páginas y recuperación de información que permita determinar qué términos son empleados para representar los conceptos que giran en torno a esta pseudociencia.

Para satisfacer los aspectos indicados, se ha desarrollado un rastreador, “Crawler by domain”, que emplea una estrategia centrada exclusivamente en una lista de términos del área de la homeopatía, configurada por documentalistas. En cada visita del crawler, consulta un vocabulario controlado con el fin de identificar uno o más términos que determinen la posible relevancia del contenido en relación con la temática analizada.

3. Metodología

La investigación arrancó con un análisis bibliográfico y documental con el propósito de obtener una primera lista de términos asociados con la temática. Se recurrió a bibliografía especializada en homeopatía disponible en Web of Science

(WoS) y Google Scholar. También se empleó la taxonomía DeCS, como instrumento terminológico validado. Con ello, se obtuvo una lista de las palabras clave más representativas del resumen de cada publicación, que dieron lugar a un vocabulario controlado. Este producto es empleado en dos ocasiones en este trabajo.

```

60 def processUrl(url):
61     if urlNotVisited(url):
62         page = contentFromUrl(url)
63         keywordsFound = (keywordsInDocument(page))
64         allURLs = utils.allURLsFromDocument(page)
65         # allURLs = utils.allURLsFromPage(page)
66         if (len(keywordsFound) > 0):
67             print(colors.bcolors.OKGREEN + "% URL added: ", url + colors.bcolors.ENDC)
68             utils.insertDataInDB(url, page, keywordsFound)
69         else:
70             print(colors.bcolors.FAIL + "% URL without keywords: ", url + colors.bcolors.ENDC)
71             print(colors.bcolors.OKCYAN + "% URL already visited: ", url + colors.bcolors.ENDC)
72
73 def queueURLsToVisit(append(url):
74     url = queueURLsToVisit(url)
75     allURLs.clear()
76
77 def processQueue(counter):
78     """ Process the queue with seeds and prior address """
79     while (counter < consts.VISIT_LIMIT) and (not queueURLsToVisit.empty()):
80         counter += 1
81         url = queueURLsToVisit.get()
82         if (utils.isValidUrl(url) and (url is not None)):
83             print(colors.bcolors.OKBLUE + "% Procesando: " + str(counter) + " - " + url + color
84             processURL(url)
85         else:
86             print(colors.bcolors.WARNING + "% Archivo inválido: " + url + colors.bcolors.ENDC)
87     utils.saveQueueToVisit(queueURLsToVisit)
88     utils.saveURLsVisited(queueURLsVisited)
89
90 if __name__ == "__main__":
91     utils.initialize(queueURLsToVisit)
92     keywordsList = utils.loadFileLikeArray(consts.ARQ_KEYWORDS)
93     queueURLsVisited = utils.loadURLsVisited(consts.ARQ_DATABASE)
94     processQueue(counter)
95     utils.exportDatabaseToLxml(consts.ARQ_DATABASE)
96     utils.countKeywordsFromDB()
97     utils.makeCloudOfWords()
    
```

Figura 1. Código fuente en que se muestra uno de los principales procesos del programa

Así, fue empleado en diversas búsquedas en Google, con el propósito de obtener las páginas-semilla que se utilizarán después en el Crawler. Se seleccionaron los 10 primeros resultados de cada búsqueda realizada, descartando aquellas URLs que no estaban relacionadas con el tema de estudio o no resultaban útiles para el propósito del Crawler. Por ejemplo, diccionarios en línea no especializados (The Free Dictionary o WordReference), así como repositorios o bases de datos especializadas en información científica, como Elsevier. Además, varias páginas tenían un origen en común y fueron contabilizadas como un resultado único. Estos términos de búsqueda son empleados después en el crawler desarrollado.

Simultáneamente, se desarrolló "Crawler by domain", con el fin de permitir filtros específicos para la captación de páginas web de la temática escogida. Un esbozo del funcionamiento de este software se muestra en la Figura 1, donde se puede visualizar un proceso importante del programa para su funcionamiento. Esta herramienta permite utilizar las palabras clave detectadas inicialmente como representativas del tema "homeopatía" y las páginas-semilla obtenidas para recolectar otras páginas web del dominio analizado.

Finalmente, se analizó la información obtenida por el software por documentalistas, excluyendo miles

de materias diversas que no estaban relacionadas con la temática propuesta. Así, se revisaron las palabras clave del total de páginas web captadas con el crawler, el tipo de página web y algunas de sus características, clasificándolas por su intencionalidad (divulgación, formativa, venta, etc.).

3.1. Crawler by domain

Los rastreadores son un software especializado que permite la recuperación automática de páginas web empleando un mecanismo de filtro para seleccionar sólo aquellas páginas web del dominio analizado. En otras palabras, a partir de un sitio y sus hipervínculos, se genera una lista de visitas para los demás sitios, hasta que un criterio de parada es satisfecho (Yu, 2011).

La herramienta "Crawler by domain" (<https://github.com/erfelipe/crawlerByDomain>), es sucesora de MetadadosHTML, un programa para la extracción de metadatos de páginas web (Baños-Moreno et al., 2017). Se trata de una nueva herramienta, escrita en Python (versión 3), orientada a la recolección del código fuente de páginas web especializadas en una temática para, entre otros, extraer datos, como el título o las palabras clave. Entre las diferencias que encontramos con su antecesor, destaca su comportamiento: visitar páginas web a partir de una lista de URL iniciales (semillas), filtrando la recolección de páginas a partir de un vocabulario controlado propio del tema de estudio, eliminado así horas de evaluación humana. Además, el programa es capaz de mostrar al usuario toda la información recolectada y facilita unos resultados más pertinentes. En la Figura 2 se resume su funcionamiento.

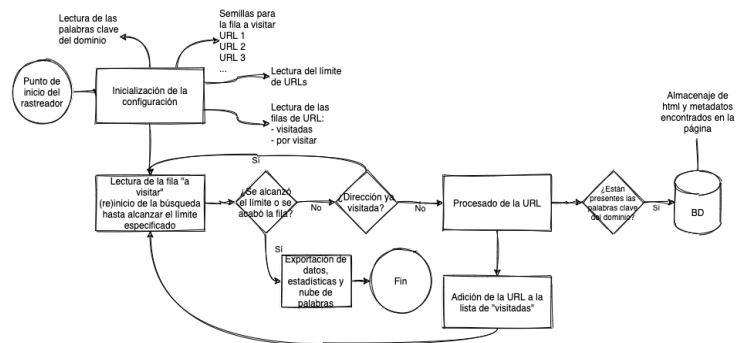


Figura 2. Funcionamiento de la herramienta

Aunque del software se haya incidido particularmente en el modelo de crawler, su parte analítica es similar a un programa orientado al web scraping. Esta técnica permite extraer datos mediante algoritmos específicos con el fin de identificar información relevante para el contexto que se estu-

dia. Este proceso hace posible que, mediante análisis computacional, se pueda analizar grandes volúmenes de datos de forma estandarizada en escenarios donde la evaluación humana se demoraría durante mucho tiempo y, además, podría conducir a errores de interpretación o de análisis.

En el programa, se emplearon diversas bibliotecas de software (documentadas en el github del proyecto), destacando: 1) *requests* para solicitar el contenido de la URL y permitir el análisis de código fuente local; 2) *urllib* para verificar el código fuente de las páginas, con el propósito de alimentar la fila de visitas con las urls de la página visitada; 3) *bs4* (*beautifulsoup*) es una biblioteca multifuncional para el parseo y análisis de código html. En todo momento, se procuró escribir el código siguiendo las buenas prácticas del "Clean Code" (Martin, 2009).

En cuanto a sus componentes, destacan los siguientes ficheros:

1. *consts.py*: permite indicar dónde se ubican los archivos de las URLs de las semillas iniciales, la lista de palabras clave y la ruta en la que se deben guardar los ficheros resultantes. Por defecto, todos los archivos estarán en el mismo directorio del programa. También permite definir el límite de visitas para el procesamiento del crawler (en la constante VISIT_LIMIT), que por defecto será de 1.000 páginas recuperadas y el número de grabación parcial (en la constante RECORD_EACH), que permite detener la recolección en un momento dado y retomarla posteriormente.
2. *urls-seeds.txt*: en este fichero se especifican las semillas. El proceso recolecta las direcciones a las que apuntan estas URLs. El proceso se repite con cada nueva colección de páginas, que entran en una cola de espera para su análisis. Así hasta que se detenga de forma activa o se alcance el límite especificado.
3. *keywords.txt*: es la lista de palabras clave que enriquece la recolección. Este fichero, si bien no es obligatorio, es muy recomendable. Su inclusión en la actividad del crawler evita analizar páginas descontextualizadas, que no tienen nada que ver con el tema de estudio. El rastreador busca esos términos en el contenido de cada página web. Si no encuentra ninguno, considera que es ruido y lo rechaza. Para asegurar un buen funcionamiento del proceso, las palabras clave, en el caso de que utilicemos terminología en castellano y/o en portugués, deben introducirse con y sin tildes.
4. *black-list-domain.txt*: es una lista editable de términos que harán que una URL sea desechada por incluir alguno de ellos. Por

ejemplo, se puede eliminar todas las referencias a *youtube*, *facebook*, *pinterest*, *twitter*, *linkedin*, *github*, entre otras.

A continuación, se describe cómo se configura y funciona la aplicación.

3.1.1. Configuración y puesta en marcha de Crawler by domain

Antes de iniciar su primera iteración, el usuario debe especificar las semillas (en el archivo *url-seeds.txt*) y el vocabulario especializado en la temática que se estudia (en *keywords.txt*). También se puede definir el número de visitas que realizar en la constante VISIT_LIMIT del fichero *consts.py*.

Una vez instaladas las bibliotecas por el administrador PIP y configurados los archivos anteriores, se puede lanzar el Crawler. Para ello, basta con escribir en la línea de comando: *Python crawlDomain.py*. Éste puede funcionar durante el tiempo que consideremos, pero, como mínimo, para que recolecte una cantidad relevante de páginas, tendría que alcanzar las 3.000 URLs de contexto. El inicio del procesamiento permite que se generen una base de datos relacional *sqlite* y ficheros relacionados.

Si es un proceso nuevo, se partirá de las semillas definidas por el usuario. Si se detiene el procesamiento de páginas, después se puede continuar con la recolección respondiendo "N" a la pregunta de construir un nuevo proceso. Las visitas continúan a partir de la última URL tratada, dando continuidad a iteraciones pasadas y permitiendo agregar más páginas al conjunto. Podemos hacerlo en momentos diferentes, con el propósito de incrementar el número de páginas durante un periodo.

Para cada página web visitada se verifica si posee al menos un término del vocabulario controlado entre las palabras clave (etiqueta `<meta keywords>`), en el cuerpo (`<body>`) o en el título (`<title>`) del código fuente. En caso afirmativo, se procesa la web, extrayendo sus metadatos y almacenándolos en el banco de datos junto con su contenido. En caso contrario, la página es descartada, aunque la URL es almacenada de cara a futuros accesos. La iteración continúa hasta que la fila que se procesa quede vacía (aunque es improbable que esto ocurra), o bien se alcance el número de visitas anteriormente definido como límite.

Al acabar, el programa consulta la base de datos generada y exporta su contenido en un documento con formato de hoja de cálculo, que facilita el análisis de los datos obtenidos, que incluye las páginas filtradas y diversos metadatos (se indican en la Tabla I).

| Metadato | Descripción |
|----------------|---|
| domain | Dominio base de la dirección analizada |
| url | Dirección completa analizada |
| keys | Palabras clave del vocabulario recuperadas de la página analizada |
| quantKeys | Cantidad de palabras clave encontradas con respecto al vocabulario controlado |
| keysFromPage | Palabras clave de la página, declaradas en "meta keywords" |
| langFromPage | Idioma de la página, declarado en "meta lang" o "language" |
| descFromPage | Descripción de la página, declarado en "meta description" |
| authorFromPage | Autor de la página, declarado en tag "meta author" |

Tabla I. Metadatos proporcionados por "Crawler by domain"

| ID | Referencia bibliográfica |
|----|---|
| 1 | Bornhöfta, Gudrun; Wolf, Ursula; Ammon, Klaus von; et al. (2006). Effectiveness, Safety and Cost-Effectiveness of Homeopathy in General Practice – Summarized Health Technology Assessment. // Research in Complementary Medicine. 13 (2), 19-29. pp. https://www.ncbi.nlm.nih.gov/pubmed/16883077 |
| 2 | Cano-Orón, Lorena; Mendoza-Poudereux, Isabel; Moreno-Castro, Carolina (2018). Perfil sociodemográfico del usuario de la homeopatía en España. // Atención Primaria. 1615. https://doi.org/10.1016/j.aprim.2018.07.006 |
| 3 | Descriptores en Ciencias de la Salud: DeCS. ed. 2018. São Paulo: BIREME / OPS / OMS [actualizado en abril de 2018]. http://decs.bvsalud.org/E/homepagee.htm |
| 4 | Ernst, Edzard (2018). Homeopathy: what does the "best" evidence tell us? // Systematic review. 192:8, 458-460. https://www.ncbi.nlm.nih.gov/pubmed/20402610 |
| 5 | López Espinosa, José Antonio (1999). Notas para la historia de la homeopatía. Revista Cubana de Medicina General Integral. 1999. 15:5, 587-590. http://www.bvs.sld.cu/revistas/mgi/vol15_5_99/mgi17599.htm |
| 6 | Nayernouri, Touraj (2017). Homeopathy, Ritual and Magic. Archives of Iranian Medicine. 20:11, 718-722. http://www.aimjournal.ir/PDF/aim-1701 |
| 7 | Ochoa Ortega, Max Ramiro (2018). Análisis sobre la homeopatía como ciencia o pseudociencia. // Archivo Médico Camagüey. 22:3, 381-392. http://ref.scielo.org/gnx7g9 |
| 8 | Young, Pablo (2014). La farsa de la homeopatía. // Revista Médica de Chile. 142:2. http://dx.doi.org/10.4067/S0034-98872014000200021 |

Tabla II. Bibliografía analizada para la selección de términos relacionados con la homeopatía

La herramienta también puede proporcionar el número total de palabras clave y una nube de palabras en formato gráfico.

4. Resultados

Se muestran los resultados obtenidos, organizados por bloques para facilitar su interpretación.

4.1. Obtención de la lista de palabras clave

Como se indicaba en la metodología, primero se realizó una revisión bibliográfica relacionada con la homeopatía. Se recoge la selección de publicaciones obtenida tras la búsqueda bibliográfica en WoS y Google Scholar, así como la taxonomía DeCS en la Tabla II. Esta búsqueda fue realizada en enero y marzo de 2019. En segundo lugar, mediante indización humana se extrajo de la bibliografía anterior una lista de palabras clave relacionadas con la homeopatía. Se indican a continuación, estos términos, organizados por categorías:

- Términos genéricos: Medicina alternativa, Pseudociencia, Medicina natural, Terapias alternativas, Terapias complementarias, Práctica médica alternativa, Sistema terapéutico.
- Términos especializados: Nosode.
- Términos de ámbito organizativo: Industria homeopática, Boiron, COFENAT, Homeopathy Research Institute, Observatorio de Terapias Naturales, Swiss Association of Homeopathic Physicians, SAHOP/SVHA.
- Personas: Samuel Hahnemann (creador); Homeópata (profesional).
- Principios de la homeopatía: Ley de los similares, Ley de los infinitesimales, Agua polimerizada, Memoria del agua, Dilución homeopática, Número de Avogadro, Constante de Avogadro, Moléculas activas.
- Relativos a tratamientos homeopáticos: Prueba homeopática, Repertorios homeopáticos, Homeopathic Materia Medica, Productos homeopáticos, Tratamientos homeopáticos, Medicina homeopática, Remedios homeopáticos.

Estas relaciones son importantes para conocer con mayor profundidad los elementos relacionados con la homeopatía. Así, por ejemplo, podemos ver en qué principios está basada o qué tratamientos pueden establecer sus especialistas. También permiten determinar cuál es su marco general como pseudociencia o terapia alternativa.

4.2. Obtención de las semillas

Se realizaron las siguientes búsquedas en Google, a partir de los términos recogidos en el apartado anterior: "Homeopatía"; "Nosode"; "Samuel Hahnemann"; "Industria homeopática y Boiron"; "Industria homeopática y COFENAT"; "Industria

homeopática y Homeopathy Research Institute”; “Industria homeopática y Observatorio de Terapias Naturales”; “Industria homeopática y Swiss Association of Homeopathic Physicians”; “Industria homeopática y SAHOP/SVHA”; “Miasmas”; “Ley de los similares”; “Ley de los infinitesimales”; “Agua polimerizada”; “Memoria del agua”; “Dilución homeopática”; “Número de Avogadro”; “Constante de Avogadro”; “Prueba homeopática”; “Repertorios homeopáticos”; “Homeopathic Materia Medica”; “Productos homeopáticos”; “Tratamientos homeopáticos”; “Medicina homeopática”; “Remedios homeopáticos”; “Homeópata”.

Estas búsquedas fueron lanzadas entre junio y julio de 2020. Los términos genéricos, como “Medicina alternativa”, fueron excluidos de las búsquedas por ser menos específicos que “homeopatía”. Tras el análisis de los datos, se consiguieron 12 semillas distintas relacionadas con el tema de estudio (250 en bruto). Como requisito se impuso que el dominio de las URLs debía aparecer como mínimo en tres consultas diferentes (ver la Tabla III).

| ID | Página semilla y búsqueda en que aparece |
|-----|--|
| 001 | https://cuidateplus.marca.com (01, 18, 19, 21, 22, 23) |
| 002 | http://semh.org (01, 22, 23, 25) |
| 003 | https://www.boiron.es (01, 04, 15, 21, 22, 24, 25) |
| 004 | http://queeslahomeopatia.com/ (01, 18, 21, 22, 23, 24, 03, 15) |
| 005 | https://elpais.com/ (01, 18, 23, 04, 09, 05, 07, 18, 01, 09) |
| 007 | http://www.homeopatia.net/ (01, 15, 18, 22, 24, 25) |
| 016 | https://es.wikipedia.org (03, 04, 10, 13, 14, 15, 16, 17, 18, 19, 20, 23, 25) |
| 028 | https://www.redaccionmedica.com/ (04, 05, 18, 21) |
| 032 | https://www.elmundo.es/ (04, 09, 18) |
| 033 | https://www.eldiario.es/ (04, 05, 09) |
| 148 | https://www.portalfarma.com/ (21, 22, 23, 24) |
| 159 | https://www.hablandodehomeopatia.com/ (22, 23, 25) |

Tabla III. Semillas y búsquedas en que aparecen

Se desecharon sitios de venta de productos con escasa repercusión, blogs basados en la opinión particular de personas, etc. Un caso excepcional es Boiron (<https://www.boiron.es>), un laboratorio dedicado exclusivamente a la venta de productos homeopáticos, que apareció en casi un tercio de las búsquedas (7/25) y que indica, sin dudas, un buen posicionamiento en Google, pero también la especificidad de los términos empleados por dicho sitio web. Posteriormente, con el objetivo de incrementar el número de semillas, se utilizó el propio crawler hasta recolectar 238 URLs que,

tras su revisión, fueron incorporadas al archivo url-seeds.txt.

4.3. Recolección de páginas sobre homeopatía

Con la lista de palabras clave y semillas obtenidas en los procesos descritos en los apartados 4.1 y 4.2 respectivamente, se configuró la aplicación Crawler by domain para el rastreo de páginas web sobre homeopatía.

El rastreador estuvo funcionando durante varias horas en distintos periodos de tiempo, entre el 19 y el 23 de diciembre de 2021, visitando un total de 141.702 páginas. 5.108 páginas fueron procesadas como válidas para el objeto de estudio, descartando todas aquellas que no tuvieran al menos uno de los términos recogidos en el archivo keywords.txt del crawler en su código fuente (palabras clave, título o cuerpo). El producto resultante fue exportado en formato Excel.

Tras un análisis inicial, se desecharon: páginas sin acceso, de los que no se puede acreditar su contenido; páginas repetidas con distintas casuísticas (URLs recortadas; páginas móviles aceleradas; versiones para impresión, sin certificado de seguridad, etc.); referencias a esquemas, como XML o JSON; agregadores de contenido (*feeds*); información de suscripciones web; avisos de cookies; recuperación de contraseñas; gestión de cuentas de usuario; avisos legales; contacto; políticas de privacidad, carritos de compra o listas de deseos vacías; listas de deseos vacías; búsquedas sin resultados; páginas de bancos de imágenes, etc.

Además, se encontraron dos tipos de falsos positivos: páginas con el término “homeopatía” (o cualquiera de los identificados durante la indización humana) en un enlace, pero no en el propio contenido; páginas con palabras clave polisémicas, como “memoria del agua”, que nada tienen que ver con el área en que se centra este trabajo.

Finalmente, la muestra quedó compuesta por 2715 páginas recolectadas unívocas.

4.4. Análisis de los datos obtenidos

4.4.1. Tendencia de la página web

Por tendencia se hace referencia a la consideración que se hace de la homeopatía en la página web. En la Figura 3 se identifican las siguientes tendencias:

1. Positiva: corresponde a páginas en las que se habla bien de la homeopatía, ya sea con intención de vender un producto homeopático, ofrecer un servicio o simplemente convencer

de sus beneficios). Se atribuye esta tendencia a 2238 páginas (82,43%).

2. Negativa: se refiere a aquellas páginas que analizan la homeopatía desde un punto de vista crítico y centrado en la exposición de problemas asociados al lector común. Constituyen el segundo tipo de fuente más abundante (15,69%), aunque muy lejos de la anterior.
3. Sin inclinación: estas páginas se centran prácticamente en un aporte histórico de sus inicios o la descripción de algunos de sus referentes básicos, como qué es la ley de los similares, por ejemplo (1,88%).

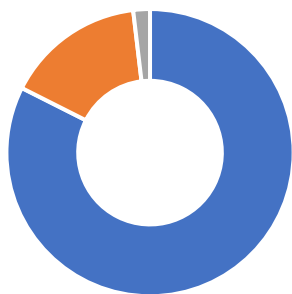


Figura 3. Páginas web según la intencionalidad de sus contenidos

Los datos muestran una clara tendencia de las fuentes a ofrecer una visión de la homeopatía como una terapia con visos de fiabilidad. Teniendo en cuenta que su público objetivo son los usuarios con un conocimiento medio o bajo de medicina, pueden ser un punto de apoyo para el crecimiento que la homeopatía ha vivido en las últimas décadas.

4.4.2. Tipos de página

La tipología de las páginas es diversa, desde aquellas dedicadas a la venta hasta páginas de asociaciones de distintos colectivos.

La Tabla IV agrupa las páginas analizadas por tipos. Los tipos de la tabla anterior fueron identificados *ad hoc* para este estudio.

Destacan muy especialmente los contenidos enfocados a la venta de productos y/o la oferta de servicios propios de la homeopatía. De hecho, constituyen el 50% de las URLs analizadas. En una cantidad bastante inferior (17,5%), también se recolectaron páginas de portales especializados en homeopatía, dedicadas principalmente a informar de sus características, de los beneficios que aportaría, de sus bases, etc.

El tercer tipo de página es el blog (8,18%), generalmente de tipo personal y casi siempre orientado a proporcionar una opinión sobre la homeopatía en un sentido negativo (96%).

También se detectó que un 6,7% del total de páginas se centraban en la formación en homeopatía y pseudoterapias relacionadas, proporcionada por instituciones de todo tipo (universidades, institutos, etc.). Se ha considerado interesante separar este dato de la oferta de servicios, aunque ciertamente, se podría haber considerado un servicio más. Que casi una décima parte de las URLs estudiadas se centren en formar homeópatas o profesionales afines y otro 50% en la venta y servicios da una idea del volumen de negocio de este tipo de producto.

| Tipos de páginas | Frec. | % |
|---|-------|-------|
| Venta de productos y prestación de servicios | 1359 | 50,05 |
| Portal especializado en homeopatía | 476 | 17,53 |
| Blog | 222 | 8,18 |
| Entidad educativa | 182 | 6,70 |
| Prensa generalista, deportes y economía | 128 | 4,71 |
| Asociación de profesionales de homeopatía o terapias alternativas | 98 | 3,61 |
| Publicación o sitio de información científica | 54 | 1,99 |
| Prensa especializada en medicina y salud | 52 | 1,92 |
| Prensa especializada en ciencia, cultura y tecnología | 40 | 1,47 |
| Portal especializado en medicina, farmacia y/o salud | 24 | 0,88 |
| Otro tipo de página | 22 | 0,81 |
| Portal de información pública oficial | 16 | 0,59 |
| Asociación de profesionales de farmacia | 15 | 0,55 |
| Otro tipo de asociación | 8 | 0,29 |
| Asociación de empresas de homeopatía o terapias alternativas | 8 | 0,29 |
| Asociación de profesionales de medicina | 6 | 0,22 |
| Venta de otro tipo de productos | 4 | 0,15 |
| Prensa especializada en terapias alternativas | 1 | 0,04 |

Tabla IV. Tipología de páginas recolectadas

La siguiente figura muestra los datos de la Tabla IV, agrupando en bloques mayores algunos de los tipos detectados. Así, es interesante observar que algo más de un 8% de las URL recuperadas (221 ítems) corresponden a noticias publicadas en prensa generalista o especializada, formando parte de esa variedad de temas que cubren los medios de comunicación (Baños-Moreno, 2017). Esto indica que el tema objeto de estudio constituye una fuente de interés para la sociedad. Casi

un 79% de las noticias tenían una tendencia negativa, desmitificando los beneficios de la homeopatía. Sería de interés para trabajos futuros profundizar en la información que ofrecen los medios generalistas, pero también los especializados en salud y medicina.

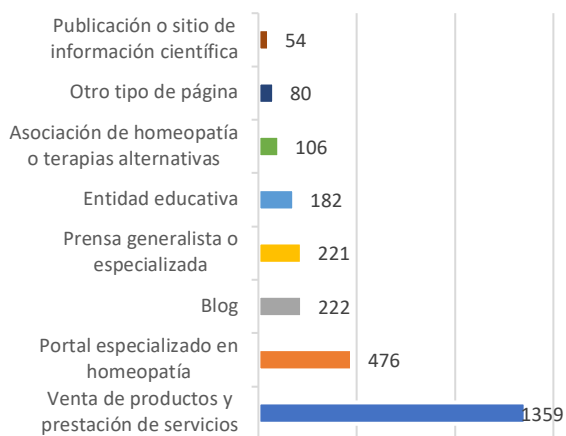


Figura 4. Tipología de páginas recolectadas

En cuanto a publicaciones científicas (casi 2%), se han identificado varios trabajos que analizan los efectos de la homeopatía ante diferentes afecciones. Según algunos estudios, el remedio funcionó positivamente (66,6%), para otros tuvo consecuencias negativas (casi 15%) y para el porcentaje restante no hizo nada (18,51%).

4.4.3. Dominios de las páginas recolectadas

Las 2715 páginas recolectadas corresponden a 196 dominios distintos. Esto podría indicar que las páginas sobre homeopatía tienden a referenciar contenidos de su propio dominio.

En la Tabla V se recogen 16 dominios con mayor frecuencia en la muestra (una ocurrencia mínima de 50), que representan más de un 71% del total de páginas recolectadas. La mayor parte de dominios se centran, como era lógico, a la vista de datos de apartados anteriores, en la venta y/o servicios de productos homeopáticos, o bien a la promoción de la homeopatía.

De los 16 dominios, sólo tres son ajenos al área y, además, tienen una tendencia negativa respecto a esta pseudociencia: “La lista de la vergüenza (blog Naukas)” (3,87%), “El País” (2,25%) y “El blog de búho gris” (1,84%). Resulta llamativo que un medio de comunicación generalista y de gran alcance, como es el periódico español “El País”, aparezca más veces que la página web de Boiron, la citada empresa que ofrece productos homeopáticos.

| Dominio | Frec. | % |
|--|-------|-------|
| Hablando de homeopatía | 289 | 10,64 |
| Quanta. Farmacia homeopática y boutique de la vida saludable | 239 | 8,80 |
| B JAIN | 211 | 7,77 |
| Clínica Medicina Integrativa | 178 | 6,56 |
| Consultorio médico-homeopático Doctores Eizayaga | 162 | 5,97 |
| Satisfarma | 151 | 5,56 |
| La lista de la vergüenza (blog Naukas) | 105 | 3,87 |
| Farmacia Coliseum | 97 | 3,57 |
| Elaesi CDMX | 89 | 3,28 |
| Jacomart Farmacia | 79 | 2,91 |
| Homeopatía Pura | 67 | 2,47 |
| El País | 61 | 2,25 |
| Sociedad Española de Medicina Homeopática | 55 | 2,03 |
| Homeopatía suma. Asamblea Nacional de Homeopatía (ANH) | 51 | 1,88 |
| El blog de búho gris | 50 | 1,84 |
| Instituto Argentino de Terapia Neural (Neuralterapia) | 50 | 1,84 |

Tabla V. Dominios con mayor frecuencia

4.4.4. Palabras clave de las páginas

Se localizaron 6223 palabras clave, de las cuales, “homeopatía” (2320) es la más frecuente, seguida de “homeopatía” (1523).

| Palabra clave | Frec. | % |
|----------------------------|-------|-------|
| homeopatía | 2320 | 37,28 |
| homeopatía | 1523 | 24,47 |
| homeópata | 884 | 14,21 |
| medicina homeopática | 461 | 7,41 |
| productos homeopáticos | 366 | 5,88 |
| nosode | 353 | 5,67 |
| remedios homeopáticos | 88 | 1,41 |
| memoria del agua | 83 | 1,33 |
| tratamientos homeopáticos | 62 | 1,00 |
| miasmas | 30 | 0,48 |
| dilución homeopática | 20 | 0,32 |
| ley de los infinitesimales | 16 | 0,26 |
| ley de los similares | 8 | 0,13 |
| agua polimerizada | 4 | 0,06 |
| homeopathic materia medica | 3 | 0,05 |
| constante de avogadro | 1 | 0,02 |
| samue hahnemann | 1 | 0,02 |

Tabla VI. Palabras clave identificadas

Estos dos términos constituyen casi el 62% del total. De hecho, juntas (“homeopatía, homeopatía”) aparecen en 1227 ocasiones. También se emplean otras palabras clave relevantes para el área objeto de estudio, como “homeópata” (14,21%), “medicina homeopática” (7,41%) o “productos homeopáticos” (5,88%).

Se observa que los términos propios de la jerga homeopática, como “Ley de los infinitesimales” o “Miasmas”, son menos representativos en las páginas recolectadas. Se trata de un resultado lógico si se tiene en cuenta que la gran mayoría de éstas se orientan a la venta o divulgación, reafirmando una imagen positiva de la homeopatía. Los conceptos más complejos relacionados con esta pseudoterapia están fuera de su ámbito de interés. Se entiende que el usuario medio desconoce las bases teóricas de la homeopatía, por lo que es comprensible que este tipo de conceptos no aparezcan en las páginas analizadas.

Una excepción es el término “Nosode” (5,67%), utilizado fundamentalmente en las páginas categorizadas como publicaciones o sitios de información científica. Es decir, se inserta en un ámbito científico y muy especializado.

Un posible trabajo futuro podría ser el análisis del vocabulario de las páginas tanto de tendencia positiva como negativa, para conocer su posible influencia en la construcción de la opinión de los usuarios.

5. Conclusiones y trabajos futuros

La homeopatía resulta atractiva para un gran número de pacientes (López Espinosa, 1999) que, si bien se sirven de amigos y conocidos como principal elemento de información, acuden a Internet como segunda fuente (Cano-Orón et al., 2018). Ahora bien, ni la abundancia de páginas especializadas y tampoco un uso extendido aseguran que la información sea de calidad ni fiable.

Este trabajo ofrece información acerca de las posibles vías que los usuarios pueden tener para obtener información sobre este tipo de pseudoterapias. Los resultados muestran una tendencia de las fuentes a ofrecer un contenido con sesgo positivo y fácilmente comprensible por el usuario medio. La relativa escasez de páginas con un sentido crítico o incluso sin tendencia puede ser un factor que impulse a los usuarios a inclinarse por la utilización de esta pseudoterapia, ya que es posible que sea interpretada como prueba efectiva de sus beneficios.

Si bien resulta lógico que las páginas web creadas expresamente para apoyar la homeopatía muestren esa tendencia positiva, la capacidad de los medios para influir en la opinión pública hace

que su responsabilidad ética y social sea mayor. De hecho, como se ha observado, sus publicaciones tienen casi siempre una postura crítica con la homeopatía. Resulta evidente que los medios generalistas influyen de forma decisiva en la opinión pública; un usuario que quizás no se haya sentido atraído por la homeopatía o la desconozca, es posible que cambie de punto de vista si una fuente a priori fiable le ofrece una visión positiva o neutra de este tipo de productos.

Tanto los términos más frecuentes (muy poco especializados), como las fuentes recolectadas con mayores ocurrencias revelan la predisposición a ofrecer información con la finalidad de influir con un sesgo positivo que de ofrecer contenidos objetivos y basados en la evidencia científica.

Un análisis pormenorizado de los términos empleados en los contenidos de páginas recolectadas podría mejorar el propio vocabulario controlado, enriqueciendo así los procesos de captación de nuevas páginas visitadas. La extracción de palabras clave del dominio más allá de las páginas visitadas también podría contribuir a mejorar esa lista de términos.

Teniendo en cuenta los riesgos de la popularización de las pseudoterapias, es clara la necesidad de una alfabetización informacional en los usuarios para mejorar su capacidad de detectar fuentes sesgadas. Asimismo, sería interesante un “contraataque” informativo por parte de organismos oficiales y medios de comunicación generalistas que, en un lenguaje fácilmente comprensible y con un vocabulario cercano, explicasen la realidad de la homeopatía como pseudoterapia y los beneficios de la medicina basada en la evidencia.

Por otro lado, el comportamiento del rastreador fue el esperado. Las páginas que no pertenecían a la temática de este estudio, filtradas gracias al empleo de un vocabulario controlado, fueron descartadas. Las páginas captadas presentaban, como mínimo, un término de la lista de palabras relevantes para el dominio analizado y fueron almacenadas en un formato que hace posible su análisis, junto con sus metadatos. Todavía se puede hacer evolucionar el algoritmo del software, programando su mecanismo de captación de páginas hacia una modalidad *multiprocessing*, donde varias páginas se pueden visitar “al mismo tiempo”, acelerando o proceso de evaluación y captación (Python, 2022).

“Crawler by domain” ha resultado útil para el propósito de recolección de páginas cuyo contenido trata algún aspecto relacionado con la homeopatía. El proceso de ejecución es sencillo, aunque puede mejorarse y la obtención de un documento en formato de hoja de cálculo facilita el análisis

de los datos. Sin embargo, hay dos mejoras obvias que hay que realizar: a) durante el procesamiento, se deben desechar las páginas que son recolectadas cuando el término del vocabulario controlado se localiza en el cuerpo de la página web pero como parte de un enlace ya que, en realidad, no forman parte de su contenido; b) es necesario que, de alguna forma, se comprueben si los enlaces recortados o que hacen referencias a partes de una página web han sido ya incluidos en la lista de URLs recolectadas, pues estaríamos hablando de una misma repetición.

Finalmente, esta aplicación podría utilizarse para analizar otros temas de interés social, como puede ser el SARS-CoV-2. Esta cuestión está teniendo repercusiones mundiales y, desde hace tiempo se observan movimientos negacionistas en contra de las evidencias científicas. En España, diversos equipos, integrados por documentalistas y periodistas, como VerificaRTVE se encargan de analizar los datos y arrojar luz sobre la veracidad de la información aportada. Un estudio basado en el análisis conceptual de fuentes sobre este tema, en combinación con la “Crawler by domain”, aportaría nuevos datos sobre la intencionalidad de las fuentes en línea sobre dicho virus y sus consecuencias para el ser humano.

Referencias

- Alcott, H.; Gentzklow, M. (2017). Social Media and Fake News in the 2016 Election. // *Journal of economic perspectives*. 31:2, 211-36. <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211>
- Baños-Moreno, M.-J. (2017). Propuesta de modelado de una ontología de dominio para la representación de acciones en política-economía. Universidad de Murcia. <http://hdl.handle.net/10201/56661>
- Baños-Moreno, M.J.; Felipe, E. R.; Pastor-Sánchez, J. A.; Lima, G.; Martínez-Béjar, R. (2017). Análisis de metadatos de noticias para la extracción de información del código fuente: El software METADADOSHTML. // *Information Research*. 22:1. <http://InformationR.net/ir/22-1/paper740.html>
- Bedi, P.; Thukral, A.; Banati, H.; Behl, A.; Mendiratta, V. A. (2012). Multi-Threaded Semantic Focused Crawler. // *Journal of Computer Science and Technology*. 27:6, 1233–1242. <https://doi.org/10.1007/s11390-012-1299-8>.
- Cano-Orón, L.; Mendoza-Poudereux, I.; Moreno-Castro, C. (2018). Perfil sociodemográfico del usuario de la homeopatía en España. // *Atención Primaria*. 51:8, 499-505. <https://doi.org/10.1016/j.aprim.2018.07.006>
- Chakrabarti, Soumen (2003). *Mining the Web: discovering knowledge from hypertext data*. San Francisco, CA: Morgan Kaufmann Publishers.
- Chowdhury, G. G. (2004). *Introduction to modern information retrieval*. 2nd ed. London: Facet.
- Dahiwale, P.; Raghuvanshi, M. M.; Malik, L. (2014). PDD Crawler: A Focused Web Crawler Using Link and Content Analysis for Relevance Prediction. // *Third International Conference on Advanced Information Technologies & Applications*, 7 nov. 2014. *Comp. Science & Information Technology (CS & IT)*. 245–253. <https://doi.org/10.5121/csit.2014.41123>.
- Du, Y.; Liu, W.; LV, X.; Peng, G. (2015). An improved focused crawler based on Semantic Similarity Vector Space Model. // *Applied Soft Computing*. 36, 392–407. <https://doi.org/10.1016/j.asoc.2015.07.026>.
- Europa Press (2021). Sanidad califica como pseudoterapias a la dieta macrobiótica, al masaje tailandés y a la magnetoterapia estática. <https://www.infosalus.com/actualidad/noticia-sanidad-califica-pseudoterapias-dieta-macrobiotica-masaje-tailandes-magnetoterapia-estatica-20210219132911.html> (19/02/21).
- Farag, M.; Lee, S.; Fox, E.A. (2018). Focused crawler for events. // *International Journal on Digital Libraries*. 19:1, 3–19. <https://doi.org/10.1007/s00799-016-0207-1>.
- Groc, Clement, Babouk de (2011). Focused Web Crawling for Corpus Compilation and Automatic Terminology Extraction. // *2011 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, ago. 2011. Lyon, France: IEEE, ago. 2011. 497–498. <https://doi.org/10.1109/WI-IAT.2011.253>. <http://ieeexplore.ieee.org/document/6040719/>.
- Hegade, P.; Lingadhal, N.; Jain, S.; Khan, U.; Vijeth, K. L. (2021). Crawler by Contextual Inference. // *SN Computer Science*. 2:3, 216. <https://doi.org/10.1007/s42979-021-00574-z>.
- Lee, J. G.; Bae, D.; Kim, S.; et al. (2020). An effective approach to enhancing a focused crawler using Google. // *The Journal of Supercomputing*. 76:10, 8175–8192. <https://doi.org/10.1007/s11227-019-02787-9>.
- López Espinosa, J.A. Notas para la historia de la homeopatía. (1999). *Revista Cubana de Medicina General Integral*. 1999. 587-590. http://www.bvs.sld.cu/revistas/mgi/vol15_5_99/mgi17599.htm
- Martin, Robert C. (Org.). (2009). *Clean code: a handbook of agile software craftsmanship*. Upper Saddle River, NJ: Prentice Hall.
- Nayernouri, T. (2017). Homeopathy, Ritual and Magic. // *Archives of Iranian Medicine*. 20:11, 718-722. <http://www.aimjournal.ir/PDF/aim-1701>
- Python (2022). multiprocessing — Process-based parallelism. // *Documentation*. <https://docs.python.org/3/library/multiprocessing.html>
- Safran, M. S.; Althagafi, A.; Dunren C. Improving Relevance Prediction for Focused Web Crawlers. In: *2012 IEEE/ACIS 11TH International Conference on Computer and Information Science (ICIS)*, maio 2012. 2012 IEEE/ACIS 11th International Conference on Computer and Information Science [...]. Shanghai: IEEE, maio 2012. p. 161–166. DOI 10.1109/ICIS.2012.61. <http://ieeexplore.ieee.org/document/6211091/>.
- Souza, R. R.; Tudhope, D.; Almeida, M. B. (2012). Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. // *Knowledge Organization*. 39:3, 180. <https://www.nomos-elibrary.de/10.5771/0943-7444-2012-3-179.pdf>
- Redacción Médica. (2019). La homeopatía no cura ya ni según la definición de la RAE. <https://www.redaccionmedica.com/virico/noticias/la-homeopatia-no-cura-ya-ni-segun-la-definicion-de-la-rae-2633> (28-12-2019).
- Young, P. (2014). La farsa de la homeopatía. // *Revista Médica de Chile*. 142:2. <http://dx.doi.org/10.4067/S0034-98872014000200021>
- Yu, Liyang (2011). *A developer's guide to the semantic web*. Berlin: Springer.

Enviado: 2021-03-27. Segunda versión: 2022-05-01.
Aceptado: 2022-05-13.

Representação temática de material audiovisual em repositórios multimídia de instituições federais brasileiras: um estudo de caso com conceito paradigmático

Representación temática del material audiovisual en repositorios multimedia de instituciones federales brasileñas: un estudio de caso con un concepto paradigmático

Thematic representation of audiovisual material in multimedia repositories of Brazilian federal institutions: a case study with a paradigmatic concept

**Suelene Santana ASSUNÇÃO (1) , João Vitor Correa DINIZ (1),
Franciele Marques REDIGOLO(1), Mariângela Spotti Lopes FUJITA (2)**

(1) Universidade Federal do Pará, R. Augusto Corrêa, 01 - Guamá, Belém - PA, 66075-110, Brasil, suelene.1410@gmail.com, joaovitorcorrea@gmail.com, franciele@ufpa.br (2) Universidade Estadual Paulista "Júlio de Mesquita Filho" - Unesp, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP.17525-900, Marília, SP, Brasil, mariangela.fujita@unesp.br

Resumen

La representación temática de material audiovisual en repositorios multimedia se explora a través de un concepto paradigmático para Brasil: 'Amazônia'. Se buscaron videos sobre el tema en tres repositorios multimedia de instituciones federales brasileñas: el repositorio LUME, alojado en un dominio de la Universidad Federal de Rio Grande do Sul (UFRGS); el Repositorio Institucional de Geociencias de la Empresa de Investigación de Recursos Minerales (CPRM), conocido como RIGeo; y el Repositorio Multimedia de la Universidad Federal de Pará (UFPA). Se recuperaron 27 resultados en la UFPA, 2 en LUME y ninguno en RIGeo. Se encontró que los repositorios analizados utilizan hasta diez términos generales y específicos para la representación temática de un documento audiovisual en formato de video. Paradójicamente, un concepto tan importante para Brasil no siempre se recupera fácilmente en los repositorios audiovisuales relevantes, lo que destaca la importancia de la adecuada representación temática en estas bases de datos.

Palabras clave: Indización temática. Repositorios institucionales. Multimedia. Documentos audiovisuales. Repositorio LUME (Brasil). Repositorio Institucional de Geociencias de la Compañía de Investigación de Recursos Minerales (Brasil). Repositorio UFPA Multimedia (Brasil). Amazon. Brasil.

1. Introdução

Esta pesquisa discorre sobre a representação de assunto de materiais audiovisuais, em repositórios multimídia. Os repositórios institucionais reúnem conhecimentos organizados e a produção intelectual de uma instituição, seja de ensino ou de pesquisa. Uma característica comum dos Repositórios Institucionais (RI) é seu objetivo de

Abstract

The thematic representation of audiovisual material in multimedia repositories is explored throughout a case study with a paradigmatic concept for Brazil: 'Amazônia'. Three multimedia repositories of Brazilian federal institutions were researched for videos on the topic: the LUME repository, hosted in a domain of the Federal University of Rio Grande do Sul (UFRGS); the Institutional Repository of Geosciences of the Mineral Resources Research Company (CPRM), known as RIGeo; and the Multimedia Repository of the Federal University of Pará (UFPA). 27 results were retrieved in UFPA, 2 in LUME and none in RIGeo. It was found that the analyzed repositories use up to ten general and specific terms for the thematic representation of an audiovisual document in video format. Paradoxically, such an important concept for Brazil is not always easily retrieved in relevant audiovisual repositories, which highlights the importance of a proper thematic representation in these databases.

Keywords: Thematic indexing. Institutional repositories. Multimedia. Audiovisual documents. LUME Repository (Brazil). Institutional Repository of Geosciences of the Mineral Resources Research Company (Brazil). UFPA Multimedia Repository (Brazil). Amazon. Brazil.

oferecer serviços a partir de um coletivo organizado que disponibilize, importe, dissemine e gereencie conteúdos informacionais (Monteiro, 2009). É nesse contexto que estão localizados os repositórios multimídia, espaços onde se armazenam documentos em formato de vídeo, áudio, fotos, entre outros, disponibilizados como recursos didáticos, que se coloca a questão desta investigação: como ocorre a representação de as-

sunto de mídias em vídeo, nos repositórios institucionais multimídia das instituições federais brasileiras?

O tema de estudo é a representação temática de material audiovisual em Repositórios Multimídia, a partir de um estudo de caso com conceito paradigmático com o objetivo de analisar a representação do conteúdo pela palavra-chave 'Amazônia', no campo temático.

Ao tratar do termo Amazônia como um conceito paradigmático, pretende-se realizar uma leitura diferente de algo que vise sua historicidade, por exemplo, atravessada por extrativismo, migrações, políticas desenvolvimentistas, e questões agrárias. Propôs-se uma faceta na qual estes marcadores socioeconômicos possam estar contidos; contudo volta-se para uma abordagem que vislumbre a produção de conteúdos acadêmicos, de formações diversas. E que estes conteúdos possam servir de subsídios para outros debates, a respeito deste conceito que se faz paradigmático desde sua formação histórica (Castro; Índio, 2015). Dessa forma, foi escolhido o termo Amazônia como descritor norteador das buscas realizadas em ambientes digitais.

Desse modo, inseridos no contexto dos repositórios institucionais que possuem acervo multimídia, escolheram-se para este estudo três repositórios: o primeiro repositório foi o LUME, pertencente à Universidade Federal do Rio Grande do Sul (UFRGS); o segundo escolhido é o Repositório Institucional de Geociências da Companhia de Pesquisa de Recursos Minerais (CPRM), denominado RIGeo; o terceiro foi o Repositório UFPA Multimídia, localizado na região Norte do Brasil.

Destacam-se referenciais que abordam: perspectivas sobre a representação da informação: considerando-se a organização e recuperação do material audiovisual, no repositório (Novelino, 1996; Zarpelon, Basniak, Silva e Resende, 2018); e representação temática (Fujita, 2003; Feitosa, 2006). Ademais, há teóricos que auxiliaram nas análises, tanto de repositórios digitais institucionais (Leite, 2009; Sayão e Sales, 2016; Tartarotti, Dal'Evedove e Fujita, 2020), quanto para a questão dos repositórios multimídia, nas discussões de Zarpelon, Basniak, Silva e Resende (2018). Como suporte teórico para os Repositórios Institucionais, utilizaram-se os textos de González-Pérez, Ramírez-Montoya e García-Peñalvo (2020); igualmente, Kuri e Singh (2020). No debate sobre informação e sociedade, Buckland (2017) foi importante.

Dessa maneira, o artigo está estruturado em seis tópicos, no segundo, fundamentos teóricos sobre representação da informação em documentos

audiovisuais, no terceiro, discorre-se sobre os repositórios institucionais e os documentos multimídia, na quarta seção, encontra-se a metodologia adotada nesta pesquisa, a análise da representação da informação nos repositórios institucionais multimídias é descrita na quinta seção e, por último, as considerações finais.

2. Representação temática de documento multimídia audiovisual

Esta pesquisa dedicou-se à discussão da representação temática de material audiovisual, presente dentro e fora das salas de aula, e que aparece nos diferentes suportes de acesso digital dos discentes. Deve-se frisar que os documentos multimídia audiovisuais, como um tipo de recurso de receptividade (captação sensorial, cognitiva, didática) diversificada - e com o auxílio das TICs (compactação, descompactação, *upload download*, armazenamento em *drives* portáteis, CDs etc.) auxiliam os discentes no seu processo de ensino-aprendizagem.

Ademais, um dos desafios para lidar com vocabulários dos conceitos encontrados em distintos repositórios multimídia está, por exemplo, na disposição e escolha de seus elementos representativos e da organização de seus conteúdos audiovisuais.

Para isso, os sistemas de classificação destacam-se na organização da informação conservada em bibliotecas, por exemplo, tendo seu uso recorrente até o presente (Brascher; Monteiro, 2010). Os sistemas de classificação são considerados "o primeiro tipo de instrumento de indexação utilizado em sistemas documentários" (Chaumier, 1988, p. 67).

Os sistemas de classificação são, também, utilizados nos repositórios institucionais a fim de organizar o conhecimento produzido, disponibilizando o livre acesso ao material bibliográfico ou audiovisual, armazenando-os mediante o tratamento temático da informação, para serem recuperados pelos usuários.

Pinto Molina, Garcia Marco e Agustín Lacruz, (2002), no livro "Indexação e resumo de documentos digitais e multimídias", apresentam exemplos sobre a análise de conteúdo de documentos gráficos, cinematográficos, videográficos, televisivos e multimídia, entre outros gêneros, como guia metodológico, ressaltando a mudança documental e o novo paradigma dos multimídias, sugerindo técnicas para a indexação e elaboração de resumo.

A voz, os sons ambientes, a música e as imagens compõem as informações audiovisuais. Portanto, para Pinto Molina, Garcia Marco e Agustín Lacruz

(2002), nos documentos audiovisuais, existem três segmentos de interação entre si, os quais devem ser analisados: o sonoro, da imagem e o retórico sintático, derivado da interação entre os segmentos anteriores: a montagem e a edição.

Por sua vez, nas operações do tratamento temático da informação, estão a catalogação, a classificação, a indexação e o resumo (Feitosa, 2006). De acordo com Chaumier (1988), a indexação é peça fundamental para a análise documental. Nesse sentido, a indexação “[...] consiste em atribuir um ou vários descritores, que podem ter ligações entre si, para descrever o conteúdo do documento [...]” (Feitosa, 2006, p. 21). Para realizar a indexação, é preciso fazer a análise de assunto e utilizar a linguagem documental para a seleção dos descritores que representam o documento (Fujita, 2003), de sorte a proporcionar a recuperação da informação pelos usuários.

A representação da informação, conforme Novelino (1996, p. 38), abrange duas etapas principais: análise de assunto do material em uma expressão linguística e a atribuição de conceitos do documento analisado. A descrição de um documento deve ser realizada de forma padronizada, por intermédio do uso de instrumentos de controle de vocabulário para que a recuperação da informação aconteça sem inconsistências.

Observa-se, por exemplo na Figura 1, que o repositório Audiovisual da UFC possui seis categorias (cinematográfico, televisivo, sonoro, institucional, eventos, cultural e artístico), em que podem ser classificados os documentos audiovisuais. Dentro de cada uma dessas categorias, há as subcategorias nas quais o material pode ser classificado. Santos (2018b, p. 62) elaborou o mapeamento das categorias que compõem o acervo do repositório audiovisual da Universidade Federal do Ceará (UFC).

La Barre e Cordeiro (2012, p. 241) analisaram *sites* de pesquisa de imagem em movimento, com realce para os seguintes pontos principais: “[...] resumo, data, título, pessoas associadas, gênero, direitos, duração, cor, formato, palavra-chave, assunto e fonte [...]”.

Nessa perspectiva, Santos (2013, p. 6) assinala que a representação temática de material audiovisual consiste na atribuição dos termos, de forma hierarquizada; assim, a representação da informação acontece “[...] sempre indexando os termos do mais geral para o específico, de acordo com a imagem ou a política da instituição exigir [...]”.

Percebe-se que existe uma diversidade de tipologia documental multimídia, dentro de cada categoria, por isso, é importante que esse material

multimídia seja representado de maneira fidedigna, a fim de que o usuário final tenha êxito na realização da sua pesquisa. Para isso, a informação audiovisual “traz em seu conteúdo texto, áudio (sons, ruídos etc.) e vídeo (imagens em movimento) sincronizados e convertidos para uma nova dimensão textual” (Santos, 2018a, p. 255).

Por conseguinte, no próximo tópico, apresenta-se a contextualização sobre os repositórios institucionais multimídia.

2.1. Contextualizações sobre repositórios institucionais multimídia

Para a elaboração deste item buscou-se os termos “repositórios institucionais, repositórios institucionais federais, repositórios digitais, repositórios multimídia, repositórios audiovisuais, repositórios multimídia no Brasil, *institutional repositories, digital repositories, multimídia repositories, audiovisual repositories*”, nas bases BRAPCI, Scopus e Web of Science.

Após a Segunda Guerra Mundial, o desenvolvimento da ciência delineou-se através da informação em ciência e tecnologia para o avanço de um país, por conseguinte, coletar, tratar, organizar e recuperar a informação tornou-se necessário para a tomada de decisões nos mais variados segmentos da sociedade (Buckland, 2017).

Com o advento das tecnologias de informação e comunicação, o fluxo da informação também acontece pela rede de comunicação eletrônica. Dessa forma, o uso da internet diminuiu as barreiras geográficas, otimizou o tempo de pesquisa e ampliou o compartilhamento de conhecimentos. Ampliaram-se os recursos didáticos utilizados para compreensão dos conteúdos educacionais nas diversas áreas; com isso, surgiram os repositórios de Objetos de Aprendizagem (OA) (López Guzmán, 2005; Tortosa, 2006).

Nesse formato de navegação eletrônica, tem-se o acesso aos Repositórios Digitais, os quais podem ser: Institucionais, Temáticos, Disciplinares e de Dados de Pesquisa (Marcondes; Sayão, 2009; Sayão; Sales, 2016). Os Repositórios Digitais se destacam com processos que os caracterizam, como organização, armazenamento, disseminação, implementação de políticas, acessibilidade, atemporalidade, recuperação e preservação – por exemplo – da produção científica de uma comunidade ou instituição de ensino e pesquisa (Leite, 2009).

Não obstante, o perfil prático da tecnologia adotada pelos repositórios deve ser levado em conta, para a identificação de ferramentas tecnológicas cabíveis, tanto ao público quanto às instituições

mantenedoras, orientando-se pelas novas tendências de tecnologia e gestão da informação (González-Pérez, Ramírez-Montoya e García-Peñalvo, 2020).

Marcondes e Sayão (2009, p. 9) definem o repositório institucional como “[...] uma biblioteca digital destinada a guardar, preservar e garantir livre acesso via internet, à produção científica no âmbito de uma dada instituição.”

Tartarotti, Dal’Evedove e Fujita (2020, p. 13) conceituam os repositórios digitais como “[...] sistemas de recuperação da informação pensados para atender os modos contemporâneos de criar conhecimento e gerir informações científicas de acesso aberto.” Os repositórios digitais são fontes de informações confiáveis, para a realização de pesquisas e disseminação do acervo digital que compõem o repertório de uma determinada instituição.

Ademais, esses espaços tem potencial de transformação social, pois os repositórios armazenam quantidades expressivas de conteúdos informacionais, com possibilidade de acesso aberto, e são desenvolvidos como parte da gestão do conhecimento das instituições que os gerem (Kuri; Singh, 2020).

Como reforço à relevância desse tipo de conjunto documental, e devido à produção científica de uma instituição ser composta de materiais bibliográficos (e multimídia), Cunha e Cavalcante (2008, p. 131) reforçam que o documento audiovisual é um “[...] registro associado de som e imagem, para formar um tipo específico de documento”.

Sobre documentos multimídia, aborda-se o conceito de multimídia relacionado à integração e controle dos computadores. Como textos, imagens, vídeos, áudios e suportes de informação podem ser armazenados, transmitidos e processados de forma digital (Marshall, 2001). Quanto à característica audiovisual, trata-se de materiais instrutivos não impressos, os quais auxiliam na apreensão dos sentidos, no processo de ensino e aprendizagem (Adebowale e Adekanye, 2013).

Os objetos educacionais se valem de textos, imagens, vídeos e áudios, como recursos de aprendizagem, e uso de computador. Por integrar mais de uma mídia em um documento, esse conjunto é denominado multimídia. E o material audiovisual dispõe de dois recursos, que são o áudio e a imagem, para a instrução de determinado assunto. Assim, os objetos educacionais encontrados nos repositórios digitais inserem-se tanto na qualidade de multimídias, por utilizarem textos, imagens, vídeos e áudios, como recursos didáticos. Quanto têm especificado audiovisual, a qual se dedica esta

pesquisa, é por serem materiais circunscritos nesse conjunto maior de variedades digitais.

Ao longo da leitura dos materiais bibliográficos, perceberam-se equivalências e hierarquias quanto aos usos dos termos “Objetos de Aprendizagem” (OA) e “Objetos Educacionais”, empregados em contextos equivalentes, por outro lado, “Objetos de Aprendizagem Multimídia” vêm a ser uma espécie de subcategoria contemplada dentro dos OA, conforme constatado em leituras e acessos feitos nos repositórios institucionais. Verificaram-se outras variações terminológicas para representar os materiais multimídia disponibilizados, como: documentos multimídia, arquivo multimídia, objetos multimídia, linguagens multimídia ou recursos multimídia (López Guzmán, 2005; Dias, 2012; Dias e Moreira, 2014; Santos, Pavão e Moura, 2016; UFPA, 2020).

Entre o domínio de estudos voltados para relação e eficácia de repositórios multimídia, Zarpelon Basniak, Silva e Resende (2018), por exemplo, caracterizam três repositórios de matemática e seus objetos de aprendizagem. Ao longo desse estudo, Zarpelon Basniak, Silva e Resende (2018, p. 52) elaboram um quadro com a relação de alguns repositórios de reconhecimento nacional e internacional. Observa-se, a existência de 10 repositórios com acervos multimídia nacionais e 2 repositórios internacionais, totalizando 12 repositórios; dentre os repositórios nacionais, apenas 3 são de universidades localizadas em regiões diferentes, a saber: no Sul (Universidade Federal do Rio Grande do Sul), no Nordeste (Universidade Federal do Ceará) e no Sudeste (Universidade de São Paulo).

Todavia, mesmo que os repositórios institucionais possam apresentar vias de acessibilidade, o momento de privações, seja pela pandemia, seja pelas condições sociais do Brasil, apresentam adversidades díspares referentes ao acesso dos meios e produtos informacionais, o que vem exigindo de os profissionais repensarem sua atuação frente à realidade informacional, bem como da sociedade à qual pertencem (Santos e Carvalho, 2009).

Ao entender que os documentos audiovisuais possuem diversidade de usos e de metadados a serem acessados, acredita-se que esses documentos podem compor, de forma eficaz, a organização de um repositório multimídia, para busca e localização por parte de seus usuários.

Nesse sentido, para López Guzmán (2005), a elaboração de metadados torna-se necessária em face de recursos apresentados inicialmente, de forma não textual, sendo inviável sua indexação por sistemas de automação; como exemplo

desses recursos, têm-se os materiais registrados em multimídias ou audiovisuais.

Em relação ao campo educacional, os documentos audiovisuais são adotados na representação sistêmica dos recursos disponíveis em repositórios institucionais, por exemplo. Assim, ao estabelecer padrões de organização para disponibilização de materiais audiovisuais, facilita-se a sua recuperação, pois um conjunto de elementos ordenados permite a identificação das principais características dos recursos disponíveis (Chauvier, 1998).

Na esteira da organização de conteúdos e disponibilização sistemática aos usuários, traz-se, no item seguinte, uma breve descrição da representação temática da informação em materiais audiovisuais. Tendo feito até aqui considerações sobre os materiais audiovisuais e seus repositórios, percebe-se o quanto o trabalho com seus componentes representados pelos mencionados metadados e os conceitos que abarcam demandam uma ordenação sistemática, para fins de recuperação de suas informações.

3. Metodologia

Esta pesquisa caracteriza-se como exploratória, com abordagem qualitativa e seu desenvolvimento se realizou em duas etapas: análise documental e análise temática. A análise documental foi realizada por meio de pesquisa eletrônica na internet para levantamento, análise e seleção de repositórios institucionais brasileiros. A análise temática utilizou procedimentos de busca por assunto nos repositórios selecionados.

3.1. Análise documental

Para a escolha dos repositórios procurou-se no site do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT, 2018), na lista de repositórios digitais brasileiros, por repositórios institucionais que atendessem aos seguintes critérios pré-selecionados: ser de instituição federal brasileira; possuir material audiovisual em formato de vídeo; serem de regiões diferentes do Brasil e utilizarem o software DSpace. O *software DSpace*, possui acesso aberto e está estruturado em comunidades e coleções, satisfazendo assim às necessidades informacionais, tanto em nível institucional quanto social.

Logo, a busca dedicou-se aos repositórios de instituições federais de ensino e pesquisa, os quais possuíssem um repertório audiovisual. Na base do IBICT, o universo de repositórios federais no Brasil é de 76, distribuídos em 66 instituições públicas e 10 instituições privadas. Destaca-se que

3 repositórios institucionais atenderam aos parâmetros pré-estabelecidos para esta pesquisa: 1 dos repositórios selecionados encontra-se na região Sul, outro da região Centro-Oeste e outro da região Norte do Brasil.

Na região Sul, o repositório LUME (<https://www.lume.ufrgs.br/>), pertencente à Universidade Federal do Rio Grande do Sul (UFRGS), contém materiais audiovisuais, em seu acervo digital. O repositório LUME foi criado em 2008, a partir de uma iniciativa do sistema de bibliotecas e do Centro de Processamento de Dados da UFRGS, cujo objetivo é reunir o acervo digital das bibliotecas e demais órgãos da UFRGS, otimizando o tempo do usuário final (Santos, Pavão e Moura, 2016).

Na região Centro-Oeste, destaca-se o repositório Institucional de Geociências, da Companhia de Pesquisa de Recursos Minerais (CPRM). Foi criado para armazenar a produção da instituição e disponibilizá-la à sociedade, através do repositório de acesso aberto, contribuindo para a democratização da informação, no Brasil.

Na região Norte, tem-se o repositório UFPA Multimídia (2012) vinculado à Universidade Federal do Pará (<http://rigeo.cprm.gov.br/jspui/>). O Repositório UFPA Multimídia foi idealizado em função de um edital da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), em 2010, originando o projeto de resgate e sistematização da trajetória da educação à distância, na UFPA, tendo à frente a docente Selma Leite. Nesse projeto, além de produzir material didático, descobriu-se um rico acervo da década de 1990 que estava disperso, sendo necessário criar um local para o armazenamento e socialização desses recursos midiáticos, a partir da visão local da Região Amazônica; com isso, o repositório começou a ser pensado como um ambiente que iria reunir essa produção amazônica e disponibilizá-la, inclusive em outros idiomas, como inglês e espanhol (Malcher, 2012).

3.2. Análise temática

Nos repositórios LUME, RIGeo e UFPA Multimídia, a busca foi realizada a partir do campo assunto, pela categoria de material “vídeo” e uso do descritor “Amazônia”. A escolha pelo descritor “Amazônia” é conceitual e se explica em virtude de ser uma região rica de informações – ainda não registradas ou compartilhadas – e que está em evidência, seja pela diversidade da sua fauna e flora, seja por problemas específicos que a população amazônica vivencia. A seleção do material de vídeo deu-se por ser uma mídia que possui pontos de acesso diversificados, os quais devem ser descritos na sua representação de conteúdo.

Dessa forma, o percurso de busca no Repositório LUME iniciou-se no campo “assunto”, onde se digitou o descritor “Amazônia”, recuperando-se como resultado 57 documentos, após a verificação do tipo de material audiovisual selecionado para esta pesquisa (vídeo). Dos 57 documentos recuperados, apenas 2 são materiais multimídia audiovisual, enquanto os 55 documentos restantes estão distribuídos, por tipologia documental, em Teses, Dissertações, Trabalhos de Conclusão de Curso e Artigos Científicos.

No RIGeo, por ter um extenso acervo da região amazônica, iniciou-se a busca por “Comunidades e coleções”, escolheu-se o tipo de material “multimídia”, observou-se a quantidade de 12 documentos no formato multimídia, porém, ao se digitar o termo “Amazônia”, não foi recuperado nenhum documento multimídia no formato de vídeo relacionado a Amazônia.

Já no Repositório UFPA Multimídia, ocorreu pela escolha no campo “assunto”, digitação do descritor “Amazônia”, obtendo-se 27 documentos, cujo tipo de material está em formato de vídeo, 19 documentos em áudio e 1 fotografia, totalizando 47 resultados para o termo “Amazônia”.

Além da busca por assunto, identificou-se a recuperação de material com o termo ‘Amazônia’ em outros campos de busca, por exemplo: resumo, descrição do vídeo e ao observar todos os metadados do material indexado.

No que concerne aos repositórios selecionados para pesquisa (LUME, RIGeo e UFPA Multimídia), entende-se que suas formas de gestão documental dialogam tanto com o momento atual – relativo ao desenvolvimento das Tecnologias da Informação e Comunicação (TICS) –, pois suas plataformas se assemelham aos demais modelos de ambientes da rede, quanto as tradicionais instituições de memória, devido as suas formas de gerirem os patrimônios documentais.

Após a identificação do material no formato de vídeo, verificamos o registro completo dos documentos selecionados, analisando os descritores utilizados na indexação do material audiovisual e observando a representação temática do material selecionado. No próximo item, apresentamos o resultado da busca realizada nos repositórios multimídia.

4. Análise da representação de assunto nos repositórios LUME, RIGeo e UFPA Multimídia

Neste item, explicita-se a análise referente à busca, nos repositórios institucionais LUME, RIGeo e UFPA Multimídia, tendo em vista que se deu a partir do campo “assunto”, recorrendo-se

ao descritor “Amazônia” cujo tipo do documento seja “vídeo”. Após os resultados da busca, realizou-se a seleção do material de tipologia audiovisual, para analisar como é feita a representação temática desse tipo de documento.

No Repositório LUME, recuperaram-se 2 materiais audiovisuais, o primeiro intitulado “Conexões Amazônicas” e o segundo, “Núcleo de Estudos Geografia e Ambiente (NEGA)”. Na Figura 1, tem-se a representação temática do documento “Conexões Amazônicas”.

The screenshot displays the LUME Digital Repository interface. At the top, there is a navigation bar with the LUME logo and the text 'Repositório Digital'. Below this, a breadcrumb trail shows the path: 'Lume Início > Acervo > UFRGS TV > Conteúdo a UFRGS > Visualizar item'. The main content area is titled 'Núcleo de Estudos Geografia e Ambiente (NEGA)'. It features a 'Resumo' section with a brief description of the NEGA, followed by a 'Data de publicação' (2014-10-07) and an 'Assunto' (Amazônia). The 'Entrevistado' section lists several names, and the 'Repórter/Entrevistador' section lists 'Carvalho, Luana'. There are also sections for 'Produtor', 'Realizada', 'Diretor Técnico', and 'Editor'. The 'Outras opções' section includes a link to the full metadata. On the right side, there is a search bar and a 'Navegar' menu with options like 'Todo o repositório', 'Coleções e comunidades', and 'Entrevistação'. At the bottom right, there are social media icons for Facebook, Google+, Twitter, LinkedIn, and Email.

Figura 1. Visualização da descrição de vídeo no Repositório LUME (<https://lume.ufrgs.br/handle/10183/104500>)

Observou-se que as descrições dos metadados remetem à equipe técnica que produziu o vídeo, com seu tempo de duração; no campo assunto, os termos correspondem à linguagem controlada e à linguagem natural. Verifica-se ainda que o Repositório LUME utiliza dez termos, distribuídos em gerais e específicos, para descrever os documentos, seja material audiovisual, seja bibliográfico.

No vídeo “Conexões Amazônicas” a abordagem é sobre um evento ocorrido em 2019, na UFRGS com apresentação das pesquisas com a temática Amazônia nas diversas áreas do conhecimento que são realizadas na UFRGS. O vídeo produzido compõe a coleção da UFRGS TV.

O segundo vídeo recuperado, apresenta o “Núcleo de Estudos Geografia e Ambiente (NEGA)”, vinculado ao Programa de Pós-Graduação em Geografia da UFRGS, que foi fundado em 2003 com o objetivo de trabalhar o ambiente geográfico com a participação da sociedade.

No caso dos materiais bibliográficos, geralmente se empregam cinco termos em língua portuguesa e na língua inglesa, totalizando dez termos para cada material bibliográfico.

A partir do resultado da busca no Repositório Institucional de Geociências da CPRM, verificou-se que o termo “Amazônia” não apresenta resultados no acervo do RIGeo correspondentes ao formato em “vídeo”, entretanto, há um extenso acervo, distribuído em mapas, dissertações, teses, artigos científicos, relatórios técnicos, trabalhos apresentados em eventos (recuperados com o descritor “Amazônia”). Dentre esses materiais bibliográficos, observou-se que são usados até a quantidade de dez termos para a descrição temática dos materiais, sendo que os termos estão apenas na língua portuguesa, limitando dessa forma a recuperação da informação no RIGeo.

The screenshot shows the UFFA Multímedia repository interface. The video title is "Geociências: O Museu de Geociências da UFFA e os minerais em nosso cotidiano". The author is Dall'Agnol, Roberto, Costa, Marcondes Lima da, and Natchez, Maria Azaide. The description states it is a product developed by the Amazonian Academy in partnership with researchers from the Instituto Nacional de Ciência e Tecnologia de Geociências da Amazônia. The video is 11 minutes and 18 seconds long. The license is Attribution-NonCommercial-ShareAlike 3.0 Unported by Creative Commons.

Figura 2. Visualização da descrição de vídeo no Repositório UFFA Multímedia (<http://www.multimedia.ufpa.br/jspui/handle/321654/858>)

No Repositório UFFA Multímedia, como visto na figura 3 recuperou-se 47 resultados para o termo “Amazônia”, distribuídos em 27 documentos em formato de vídeo, 19 documentos em áudio e 1 fotografia. Dos 27 documentos recuperados em formato de vídeo, 17 documentos trata-se de entrevistas com diversas personalidades paraenses e acadêmicas que participaram em 1992, do Projeto Educação sem Fronteiras na Amazônia (UFFA, 2010), em que foi indexado com 8 assuntos, a saber: Educação a distância, Ensino superior, Implantação, Amazônia, Matemática, Universidade Federal do Pará (UFFA), Matemática, Graduação, o conteúdo do material, aproxima-se da questão territorial, quanto a dimensão da região Amazônica, e as questões de implementação de um programa educacional à distância nesta região. Ainda dos 27 vídeos, identificou-

se 5 vídeos de entrevistas sobre “A Luta dos Soldados da Borracha por Justiça e Reparação” (UFFA, 2012), que foram indexados com 7 assuntos no repositório, a saber: Soldados da borracha; História; Nordeste; Amazônia; Justiça e reparação; Cinema; Direitos humanos. Identificou-se no conteúdo dos vídeos, pelo entrevistado cineasta César Garcia Lima, que faz referência não somente às migrações de nordestinos recrutados como soldados da borracha para a região, mas também como um local de impasse socioeconômico dentro do período destacado em sua fala: entre os anos de 1942 e 1945, durante o Governo de Getúlio Vargas quando homens nordestinos foram recrutados para trabalharem nos seringais, compreendidos dentro do espaço da Amazônia brasileira. Os demais vídeos são 4 documentos referente ao Programa de Pós-Graduação Comunicação, Cultura e Amazônia, no qual contém a participação de pesquisadores na aula inaugural de em 2010 e o Seminário de Metodologia de Pesquisa. E, por último, 1 vídeo apresentando o Museu de Geociências da UFFA e os minerais encontrados na região amazônica.

Após análise da representação da informação no campo “assunto”, no Repositório UFFA Multímedia, percebeu-se que há utilização de até dez termos para identificar um documento multimídia, adotando-se a linguagem controlada com o uso dos termos gerais e específicos.

| Repositório | Título do Vídeo | Campo de busca |
|-----------------|--|--------------------|
| LUME | Catarina Domenici | Resumo |
| LUME | Sanidade Avícola | Descrição do vídeo |
| RIGeo | - | - |
| UFFA Multímedia | Alfabetização com Base Linguística II | Descrição do vídeo |
| UFFA Multímedia | Vídeo-relatório Música e Mito no Alto Rio Negro: criação e transformação da humanidade | Descrição do vídeo |

Tabela 1. Documentos recuperados a partir do termo ‘Amazônia’ em outros campos de busca

Além do critério de busca pelo campo assunto, realizou-se busca com o termo “Amazônia” em outros campos de metadados que recuperou vídeos diferentes dos recuperados no campo assunto encontrados anteriormente, listados na Tabela 1.

Como pode ser observado, existem documentos sobre a Amazônia que não estão indexados com

o conceito “Amazônia”. No Repositório LUME, recuperou-se 2 vídeos que não possuem no assunto o termo Amazônia. No primeiro vídeo o termo consta no resumo por ser “Amazônia” o título de uma canção de autoria de Catarina Domenici, pianista, compositora, pesquisadora e professora do Departamento de Música da Universidade Federal do Rio Grande Sul (UFRGS). No segundo vídeo a temática é sobre as redes neurais artificiais das aves e não há indicação do termo Amazônia no resumo e nem no campo assunto, porém, ao verificar os demais campos dos metadados que ficam ocultos, identificou-se no campo descrição do vídeo o termo Amazônia Rural.

No repositório RIGeo não foram recuperados vídeos para o termo ‘Amazônia’.

Quanto aos documentos recuperados no Repositório UFPA Multimídia, identificou-se o termo ‘Amazônia’ nas descrições dos vídeos. O primeiro documento denominado ‘Alfabetização com Base Linguística II’ é o material didático utilizado no Programa de Educação à Distância da UFPA oriundo do Projeto Educação sem Fronteiras na Amazônia. O segundo material é um vídeo-relatório do Grupo de Pesquisa Música e Identidade na Amazônia, no qual apresenta o clã Desana Guahari Diputiro Porã e sua relação com os repertórios musicais cariço, japurutú e kapiwayá. Os *Guahari Diputiro Porã* tradicionalmente habitam a comunidade de Santa Marta, no rio Papuri. Atualmente, membros deste clã moram no distrito de lauretê, no rio Uaupés, fronteira com a Colômbia (Barros, 2012, sem paginação).

Destaca-se a similaridade na representação temática entre os três repositórios e a presença de linguagem padronizada, assim como o emprego do Código de Catalogação Anglo-Americano (AACR2), para realizar a descrição de documento em formato multimídia e, no caso do vídeo, auxilia a descrição desse recurso eletrônico, assim como na definição dos metadados e na identificação dos pontos de acesso para a recuperação da informação. Ademais, a disponibilidade de efetuar o *download* dos documentos pesquisados nos repositórios é uma opção presente nos três repositórios. A seguir, alinhavam-se as considerações finais desta pesquisa.

5. Considerações finais

Ao se levar em conta a riqueza de informações e potencial de expansão dos repositórios institucionais brasileiros, acredita-se que também é possível considerar a riqueza informacional da Amazônia, tanto sobre o viés de seus conteúdos de complexidade social e biodiversidade, quanto ao viés de conceito paradigmático, atravessando diver-

sos trabalhos, instituições e campos, como apresentado nesta pesquisa. Assim, esta pesquisa analisou a representação de conteúdo, a partir do descritor “Amazônia”, no campo assunto, tendo como tipo de documento o “vídeo”, nos repositórios LUME, RIGeo e UFPA Multimídia. Ressalta-se que há uma grande quantidade de material bibliográfico com o descritor “Amazônia”, entretanto, para o tipo de material “vídeo”, recuperaram-se poucos documentos em relação ao resultado para material bibliográfico (Tese, Dissertação, Artigo Científico, Relatório Técnico etc.).

Dos três repositórios analisados, o que recuperou mais resultados em formato de vídeo para o termo “Amazônia” foi o Repositório UFPA Multimídia, com 27 resultados. Destaca-se que o maior volume de material encontrado no Repositório UFPA Multimídia, diz respeito à luta dos soldados da borracha e o projeto de educação sem fronteiras, ambos tendo a Amazônia como espaço central de seu conteúdo. Quanto aos dois resultados obtidos no Repositório LUME, não houve uma recuperação imediata como no repositório da UFPA, pois o termo Amazônia constava de forma adjacente, sem a mesma relevância que os resultados obtidos no Repositório UFPA Multimídia. O RIGeo não recuperou nenhum material com o termo “Amazônia” por tipologia multimídia, contudo, foram recuperados relatórios técnicos, monografias e outros materiais armazenados no RIGeo indexados com o descritor ‘Amazônia’.

Enfatiza-se que foram recuperados vídeos no qual o termo ‘Amazônia’ estava em outros campos de busca, como por exemplo, no resumo referente ao documento e na descrição do vídeo. Com isso, recuperou-se pelo resumo do vídeo da canção ‘Amazônia’ de autoria de Catarina Domenici, sendo os demais vídeos recuperados ao observarmos na descrição do vídeo o descritor ‘Amazônia’.

A descrição dos metadados em ambientes digitais é realizada iniciando-se pelos termos mais gerais até chegar aos termos específicos que correspondem ao documento audiovisual (Santos, 2013; 2018b).

A temática de representação de material audiovisual em repositórios ainda possui poucas produções, mas os materiais bibliográficos que identificamos, pertinentes para a compreensão e desenvolvimento da temática, foram de grande contribuição na produção deste trabalho científico.

Portanto, percebe-se que não apenas a abordagem dos documentos multimídia, com seus componentes representados pelos metadados, mas também os conceitos que abarcam demandam

uma ordenação sistemática, para fins de recuperação de suas informações. Ademais, um dos desafios para lidar com a realidade complexa dos conceitos presentes em distintos repositórios multimídia está, por exemplo, na disposição de escolha dos seus elementos representativos e da organização dos conteúdos em formatos variados. Desse modo, a adoção de técnicas diversificadas, atualizadas e, talvez, interdisciplinares faz-se necessária para o tratamento temático das informações contidas em repositórios institucionais multimídia nos formatos de documentos audiovisuais.

Referências

- Adebowale, A. J. E.; Adekanye, A. (2013). Audio-visual resources availability and use for library services among colleges of education in Lagos State Nigeria. // *International Journal of Library and Information Science*. 5:10, 417-425, <http://www.academicjournals.org/IJLIS>
- Barros, L. C. da S. (2012). O Kapiwayá e seu lugar no universo músico-coreográfico-ritual em um clã Desana, alto Rio Negro, Amazonas. // *Bol. Mus. Par. Emílio Goeldi Ciênc. Hum.* 7:2. <https://www.scielo.br/lj/bgoeldi/a/QYy5kYbt9hgksLxPpc3t4nG/?lang=pt>.
- Bräscher, M.; Monteiro, F. S. (2010). Organização da informação em repositórios digitais. // *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 15:29.
- Buckland, M. *Information and society*. Cambridge: MIT Press, 2017.
- Castro, E.; Índio, C. (org.) (2015). *Formação socioeconômica da Amazônia*. Belém: NAEA/UFPA. 640 p. (Coleção Formação Regional da Amazônia, 2). <http://livroaberto.ufpa.br/jspui/handle/prefix/164>.
- Chaumier, J. (1988). Indexação: conceito, etapas e instrumentos. // *Revista Brasileira de Biblioteconomia e Documentação*. 21:1/2 (Janeiro/Junho 1988) 63-79.
- Cunha, M. B.; Cavalcante, C. R. O. (2008). *Dicionário de Biblioteconomia e Arquivologia*. Brasília-DF: Briquet de Lemos.
- Dias, G. (2012). Entrevista Repositório UFPA Multimídia. Belém: UFPA. <http://multimidia.ufpa.br/jspui/handle/321654/901>.
- Dias, G. S.; Moreira, S. B. (2014). Repositório Institucional UFPA Multimídia: o seu desenvolvimento sob a ótica do profissional bibliotecário. // *XVIII Seminário Nacional de Bibliotecas Universitárias*. 18, 1-17.
- Feitosa, A. (2006). *Organização da Informação na web: das tags à web semântica*. Brasília: Thesaurus.
- Fujita, M. S. L. (2003). A identificação de conceitos no processo de análise de assunto para indexação. // *Revista Digital de Biblioteconomia e Ciência da Informação*. 1:1, 60-90.
- González-Pérez, L. I.; Ramírez-Montoya, M.; García-Peñalvo, F. J. (2020). User Experience in Institutional Repositories: A Systematic Literature Review. // *I. Management Association (Ed.). Digital Libraries and Institutional Repositories: Breakthroughs in Research and Practice*. 423-440. IGI Global. <http://doi:10.4018/978-1-7998-2463-3.ch026>
- IBICT (2018). *Repositórios Digitais: Repositórios Brasileiros*. Outubro de 2018. <http://sitehistorico.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais/repositorios-brasileiros>.
- Kuri, R.; Singh, M. (2020). Indian Institutional Repositories (IRs) resected in the Directory of Open Access Repository (DOAR): A Case Study. // *Library Philosophy and Practice*. 4640. <https://digitalcommons.unl.edu/libphilprac/4640>
- La Barre, Kathryn; Cordeiro, Rosa Inês Novais de. That obscure object desire: facets for film access and discovery. In: NEAL, Diane Rasmussen (ed.). *Indexing and retrieval of nontext information*. Berlin: De Gruyter. 234-262.
- Leite, F. C. L. (2009). *Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto*. Brasília: IBICT.
- López Guzmán, C. (2005). *Los repositorios de objetos de aprendizaje como soporte a un entorno e-learning*. 2005. Tese (Doutorado). Universidad de Salamanca, Salamanca, 2005. <http://www.biblioweb.dgsca.unam.mx/libros/repositorios>.
- Malcher, M. A. (2012). Entrevista Repositório UFPA Multimídia. Belém: UFPA. <http://www.multimidia.ufpa.br/jspui/handle/321654/923>.
- Marcondes, C. H.; Sayão, L. F. (2009). À Guisa de introdução: repositórios institucionais e livre acesso. In: Sayão, Luis; Toutain, Lídia Brandão; Rosa, Flavia Garcia; Marcondes, Carlos Henrique (orgs.). *Implantação e gestão de repositórios institucionais: políticas, memórias, livre acesso e preservação*. Salvador: EDUFBA.
- Marshall, D. (2001). What is multimedia? <https://users.cs.cf.ac.uk/Dave.Marshall/Multimedia/node10.html>.
- Monteiro, F. (2009). Organização da informação em repositórios temáticos: o uso da modelagem conceitual. // *VIII ENANCIB*, 28 a 31 de outubro de 2007, Salvador, Bahia, Brasil. <http://www.enancib.ppgci.ufba.br/artigos/GT2--261.pdf>
- Novellino, M. S. F. (1996). Instrumentos e metodologia da representação da informação. // *Informação & Informação*. 1:2, 37-45.
- Pinto Molina, M.; Garcia Marco, F.J.; Agustín Lacruz, M. C. *Indización y resumen de documentos digitales y multimedia*. Técnicas y procedimientos. 1. ed. Gijón: TREA, 2002.
- Santos, F. E. P. (2013). Documentos e informações audiovisuais: a teoria arquivística da Biblioteconomia aplicadas à organização de arquivos de TV. // *DataGramaZero*. 14:5.
- Santos, F. E. P. (2018a). Documento e informação audiovisual: bases conceituais numa perspectiva neodocumentalista. // *Em Questão*. 24:2 235-259.
- Santos, F. E. P. (2018b). *Gestão de acervos em repositórios audiovisuais*. Orientadora: Maria Giovanna Guedes Farias; Co Orientador: Luiz Tadeu Feitosa. Dissertação (Mestrado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Centro de Humanidades, Universidade Federal do Ceará. <http://www.repositorio.ufc.br/handle/riufc/39305>.
- Santos, P. L. V. A. da; Carvalho, A. M. G. de. (2009). Sociedade da informação: avanços e retrocessos no acesso e no uso da informação. // *Informação & Sociedade: Estudos*, João Pessoa. 19:1, 45-55.
- Santos, D. B. dos; Pavão, C. M. G.; Moura, A. M. M. de (2016). Usabilidade do Lume – Repositório Digital da UFRGS: uma avaliação por meio das heurísticas e de testes com usuários. // *InCID: Revista de Ciência da Informação e Documentação*. 7:1, 150-166.
- Sayão, L. F.; Sales, L. F. (2016). Algumas considerações sobre os repositórios de dados de pesquisa. // *Informação & Informação*. 21:2, 90-115.
- Tartarotti, R. C. D.; Dal' evedove, P. R.; Fujita, M. S. L. Indexação em repositórios digitais: uma incursão na produção científica brasileira de Ciência da Informação. // *Scire*. 26:2, 13-18.

- Tortosa, S. O. (2006). Propuesta de una arquitectura software basada en servicios para la implementación de repositorios de objetos de aprendizaje distribuidos. 2006. Tesis (Doctoral). Escuela Técnica Superior de Ingeniería Informática, Alcalá de Henares.
- UFPA Multimídia. (2012). O projeto: qual o objetivo? 2012. <http://www.multimidia.ufpa.br/jspui/links/projeto-home.jsp>.
- UFPA (2020). Universidade Federal do Pará. Pró-Reitoria de Planejamento e Desenvolvimento Institucional. Relatório de Gestão 2019. Belém: UFPA.
- Universidade Federal do Pará. Reitoria. A Luta dos Soldados da Borracha por Justiça e Reparação: entrevista com César Garcia Lima. Belém: UFPA, 2012. 1 vídeo (10min e 44seg). <http://www.multimidia.ufpa.br/jspui/handle/321654/1098>.
- Universidade Federal do Pará. Assessoria de Educação a Distância. Faculdade de Comunicação. Educação sem fronteiras na Amazônia: entrevista com José Miguel Martins Veloso (vídeo). In: Leite, Selma Dias et al. Educação sem fronteiras na Amazônia: trajetória e perspectiva da educação a distância da UFPA. Belém: UFPA, 2010. Disco 2. (Entrevistas). (25min e 41seg). [Dhttp://www.multimidia.ufpa.br/jspui/handle/321654/714](http://www.multimidia.ufpa.br/jspui/handle/321654/714).
- Zarpelon, E. Basniak, M. I.; Silva, S. C. R.; Resende, L. M. (2018). Repositórios de objetos de aprendizagem de matemática: uma breve caracterização e discussão a partir dos recursos disponibilizados em três bancos de dados. // Revista Electrónica de investigación en Educación en Ciencias. 2 (Dezembro, 2018) 47- 61.

Enviado: 2021-04-01. Segunda versão: 2022-06-06.
Aceptado: 2022-06-09.

Vozes cultivadas *versus* tipos de vozes na CDD: o uso da metateoria para a compreensão da voz cantada

Voces cultivadas versus tipos de voces en la CDD: el uso de la metateoría para la comprensión de la voz cantada

Cultivated voices versus types of voices in the DDC: the use of metatheory for the understanding of the sung voice

Fernanda Carolina PEGORARO NOVAES

Universidade Estadual Paulista Júlio de Mesquita Filho, Avenida Hygino Muzzi Filho, 737,
Marília, SP, Brasil, fernanda.pegoraro@unesp.br

Resumen

Se pretende identificar, a través de la metateoría, las teorías aplicadas en la clasificación de voces de la Clasificación Decimal de Dewey (CDD). Para ello, se establecieron los siguientes objetivos específicos: a) realizar una revisión bibliográfica sobre la metateoría para comprenderla y aplicarla como método en la presente investigación; b) investigar, a través de la metateoría, las primeras clasificaciones sobre voces para comprender las teorías que utiliza la música; c) comparar y discutir la clasificación de voces CDD desde la perspectiva de las teorías musicales). Se concluye que las descripciones de voces del CDD parten del principio de la teoría musical, pero la forma en que se categorizan, además de difundir ideas preconcebidas, no se sustenta en la teoría de la clasificación.

Palabras clave: Metateoría. Fisiología de la voz. Voces cultivadas. Tipos de voces. Clasificación Decimal de Dewey.

1. Introdução

Atualmente exigem-se perspectivas locais e globais na representação de assuntos e com isso há o desenvolvimento de novas teorias, pois diversas questões não podem ser resolvidas com teorias pré-existentes (principalmente em se tratando de teorias aplicadas a domínios muito específicos). Desse modo, com o surgimento de inúmeras teorias e informações provindas de novas pesquisas, gerou-se um excesso de informação; tornando-se difícil organizar o conhecimento existente dentro de um próprio campo (Campbell, 2009; Wallis, 2010).

Considerando-se que a ciência compõe-se de fatos empíricos e teorias (um não existe sem o outro e ambos são necessários para estruturar uma ciência forte), o que dificulta o desenvolvimento de pesquisas é a construção de teorias sem fundamentação metodológica sólida e específica; observando-se no extremo dessa situação a formulação de teoria baseada apenas na intuição subjetiva. A ausência de uma metodologia que

Abstract

Metatheory is used to identify and understand the theories applied in the voice classification of the Dewey Decimal Classification (DDC). To this end, the following specific objectives were established: a) to carry out a bibliographic review on the meta-theory to understand it and apply it as a method in the present research; b) to investigate, through metatheory, the first classifications about voices in order to understand the theorys used by Music; c) compare and discuss the DDC voice classification from the perspective of the music theorys. It is concluded that the descriptions of voices from DDC start from the principle of music theory, but the way in which they are categorized disseminates pre-conceived ideas and is not supported by classification theory principles.

Keywords: Metatheory. Voice physiology. Cultivated voices. Voice types. Dewey Decimal Classification.

possa analisar e comparar teorias em uma pesquisa, faz com que haja uma lacuna na ciência, uma vez que não será possível avançar cientificamente sem que a teoria e a prática (fatos empíricos) estejam relacionadas (Dollan, 2007; Wallis, 2010).

Desse modo, uma alternativa que permite melhorar a eficácia das pesquisas é o desenvolvimento de abordagens mais maduras aplicadas em estudos da metateoria. Uma pesquisa metateórica (metateorização), envolve diversos processos, como: teorias de classificação, categorização, desconstrução, reconstrução, análises estatísticas literárias e entre outros (Wallace, 1992; Ritzer, 2001; Wallis, 2010). Considerando-se isso e considerando-se também que a metateoria permite a integração interdisciplinar de teorias, expõe-se o objetivo geral deste estudo, isto é, compreender por meio da metateoria, a(s) teoria(s) aplicada(s) na classificação de voz da CDD.

Particularmente, a ontogenia é considerada o estudo do tempo de vida de um determinado assunto. Esse tipo de método é comum na Biologia para analisar as transformações das etapas de vida de um organismo; porém a Ciência da Informação aplica a ontogenia para observar e compreender as mudanças proporcionadas pelo tempo em sistemas de classificação (Novaes, 2020).

Em uma análise ontogênica (1) da divisão música nas edições 19, 20 e 23 da CDD, Novaes (2020), identificou 3 tipos de mudanças que aconteceram ao longo do tempo, que foram: mudança estrutural, textual e lexical; referentes a alterações na estrutura hierárquica e associativa e no conjunto de textos e palavras. As principais mudanças identificadas, ocorreram na 20ª edição; conhecida como definidora de tendências do assunto música devido a sua reestruturação completa (Novaes, 2020).

Considerando-se as alterações identificadas, as divisões 782 a 788 (referentes a mídias de performance ou desempenho, que representam instrumentos e técnicas vocais), apresentaram o maior índice de mudanças conceituais. De acordo com Novaes (2020), o que possivelmente justifica essas mudanças, é a influência do Catálogo Britânico (para a descrição de assuntos multiculturais) e da Classificação *Hornbostel-Sachs* (para a descrição de instrumentos musicais) utilizados como referência na revisão das edições da CDD.

Apesar das alterações realizadas, ao analisar a classificação de voz (*vocal music* (782)), nota-se que ao invés de categorizar as técnicas vocais, estruturou-se a classificação da voz obedecendo-se a três critérios: 1: gênero (*woman's voices* 782.6 e *men's voices* 782.8); 2: faixa etária (*children's voices* 782.7); 3: técnicas vocais (*other types of voices* 782.8). A partir dessa observação, percebe-se que os critérios adotados reforçam estereótipos que afirmam que determinadas vozes são femininas, masculinas e infantis. Esse modelo, ainda que seja baseado na teoria musical, não se sustenta ao olhar da teoria da classificação; pois mesmo que se assente em uma perspectiva lógica e teórica da música, ao ser sistematizado em um sistema de classificação bibliográfico, estrutura-se de modo confuso (Novaes, 2020).

Sendo assim, a fim de compreender melhor a(s) teoria(s) utilizada(s) para classificar vozes, a presente pesquisa propôs os seguintes objetivos específicos: a) realizar revisão bibliográfica sobre a metateoria para compreendê-la e aplicá-la como método na presente pesquisa; b) investigar por meio da metateoria, as primeiras classificações sobre vozes a fim de compreender a(s) teoria(s)

utilizada(s) pela música; c) comparar e discutir a classificação de voz da CDD a partir do olhar da(s) teoria(s) da música.

O método aplicado para a realização deste estudo foi a metateoria por meio da dimensão interna-intelectual do quadrante de Ritzer (1990, 1991).

2. Organização do conhecimento *versus* organização do conhecimento sobre o assunto música

O campo da organização do conhecimento pode ser compreendido em sentido amplo e em sentido estrito. O sentido amplo, refere-se a divisão social do trabalho intelectual, conhecido também como a organização do conhecimento em diferentes domínios. Em se tratando do sentido estrito, refere-se a construção e uso de sistemas de organização do conhecimento e a natureza e qualidade de processos de organização do conhecimento (Hjørland, 2008; Moreira, 2018).

A organização e representação do conhecimento é estruturada nas ações de organizar e representar o conhecimento, operacionalizadas com o recurso de instrumentos, processos e produtos; preocupando-se com a aplicação e avaliação dos sistemas de organização do conhecimento (Fujita, 2008; Moreira, 2018). Desse modo, considerando que a representação se faz presente tanto nos sistemas como nos processos de organização do conhecimento, compreende-se que, conceitualmente, organização e representação tornam-se inseparáveis; pois o conhecimento é objeto da informação, representado (por meio do processo de materialização) em registro (Moreira, 2018).

Historicamente, a classificação é considerada um exemplo de sistema de organização do conhecimento mais utilizado para a organização da informação em bibliotecas. Em se tratando de bibliotecas especializadas em música, atualmente existem diversos novos sistemas de classificação estruturados para a representação de assunto com o propósito de adaptar e aprimorar antigas classificações.

Existe uma complexidade inerente à classificação do assunto música. Entre outros fatores, isso se deve ao fato de que um mesmo termo pode se referir a conceitos distintos conforme se tome como base a perspectiva da música ou a perspectiva da organização do conhecimento.

De acordo com Lee (2017), há quatro tipos de abordagens que propiciam a interlocução entre organização do conhecimento e classificação do assunto música, que são: 1) abordagem utilizando sistemas de classificação; 2) abordagem

por meio da elaboração de projetos; 3) abordagem considerando o domínio da música; 4) abordagem utilizando técnicas de domínio da música para análise de assuntos na biblioteconomia e ciência da informação.

As abordagens 1 e 2 utilizam sistemas de classificação bibliográfica para a representação do assunto música. Dessa forma, a primeira abordagem é aplicada por meio da discussão da classificação bibliográfica do assunto música caracterizada na análise tripartida ou na sugestão de um novo sistema de classificação. A análise tripartida investiga a história do sistema de classificação, descrevendo-o posteriormente e discutindo suas limitações na representação do assunto música. Como decorrência, elaboram-se propostas para a sua melhoria. Pode-se citar como exemplo, a Classificação de Dickinson, proposta por Bradley (1972) e as revisões elaboradas por Philp (1982), propondo melhorias para a CDD.

A análise tripartida é desenvolvida a partir de três categorias de análise, a fim de elaborar diferentes discussões sobre como a música é classificada em mais de um sistema de classificação – os trabalhos de Bryant e Marco (1985) e Redfern (1978) são exemplos desse tipo de análise.

Já a segunda abordagem, isto é, a abordagem mediada pela elaboração de projetos, refere-se às práticas de classificação em uma biblioteca especializada em música, realizando uma avaliação reflexiva sobre como melhorar a representação dos materiais no acervo, elaborando outros tipos de classificações profissionais. Exemplos desse tipo de abordagem são encontrados nos trabalhos de Marsh (2002), que elaborou uma adaptação para o sistema de gravações (*Alpha-Numeric System for Classification of Recordings - ANSCR*) do *Leeds College of Music*, distrito de Quarry Hill, Inglaterra, e Krohn (1970), que estruturou uma nova classificação para as partituras da biblioteca da *Washington University*, em St. Louis (Lee, 2017).

As abordagens 3 e 4 são elaboradas a partir da análise de domínio de conceitos do assunto música, para que posteriormente sejam estabelecidas discussões sobre representação para a organização do conhecimento em música e para a organização do assunto música em sistemas de classificação. Incluem-se nestas abordagens a comparação entre conceitos na perspectiva da música e na perspectiva da organização do conhecimento. Podem ser mencionados como exemplo a análise da classificação do assunto música a partir de teorias da ciência da informação, realizada por Abrahamsen (2003) e a aplicação da Análise Schenkarian em sistemas de classificação bibliográfica, realizada por Elliker (1994).

A Análise Schenkarian, como informe esse autor, é muito utilizada no domínio da música para analisar obras musicais; a “música em si”.

Desse modo, considerando-se todos os tipos de abordagens que permitem a interlocução entre a teorias da organização do conhecimento e teorias provindas do domínio da música, o presente estudo aprofundou-se na metateoria, considerada um tipo de análise de domínio, discutida a seguir.

3. Abordagem MU interna-intelectual: procedimentos metodológicos

As pesquisas metateóricas desenvolvidas pela ciência da informação, são caracterizadas de naturezas: nomotética e ideográfica. A pesquisa de natureza nomotética é responsável por estabelecer a lei geral de qualquer princípio e/ou teoria; o seu estudo permite a previsão e descrição de detalhes e por isso, é caracterizada como essencial para todas as ciências. A pesquisa de natureza ideográfica, compreende e avalia detalhes descrevendo as diferenças entre fatos de uma situação ao longo do tempo. É comum nessa abordagem, a descoberta de temas e tendências; considerada fundamental para a área de humanas (Bates, 1994, 2005; Sandstrom, 1995).

A partir das naturezas mencionadas, são descritas treze abordagens principais utilizadas para elaborar pesquisas metateóricas na ciência da informação; denominadas: ideográficas, nomotéticas e ideográficas e nomotéticas. Sendo assim, são identificadas como:

- abordagens ideográficas: histórica, construtivista, discursivo-analítica, filosófico-analítica e teoria crítica;
- abordagens nomotéticas e ideográficas: etnográfica e sócio-cognitiva;
- abordagens nomotéticas: cognitiva, bibliométrica, física, engenharia, design centrada no usuário e evolucionária (Bates, 2005).

Considerando-se as abordagens utilizadas para a elaboração de pesquisas metateóricas, a natureza da presente pesquisa caracteriza-se: mista, nomotética e ideográfica; por pertencer a abordagem sócio-cognitiva, descrita por Hjørland e Albrechtsen (1995), como análise de domínio.

Hjørland (2002), elaborou onze maneiras pelas quais a ciência da informação permite abordar um domínio específico, que são: 1) produção e avaliação de guias de literatura e entradas de assuntos; 2) produção e avaliação de classificações especiais e tesouros; 3) pesquisa sobre habilidades de indexação e recuperação de informação em especialidades; 4) conhecimento de estudos

empíricos de usuários em áreas temáticas; 5) produção e interpretação de estudos bibliométricos; 6) estudos históricos de estruturas e serviços de informação em domínios; 7) estudos de documentos e gêneros nos domínios do conhecimento; 8) estudos epistemológicos e críticos de diferentes paradigmas, pressupostos e interesses em domínios; 9) conhecimento de estudos terminológicos, linguagens para fins especiais e análise de discurso em áreas do conhecimento; 10) estudos de estruturas e instituições de comunicação científica e profissional em um domínio; 11) conhecimento de métodos e resultados de estudos analíticos de domínio sobre cognição profissional, representação do conhecimento em informática e inteligência artificial. Essas onze maneiras mencionadas permitem tanto compreender um domínio específico por meio de conhecimentos adquiridos, como também formar especialistas em áreas específicas (Hjørland, 2017).

De acordo com Ritzer (1990; 1991), os estudos metateóricos são definidos perante a natureza dos seus produtos finais e classificam-se em três abordagens, que são: 1) *metatheorizing as a means of attaining a deeper understanding of theory* – MU (metateorizar como um meio de alcançar uma compreensão mais profunda da teoria); 2) *metatheorizing as a prelude to theory development* – MP (metateorizar como um prelúdio para o desenvolvimento da teoria); 3) *metatheorizing as a source of perspectives that overarch sociological theory* – MO (metateorizar como uma fonte de perspectivas que abrangem a teoria sociológica).

Sendo assim, Ritzer (1990, p. 4, tradução livre) (1), retrata que a

[...] MU preocupa-se, mais especificamente, com o estudo de teorias, teóricos, comunicação de comunidades de teóricos, bem como os contextos intelectuais e sociais mais amplos da teoria e teóricos.

No que refere-se a MP, um estudo teórico de uma teoria existente é realizado para produzir uma nova teoria. Em se tratando da MO, Ritzer (1990, p. 4, tradução livre), menciona que trata-se de um estudo da teoria com “[...] o objetivo de produzir uma perspectiva, pode-se dizer uma metateoria, que ultrapassa alguma parte ou toda a sociologia teoria” (2).

Tendo isso, compreende-se que os três tipos de metateorização (MU, MP e MO) referem-se a diferentes formas de análise de domínio. Considerando-se que a presente pesquisa intenciona entender a(s) teoria(s) aplicada(s) na classificação de voz da da CDD, utilizará a abordagem metateórica MU.

A abordagem metateórica MU possui quatro dimensões, que são: *internal-external* (interna-externa) e *intellectual-social* (intelectual-social). Interna refere-se a tudo existente dentro de um determinado domínio e externa aos fenômenos encontrados fora do domínio, mas que exercem impacto sob o mesmo. Entende-se por intelectual, a estrutura cognitiva do domínio, como por exemplo, teorias e ferramentas teóricas. Social, como o próprio nome faz referência, trata-se da estrutura social do domínio, como as escolas e o impacto na sociedade. Quando estruturadas em uma tabela quádrupla (quadrante), há o cruzamento das quatro dimensões por meio de um corte transversal estabelecendo-se uma ponte entre dois ou mais polos. Dessa forma, originam-se as dimensões: *internal-intellectual* (interna-intelectual), *internal-social* (interna-social), *external-intellectual* (externa-intelectual) e *external-social* (externa-social) (Ritzer, 1991).

A Tabela I abaixo, demonstra o quadrante baseado em Ritzer (1991) com os polos mencionados.

| | | Intelectual | | | |
|---------|--------------------------|----------------------------------|----------------------|---------|--|
| Interna | Paradigmas Cognitivos | Uso de conceitos emprestados da: | | Externa | |
| | Escolas de pensamento | Filosofia | Economia | | |
| | Mudanças paradigmáticas | Linguística | Etc. | | |
| | Escolas de pensamento | | | | |
| | Ferramentas Metateóricas | | | | |
| | Teorias | | | | |
| | | Paradigmas comuns | Impacto da Sociedade | | |
| | | Colégios Invisíveis | Impacto do Social | | |
| | | Escolas | Instituições | | |
| | | Redes | Raízes históricas | | |
| | | Backgrounds Individuais | | | |
| | | Social | | | |

Tabela I. Quadrante de Ritzer: abordagem MU

A dimensão interna-intelectual proporciona melhor compreensão de questões intelectuais e/ou cognitivas internas de um determinado domínio; isto é, permite estudar teorias de modo sistemático, a fim de compreendê-las e analisá-las em estudos comparativos. A interna-social, observa os fatores sociais de um domínio buscando identificar as suas principais escolas; assim como realizar estudos entre os grupos do domínio e também investigar as afiliações institucionais. Já a dimensão externa-intelectual, recorre a outras disciplinas acadêmicas a fim de obter novas ideias, ferramentas e teorias do domínio. A di-

menção externa-social, busca olhar para a sociedade com um olhar mais amplo por meio de um nível macro (Ritzer, 1990, 1991).

É importante enfatizar que uma pesquisa aplicando a abordagem metateórica MU, dependendo do tamanho de seu escopo, permite utilizar mais de uma dimensão (senão as quatro). Sendo assim, considerando-se as dimensões apresentadas da MU e o objetivo central deste estudo, será aplicada apenas uma dimensão, a interna-intelectual – estruturada metodologicamente em três partes: primeira parte: teorias fundamentais; segunda parte: classificação de vozes cultivadas; terceira parte: classificação de vozes cultivadas *versus* classificação de voz de Dewey.

A primeira parte é responsável por reunir conceitos importantes provindos das teorias da fonoaudiologia e da música. Quando investiga-se sobre registros vocais, é perceptível a interdisciplinaridade de teorias entre a música e a fonoaudiologia. Isso se justifica pela necessidade em compreender a respeito da dinâmica da voz cantada (fundamentada pela fonoaudiologia) e dos tratados de registros vocais (desenvolvidos pela música). Sendo assim, inicialmente demonstra-se, a partir de um quadro, a relação terminológica e conceitual entre a fonoaudiologia e a música sobre o assunto voz cantada. Posteriormente, a fim de compreender a historicidade da classificação de vozes, apresenta-se, também por meio de quadros, os primeiros tratados italianos responsáveis por classificar os primeiros registros vocais. Os tratados italianos são de suma importância para o presente trabalho, pois influenciaram as classificações atuais utilizadas para representar os tipos de extensões vocais, isto é, tipos de vozes, como a CDD.

Em se tratando da segunda parte, entre os tratados mencionados, destaca-se o tratado de Garcia (1840), pois foi o primeiro a integrar a voz da mulher na classificação de extensões vocais, denominada pelo autor como classificação de vozes cultivadas. Sendo assim, a princípio, demonstra-se como as definições por faixa etária (infância, puberdade e maturidade) são consolidadas e subsequentemente como os tipos de vozes manifestam-se na maturidade (definida como vozes de mulheres e vozes de homens). Por fim, a partir do que Garcia (1840) representa como vozes maduras, ilustra-se como o autor classifica as extensões de voz.

A terceira parte refere-se aos resultados finais, que realiza a comparação entre a classificação de vozes cultivadas proposta por Garcia (1840) e a classificação de vozes estruturada pela 23ª edição da CDD. Esse capítulo intenciona explicar

o motivo pelo qual a CDD classifica a voz obedecendo aos critérios: 1: gênero (*woman's voices* (782.6) e *men's voices* (782.8)); 2: faixa etária (*children's voices* (782.7)); 3: técnicas vocais (*other types of voices* (782.8)). Tendo isso, primeiramente, ilustra-se, por meio de um quadro, a estrutura da classificação de voz da 23ª edição da CDD e posteriormente, são realizadas observações, baseadas na teoria musical, sobre cada termo do sistema de classificação.

4. Primeira parte: teorias fundamentais

Na teoria da música, a emoção manifesta-se por meio da dinâmica da voz cantada, caracterizada como: *pitch, loudness and quality* (*pitch, loudness* e *quality*) – também consideradas como dinâmica do som. Essa dinâmica apresenta características das vogais da língua e do texto, relacionadas ao aparato anátomo-funcional e muscular do ser humano (Laver, 1968; Salomão, 2008).

Em se tratando da teoria da fonoaudiologia, a qualidade vocal baseia-se em aspectos acústicos, fisiológicos e aerodinâmicos; que são: a variação tonal da voz cantada e práticas clínicas de reabilitação para distúrbios de voz. Os distúrbios de voz, nesse sentido, são compreendidos como quebras na qualidade da voz, conhecidos também como modificações abruptas; identificados em inúmeras situações, como um indivíduo cantar uma faixa extensa de *pitch* ou manifestar uma patologia na voz (Salomão, 2008).

| Termos | Fonoaudiologia |
|----------|--|
| Pitch | Frequência ou <i>pitch</i> , refere-se à altura do som produzido (descrito em Hertz (Hz)). A frequência permite classificar o som como agudo e grave (Silva, Dupret, 2004; Behlau, 2001). |
| Loudness | Intensidade ou também identificado na área como <i>loudness</i> , diz respeito a intensidade vocal da força da respiração (descrita em Decibéis (dB)) (Gayotto, 1997). |
| Quality | Quando investiga-se sobre o significado de qualidade, o mesmo, relaciona-se com o significado de timbre. Isso se explica pela qualidade retratar o potencial e as características de uma voz cantada e o timbre, apesar de seu conceito abranger questões fisiológicas, caracterizar-se como uma qualidade da voz cantada (Antelmi; Silveira, 2008). |

Tabela II. Terminologia de Fonoaudiologia para voz cantada

A terminologia utilizada pela fonoaudiologia para classificar a qualidade da voz cantada, é baseada em recursos vocais primários (frequência, intensidade, ressonância, articulação e respiração) e recursos vocais secundários (projeção,

volume, ritmo, velocidade, entonação e modulação) (Antelmi; Silveira, 2008; Gayotto, 1997).

Tendo isso, apesar da teoria empregada pela fonologia envolver maior número de categorias para descrever voz cantada, percebe-se que conceitualmente dialoga com a teoria da música e auxilia na compreensão do conceito de *pitch*, *loudness* e qualidade; assim demonstrado abaixo na Tabela II, na página anterior.

Desse modo, para melhor compreender como os aspectos da dinâmica da voz influenciam a extensão vocal e a classificação da voz descrita na CDD (23ª edição), é importante, primeiramente, abordar os primeiros registros vocais; observados abaixo na Tabela III.

| Autores | Descrição do tratado | Registros vocais |
|---------------|---|--|
| Zacconi, 1592 | Timbre vocal Entonação Voz de homem | Voce di petto (voz de peito) Voce di testa (voz de cabeça) |
| Caccini, 1602 | Intensidades de voz Faixas tonais Voz de homem | Voce piena e naturale (voz cheia e natural) Voce finta (voz fingida; falsete) |
| Tosi, 1723 | Registros vocais Para sopranos castratis Voz de homem | Voce di petto (voz de peito) Voce di testa (voz de cabeça) |
| Mancini, 1774 | Registros vocais Para sopranos castratis Voz de homem | Registre di petto (registro de peito) Registre di testa (registro de cabeça; falsete) |

Tabela III. Tratados italianos de dois registros vocais

Ao observar a Tabela III, nota-se que as classificações referentes a registros vocais eram divididas em duas categorias; basicamente voz de peito e voz de cabeça.

Também é perceptível que os registros vocais foram inicialmente classificados por tratados italianos e por autores homens. Isso se explica por meio do movimento *Bel Canto* (Belo Canto), repercutido no início do século XVI. O Belo Canto caracteriza-se não apenas como um movimento, mas como uma escola italiana de referência em técnicas vocais do canto solo (Salomão, 2008).

Em se tratando dos registros vocais serem classificados apenas por autores homens, a Ópera é historicamente relacionada aos cantores *castrati* (castrados). Os *castrati* eram homens cantores (e/ou professores) considerados na época fenômenos vocais, por possuírem uma extensão vocal correspondente à feminina. Os *castrati* relacionam-se ao domínio do Antigo Regime na Europa; o menino cantor da classe social baixa,

para ter permissão de conviver com a classe social alta europeia, era submetido a castração. Desse modo, preservava a sua voz aguda e demonstrava um bom comportamento social e moral (Pacheco, 2004; Salomão, 2008).

Dito isso, as classificações vocais iniciais foram escritas por homens e predominantemente para homens; ressaltando que haviam cantoras mulheres, porém os cantores homens eram mais aceitáveis socialmente (devido a era dos *castrati*). Zacconi (1592) era *maestro di cappella* (maestro de capela) e no período em que regia corais, mulheres não eram aceitas para cantar. Desse modo, tinha preferência que notas agudas fossem cantadas em voz de cabeça ou falsete, pois considerava as vozes de peito irritantes (Stark, 2003; Zacconi, 1592).

Caccini (1602) também manifestou sua preferência por vozes de cabeça, classificando os registros vocais com nomenclaturas diferentes dos demais autores dos tratados de dois registros; descrevendo voz de cabeça como voz cheia e natural e voz de peito como voz fingida. No próprio modo como refere-se a voz fingida, demonstra uma ideia pré-concebida do gênero vocal; considerando a mesma um tipo de voz soprada (Caccini, 1602; Stark, 2003).

Tosi (1723) e Mancini (1774) eram cantores *castrati* e dedicaram seus tratados a cantores sopranos. Tosi (1723) também era compositor e professor de canto e quase não faz referência sobre vozes femininas. Isso se explica pelo contexto social dos *castrati*, pois, havendo um cantor homem com a extensão vocal correspondente à feminina, não havia a necessidade de ensinar mulheres a cantar. De acordo com Pacheco (2004), Tosi considerava que as mulheres eram boas cantoras, porém não dominavam a arte do improviso musical. Dito isso, as vozes femininas eram tratadas iguais às dos *castrati*; apesar da estrutura óssea, muscular e do aparelho fonador do cantor *castrato* ser biologicamente diferente da mulher (Pacheco, 2004).

Mancini (1774) foi o primeiro autor na ordem cronológica dos tratados italianos a referir-se registro de cabeça e registro de peito ao invés de voz de cabeça e voz de peito. Registro, no contexto da voz cantada, de acordo com Large (1972), pertence a uma terminologia utilizada para representar órgãos desde o surgimento dos estudos sobre qualidade da voz. Desse modo, Mancini (1774) menciona sobre casos raros de gamas tonais de registros de peito; no qual um cantor consegue por meio da execução de um falsete, soar, de modo inexplicável, um registro de peito.

Entre os quatro autores mencionados pela Tabela 3, os tratados de Tosi (1723) e Mancini (1774) foram considerados os mais importantes da época por três razões: 1) anteriormente, na história do Belo Canto, nenhum outro cantor escreveu um tratado à sopranos castrati (o que demandava compreender sobre faixas extensas e unificação de voz do pitch); 2) O tratado de Mancini (1774) tem o mesmo segmento do tratado de Tosi (1723); 3) a tradução do tratado de Tosi (1723) para o inglês e alemão deu origem a outros registros musicais; observados abaixo na Tabela IV.

| <i>Autores</i> | <i>Descrição do tratado</i> | <i>Registros vocais</i> |
|----------------|---|---|
| Galliard, 1742 | Registros vocais Para sopranos castratis Voz de homem | Voce di petto (voz de peito) Voce di testa (voz de cabeça) Falsetto (falsete) |
| Agrícola, 1757 | Registros vocais Para sopranos castratis Voz de homem | Voce di petto (voz de peito) Voce di testa (voz de cabeça) Falsetto (falsete) |
| Nathan, 1836 | Registros vocais Divisão de vozes Voz de homem | Voce di petto (voz de peito) Voce di testa (voz de cabeça) Voce feigned (voz fingida) Falsetto (falsete) |
| Garcia, 1840 | Sons vocais Voz de mulher Voz de homem | Registre de poitrine (registro de peito) Registre de fausset-tête (registro de falsete-cabeça) Contre-basse (contrabaixo) Timbre clair (timbre claro) Timbre sombre (timbre sombrio) Volume (volume) |

Tabela IV. Tratados com mais de dois registros vocais

Galliard (1742) traduzindo o tratado de Tosi (1723) para o inglês e Agrícola (1757) para o alemão, originaram novos registros musicais a partir de três categorias de vozes masculinas: voz de peito, voz de cabeça e falsete.

Anteriormente, Mancini (1774) categorizou falsete pertencente a voz de cabeça; isso se explica por cantores líricos, na presente época, *castrati*, obterem diversas gamas tonais – também é importante enfatizar que no Barroco, voz de cabeça e falsete eram considerados sinônimos. Porém, Galliard (1742) reconhece que falsete é uma técnica vocal diferente da voz de cabeça, compreendendo que a sua formação é proveniente da garganta.

Agrícola (1757) em suas observações ao tratado de Tosi (1723), reconheceu que tanto as vozes

de peito, quanto as vozes de cabeça relacionam-se com propriedades da traqueia e relatou que o falsete poderia ocorrer em diferentes níveis de *pitch*; sendo eles mais graves ou mais agudos.

Assim, considerando os autores mencionados, o início do século XIX foi muito promissor para pesquisas relacionadas a ciência da voz, principalmente em se tratando de estudos sobre órgão vocal, produção e qualidade da voz. Até então, os cantores e professores de canto lecionavam baseando-se na própria voz e experiência, pois não havia conhecimento abrangente sobre anatomia, fisiologia e acústica da voz. Nesse sentido, percebe-se a diferença conceitual nos tratados de Nathan (1836) e principalmente no de Garcia (1840).

Nathan (1836), reconhecendo o conceito de falsete descrito por Agrícola (1757), descreveu os registros vocais em quatro categorias; duas primárias: voz de peito e voz de cabeça e duas para diferentes níveis de *pitch* mais agudos: voz fingida e falsete. De acordo com Stark (2003), a voz fingida é caracterizada como suave, melódica e produzida no peito, entre a parte de trás da garganta e cabeça.

Garcia era professor e cantor não *castrato*. O seu tratado é considerado um marco na história dos registros vocais, pois pela primeira vez na classificação de voz, foram criadas categorias vocais para vozes femininas. Algo que influenciou essa questão, é o período de sua escola, pois insere-se na Revolução Francesa e Revolução Industrial; época conhecida como o fim dos cantores *castrati*, portanto, lecionava canto tanto para homens, quanto para mulheres. O autor, baseou seus estudos na fisiologia da voz, o que contribuiu não apenas para a descrição moderna dos registros vocais, mas também para a definição de outros fenômenos da voz (como o timbre e a intensidade (*quality* e *loudness*)) (Pacheco, 2004; Salomão, 2008).

Desse modo, os registros vocais de Garcia (1840) são denominados como *espèces de sons vocaux* (espécies de sons vocais) e assim categorizados em: registros de *poitrine* (peito) e *fausset-tête* (falsete-cabeça); *contre-basse* (contrabaixo); timbres *clair* (claro) e *sombre* (escuro); *volume* (volume).

Garcia (1840) nomeia o *fausset* (falsete) e o *registre de tête* (registro de cabeça) como *registre de fausset-tête* (registro de falsete-cabeça), porque para o autor, esse registro refere-se a um novo conceito de *pitch* mais agudo. O *contre-basse* (contrabaixo) é descrito como um som mais rouco e baixo. Os timbres *clair* (claro) e *sombre* (escuro) são variações de som determi-

nadas pela vibração da laringe, podendo ser empregadas nos registros de cabeça, peito e falsete; definidos também como características dos registros vocais (qualidade). E por último, o volume caracteriza-se como *loudness*; a intensidade da voz.

5. Segunda parte: classificação de vozes cultivadas

O tratado de Garcia (1840), no presente trabalho, é considerado o mais importante para a compreensão da classificação de voz estruturada na CDD; por esse motivo, explora-se o mesmo nesse capítulo.

Ao estruturar os sons vocais, Garcia (1840, p. 10) (3) esclarece que:

A voz humana é sujeita a influência da idade, sexo, constituições e submetida a inúmeras modificações. Independentemente das marcantes diferenças que distinguem as vozes dos diversos indivíduos, há também um número ilimitado de tons pertencentes à um órgão do mesmo indivíduo.

Dito isso, Garcia (1840) retrata a singularidade da voz de cada indivíduo, pois cada voz permite exprimir diversos sons e assumir inúmeras extensões vocais; porém, a voz humana é classificada a partir de princípios primitivos e fundamentais. Desse modo, a fim de esclarecer sobre os registros vocais, o autor descreve sobre as seguintes vozes: *voix des enfants* (vozes de crianças), *la mue* (*moulting*), *femme* (mulher) e *homme* (homem).

Ao retratar vozes de crianças, Garcia (1840) explica que o período entre infância e puberdade, caracteriza as vozes de meninas e meninos de maneira semelhante e apresentando poucas distinções vocais; afirmando que as crianças geralmente falam em falsete. Isso se explica pelo órgão vocal delgado afetar o tecido torácico, pois nessa fase de vida, é pouco desenvolvido tornando-se flexível (Garcia, 1840).

A fase de puberdade, pré-adolescência, apresenta mudanças na voz de um indivíduo devido ao fortalecimento do órgão vocal delgado. Tendo isso, é realizada a *la mue* (também chamada de *moulting*, que significa fase de mudança), descrita por Garcia (1840), como um preparo para a voz de criança assumir a sua verdadeira natureza; voz de mulher ou voz de homem. Sobre a *la mue*, Garcia (1840, p. 11) (4) descreve:

Se empobrecemos o órgão vocal pelo exercício do canto, ou por qualquer excesso, esgotamos a planta antes que ela dê frutos, fazemos com que a caducidade suceda à infância: matamos a virilidade. Só depois da *la mue* é que os estudos sérios da música devem começar.

A *la mue* é desenvolvida entre a faixa etária: para meninas 14 e 16 anos e para meninos, 17 e 19 anos (Garcia, 1840).

Após o período de mudança, a voz é categorizada como voz de mulher e voz de homem; atingindo a fase de maturidade. Desse modo, Garcia (1840) esclarece que registro de peito, falsete e registro de cabeça são possíveis em ambos os gêneros; o que significa que não há uma voz específica feminina e/ou masculina (Garcia, 1840). Isso é demonstrado, a seguir, por meio da Tabela V.

| Registro | Características | Mulher | Homem |
|----------|--|---|---|
| Peito | Base essencial de qualquer voz, porém mais presente na voz do homem do que na voz da mulher. - Voz forte, redonda, clara e geralmente abrange a escala dezesseis. | Soa paralelo ao falsete; mas em notas graves, permite ter maior extensão vocal. | Ao unir a mesma escala de extremidades alcançadas por baixo e tenor, esse registro abrange três oitavas entre os sons. |
| Falsete | O falsete é mais comum entre mulheres e crianças. Considerase um registro fraco e assemelha-se ao som grave da flauta. | Por assemelhar-se a voz de peito, permite articular os sons, sem distinção, tanto em voz de peito, como em falsete. | São difíceis de articular, porém produzem os mesmos sons na voz de peito. |
| Cabeça | Voz principal da mulher e é mais presente na voz da mulher do que na voz do homem. | Voz principal de mulheres e crianças quando alcançam a extensão de gritos agudos. | Os homens perdem a voz de cabeça quando meninos por meio da <i>la mue</i> ; no entanto, alguns indivíduos retêm o primeiro terço maior. |

Tabela V. Síntese dos registros de vozes de mulheres e homens

Após observar a Tabela V, compreende-se há vozes comuns identificadas entre mulheres e homens. Desse modo, Garcia (1840, p. 12, tradução livre) afirma que “a voz de peito no homem excede pelos sons graves a da mulher. A voz em falsete é comum entre eles. A voz principal nas mulheres supera a dos homens” (5).

Dito isso, compreende-se que: 1) a emissão da voz de peito soa mais grave no homem, do que na mulher; 2) o falsete, apesar de comum entre o

homem e a mulher, é mais difícil de ser articulado na voz do homem; 3) a voz de cabeça é considerada mais comum na mulher (e por isso é chamada de voz principal) e menos presente na voz do homem, devido as mudanças de voz estabelecidas por meio da *la mue*.

Ao descrever sobre a voz de peito do homem, Garcia (1840) faz referência as extensões vocais baixo e tenor. Isso se explica pelo autor considerar a existência de *voix cultivées* (vozes cultivadas) por meio dos sons vocais anteriormente descritos. Desse modo, Garcia (1840) classifica os registros de voz em sete categorias, sendo três vozes de mulheres (contralto, meio-soprano e soprano) e quatro vozes de homens (baixo, barítono, tenor e contraltino); observados na Tabela VI, a seguir.

| | |
|---------------|---|
| Voz de mulher | Contralto (<i>contralto</i>) |
| | Registre de poitrine (registro de peito) |
| | Fausset-tête (registro de falsete-cabeça) |
| | Mezzo-soprano (meio-soprano) |
| | Soprano (soprano) |
| Voz de homem | Basse-taille (basso; baixo) |
| | Baryton (baritono; barítono) |
| | Tenore (tenor) |
| | Haute-contre (contraltino) |

Tabela VI. Classificação das vozes cultivadas

Diferente da maior parte dos autores italianos anteriores, Garcia (1840) enaltece a voz feminina definindo-a como mais bela e flexível. Apesar de, ao descrever vozes cultivadas, categorizar voz de mulher e voz de homem, em nenhum momento afirma que contralto, meio-soprano e soprano são vozes utilizadas apenas pela mulher e baixo, barítono, tenor e contraltino, utilizadas apenas pelo homem. Um exemplo dessa percepção é por meio da definição barítono; Garcia (1840) afirma que esse tipo de voz tem a capacidade de formar o mesmo tipo de som que o falsete e possui a mesma extensão vocal feminina e masculina. Desse modo, assim como nas definições de registro de peito, falsete e registro de cabeça, Garcia (1840) explicita que as vozes mais agudas, são mais presentes na voz da mulher e as mais graves, na voz do homem.

6. Terceira parte: classificação de vozes cultivadas versus classificação de voz de Dewey

A partir das considerações anteriores sobre vozes cultivadas, demonstra-se na Tabela VII, a seguir, a classificação de voz apresentada pela 23ª

edição da CDD; a fim de compreender sua estrutura por meio da teoria da música discutida até o momento no presente trabalho.

| Subdivisão Padrão | Descrição da subdivisão padrão | Extensões vocais da subdivisão padrão |
|--------------------------|--|---|
| .6 Vozes de Mulheres | Destinada igualmente a vozes femininas ou infantis | Soprano Meio-soprano Contralto |
| .7 Vozes de Crianças | Destinada igualmente a vozes femininas ou infantis | Soprano Meio-soprano Contralto Changing |
| .8 Vozes de Homens | Vozes de homens | Contratenor Falsete Castrato Tenor Barítono Baixo |
| .9 Outros tipos de vozes | Outros tipos de vozes | Vozes falando (discurso de coral) Sprechgesang (canto falado) Assobio |

Tabela VII. Tipos de vozes na 23ª da CDD (782.6 – 782.9)

Conforme mencionado anteriormente na introdução, a 20ª edição da CDD é considerada revolucionária na representação do assunto música por sua revisão basear-se no Catálogo Britânico e na Classificação *Hornbostel-Sachs*; o que significa que foram utilizadas garantias literárias da música para a construção de um sistema de classificação bibliográfico. As principais modificações que ocorreram na 20ª edição, de acordo com Novaes (2020), permaneceram na 23ª edição, havendo apenas algumas alterações terminológicas e não conceituais.

Desse modo, a partir do conhecimento adquirido pela escola de Garcia (1840) no presente trabalho, as vozes mulher, crianças e homem, não estão construídos hierarquicamente de acordo com a teoria da música. Apesar de vozes de crianças assemelharem-se à voz de mulher devido ao uso do falsete, nesse estágio de vida, as vozes de meninas e meninos são consideradas idênticas, se não muito parecidas; o que complica colocar a categoria vozes de crianças subordinada a vozes de mulheres, assim como descrevê-la como uma voz igual a feminina.

Outra questão importante, é observar entre as extensões vocais categorizadas em vozes de crianças, a inserção do termo *changing*, que faz referência ao *mouting* manifestado na fase de puberdade do indivíduo e não na infância; considerado uma técnica de amadurecimento de voz e não um tipo de extensão vocal. Tendo isso, o

ideal, ao representar vozes na perspectiva de faixa etária, seria classificar: vozes de crianças (infância) > *moulting* (puberdade) > vozes maduras (fase adulta).

Em se tratando de categorizar a voz por gênero (voz de mulher e voz de homem), não faz sentido perante a teoria da música (baseada na fisiologia da voz), pois conforme enfatizado por Garcia (1840), uma voz permite assumir inúmeras extensões vocais, independente do gênero. Tendo isso, ser consciente de que as vozes mais agudas, por questões fisiológicas, são mais presentes em mulheres e as mais graves em homens, não determina que há vozes exclusivas de ambos os gêneros.

Ademais, ao analisar as extensões vocais descritas em voz de homem, observa-se a inserção do termo *castrato*, que conforme mencionado no capítulo anterior, refere-se ao cantor soprano do período Belo Canto. Tendo isso, *castrato* não deveria estar estruturado como um tipo de extensão vocal e sim, ser inserido como um termo relacionado das vozes mais altas e agudas.

Em se tratando da subdivisão padrão outros tipos de vozes, a mesma faz referência a vozes falando (discurso em coral), *Sprechstimme* (técnica que intercala canto e fala, conhecida pela ópera alemã) e assobio. Tendo isso, compreende-se que os três termos mencionados para descrever outros tipos de vozes, expressam técnicas vocais e não extensões vocais como classificadas nas subdivisões anteriores (voz de mulher, vozes de crianças e voz de homem).

Desse modo, considerando a análise metateórica realizada, percebe-se que a classificação proposta por Garcia (1840) baseada na fisiologia da voz, traz luz para a compreensão da classificação de voz elaborada pela CDD; que apesar de estruturar de modo confuso as subdivisões padrão (obedecendo os critérios de gênero, faixa etária e técnicas vocais), ainda sim, apropriou-se da teoria da música para a sua sistematização.

7. Conclusão

Na análise metateórica realizada por meio da abordagem MU interna-intelectual de Ritzer (1990, 1991), os tratados italianos originados pelo movimento Belo Canto, foram responsáveis pelo surgimento das primeiras teorias sobre vozes humanas. Entre os oito autores mencionados, destaca-se Garcia (1840), professor e cantor não *castrato*, considerado revolucionário por basear seus estudos na fisiologia da voz, classificando não apenas os registros vocais, denominado pelo mesmo como espécies de sons vocais,

assim como as vozes cultivadas, que é a classificação fundamental para compreender os tipos de vozes classificados na 23ª edição da CDD.

Garcia (1840) ao classificar a voz humana, assume que a mesma parte de princípios primitivos e fundamentais. Desse modo, compreende-se que apesar do autor utilizar em suas definições voz de mulher e voz de homem, é ciente que essa nomenclatura não é apropriada pelas diversas extensões vocais que um indivíduo permite alcançar. Porém, ao mesmo tempo, reconhece que tais nomenclaturas são fundamentais na perspectiva do amadurecimento da voz considerada pela teoria da música; que retrata: o primeiro estágio como voz de criança, o segundo estágio como fase de mudança da voz (*la mue/moulting*), e o terceiro estágio como a verdadeira natureza da voz, onde a mulher e o homem assumem e desenvolvem as suas habilidades vocais.

Dito isso, é perceptível que os termos inseridos nas descrições das vozes na 23ª edição da CDD, partem do princípio da teoria da música, mas o modo como as subdivisões padrões são categorizadas, além de disseminarem ideias pré-concebidas, não se sustentam ao olhar da teoria da classificação.

O principal objetivo do presente estudo, foi demonstrar as teorias que fundamentam classificar vozes humanas, porém, com os resultados obtidos, questiona-se qual seria a classificação bibliográfica ideal para categorizar vozes sem demonstrar ideias pré-concebidas. Essa questão será elaborada em um trabalho futuro, mas possivelmente, um melhor modo para representar registros vocais, seria classificá-los por meio de suas extensões vocais ao invés de classificá-los por faixa etária ou gênero (mulher e homem).

Notas

- (1) “[...] MU is concerned, more specifically, with the study of theories, theorists, communities of theorists, as well as the larger intellectual and social contexts of theories and theorists.” (Ritzer, 1990, p. 4).
- (2) “[...] the goal of producing a perspective, one could say a metatheory, that overarches some part or all of sociological theory.” (Ritzer, 1990, p. 4).
- (3) “La voix humaine soumise à l'influence des âges, des sexes, des constitutions, subit des modifications innombrables. Indépendamment des différences tranche esqui distinguent les voix des divers individus, il y a aussi un nombre illimité de nuances appartenant à l'organe d'un même individu.” (Garcia, 1840, p. 10).
- (4) “Si l'on appauvrit l'organe vocal par l'exercice du chant, ou par des excès quelconques, on épuise la plante avant qu'elle soit apte à donner des fruits 5 on fait succéder la caducité à l'enfance: on tue la virilité. C'est seulement après la mue que doivent commencer les études sérieuses du chant.” (Garcia, 1840, p. 11).

- (5) "La voix de poitrine chez l'homme dépasse par les sons graves celle de la femme. La voix de fausset leur est commune. La voix de tête excède chez la femme celle de l'homme." (Garcia, 1840, p. 12).

Referências

- Abrahamsen, K. T (2003). Indexing of musical genres: an epistemological perspective. // *Knowledge Organization*. 30:3/4 (2003) 144-169.
- Agrícola, J. F (1757). *Anleitung zur Singkunst: aus dem italiinischen des Herrn Peter Franz Tosi*. Berlin: Georg Ludwig Winter, 1757.
- Antelmi, C. C. A.; Silveira, G (2008). Terminologia de recursos vocais sob o ponto de vista de fonoaudiólogos e preparadores vocais. *Anais do XVI Congresso Brasileiro de Fonoaudiologia: Campos do Jordão, 2008*. São Paulo: SBFa, 2008.
- Bates, M. J (1994). The design of databases and other information resources for humanities scholars: The Getty Online Searching Project report no. 4. // *Online and CD-ROM Review*. 18:6 (1994) 331-340.
- Bates, M. J (2005). An introduction to metatheories, theories, and models. // Fischer, K.; Erdelez, S.; Mckechnie, L. (eds). *Theories of information behavior*. Medford: Information Today, p. 1-24, 2005.
- Behlau, M. S (2001). *Voz: o livro do especialista*. Rio de Janeiro: Revinter, 2001.
- Bradley, C. J (1972). The Dickinson Classification for music: an introduction. // *Fontes Artis Musicae*. 19:1/2 (1972) 13-22.
- Bryant, E. T.; Marco, G. A (1985). *Music librarianship: a practical guide*. Metuchen: Scarecrow Press, 1985.
- Campbell, N. D (2009). Multiple paths to partial truths: a history of drug use etiology. // Scheier, L. M. (ed.). *Handbook of drug use etiology: theory, methods, and empirical findings*. Washington, DC: American Psychological Association, 2009. 29-50.
- Caccini, G (1602). *Le Nuove Musiche*. Florença: AR Editions, Inc., 1602.
- Dewey, M. (2011). *Dewey decimal classification and relative index*. Albany, New York: OCLC, 2011.
- Dolan, C (2007). Feasibility study: The evaluation and benchmarking of humanities research in Europe. // *Arts and Humanities Research Council (AHRC) (2007)*.
- Elliker, C (1994). Classification schemes for scores: analysis of structural levels. // *Notes*. 50:4 (1994) 1269-1320.
- Fujita, M. S. L (2008). Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. // *Tendências da Pesquisa Brasileira em Ciência da Informação*. 1 (2008) 1-32.
- Galliard, J. E (1742). *Observations on the florid song; or sentiments on the ancient and modern singers*. Londres: J. Wilcox, 1742.
- Garcia, M (1840). *Traite complet sur l' art du chant*. Paris: Minckoff, 1840.
- Gayotto, L. H (1997). *Voz partitura da ação*. São Paulo: Summus, 1997.
- Hjørland, B (2008). What is knowledge organization (KO)? // *Knowledge Organization*. 35:2/3 (2008) 86-101.
- Hjørland, B (2002). Epistemology and the socio-cognitive perspective in information science. // *Journal of the American Society for Information Science and Technology*. 53:4 (2002) 257-270.
- Hjørland, B (2017). Domain Analysis. // *Knowledge Organization*. 44:6 (2017) 436-464.
- Hjørland, B.; Albrechtsen, H (1995). Toward a new horizon in information science: Domain-analysis. // *Journal of the American society for information science*. 46:6 (1995) 400-425.
- Krohn, E. C (1970). On classifying sheet music. // *Notes*. 26 (1970) 473-478.
- Laver, J. D. M (1968). Voice quality and indexical information. *British Journal of Disorders of Communication*. 3:1 (1968) 43-54.
- Large, J (1972). Towards an integrated physiologic-acoustic theory of vocal registers. // *NATS Bull*. 28:3 (1972) 18-25.
- Lee, D (2017). *Modelling music: a theoretical approach to the classification of notated Western art music*. Londres: Universidade de Londres, 2017. Tese de Doutorado. <https://openaccess.city.ac.uk/id/eprint/17445> (2021-01-15).
- Mancini, G (1774). *Pensieri e riflessioni pratiche sopra il canto figurato*. Viena: Ghelen, 1774.
- Marsh, C (2002). The "ANSCR" to CD classification at Leeds College of Music // *Brio*. 39:1 (2002) 33-48.
- Moreira, W (2018). *Sistemas de organização do conhecimento: aspectos teóricos, conceituais e metodológicos*. Marília: Universidade Estadual Paulista (Unesp), 2018. Tese de Livre-docência. <http://hdl.handle.net/11449/190878> (2021-01-15).
- Nathan, I (1836). *Musurgia Vocalis*. London: Fentum, 1836.
- Novaes, F. C. P (2020). *Ontogenia do assunto música na CDD: uma análise da divisão 780*. Marília: Universidade Estadual Paulista (Unesp), 2020. Dissertação de Mestrado. <http://hdl.handle.net/11449/192277> (2021-01-15).
- Pacheco, A. J. V (2004). *Mudanças na prática vocal da escola italiana de canto: uma análise comparativa dos tratados de canto de Pier Tosi, Giambattista Mancini e Manuel P. R. Garcia*. Campinas: Universidade Estadual de Campinas, 2004. Dissertação de mestrado. <http://www.repositorio.unicamp.br/handle/REPOSIP/284287> (2021-01-15).
- Philp, G. J (1982). The proposed revision of 780 music and problems in the development of faceted classification for music. // *Brio*. 19:1 (1982) 1-13.
- Redfern, B (1978). *Organising music in libraries: arrangement and classification*. London: Clive Bingley, 1978.
- Ritzer, G (1990). Metatheorizing in Sociology. // *Sociological Forum*. 5:1 (1990) 3-15.
- Ritzer, G (1991). *Metatheorizing in Sociology*. Lexington: Lexington Books, 1991.
- Ritzer, G (2000). *Modern sociological theory*. Boston: McGraw-Hill, 2000.
- Ritzer, G (2001). *Explorations in social theory: from metatheorizing to rationalization*. London: Sage, 2001.
- Salomão, G. L (2008). *Registros vocais no canto: aspectos perceptivos, acústicos, aerodinâmicos e fisiológicos da voz modal e da voz de falsete*. São Paulo: Pontifícia Universidade Católica de São Paulo, 2008. Tese de Doutorado. <https://tede2.pucsp.br/handle/handle/14033> (2021-01-15).
- Sandstrom, A. R.; Sandstrom, P. E (1995). The use and misuse of anthropological methods in library and information science research. // *Library Quarterly*. 65:2 (1995) 161-199.
- Silva, M. A. A.; Dupret, A. C (2004). *Voz Cantada*. // Ferreira, L. P.; Befi-Lopes, D. M.; Limongi, S. C. *Tratado de fonoaudiologia*. São Paulo: Roca, 2004.
- Stark, J (2003). *Bel Canto: A History of Vocal Pedagogy*. Toronto: University of Toronto Press, 2003.
- Tosi, P. F (1723). *Opinione de' cantori antichi e moderni, o sieno osservazione sopra il canto figurato*. Bologna: 1723.

Wallace, W (1992). Metatheory, conceptual standardization and the future of sociology. // Ritzer, G. (Ed.). *Metatheorizing*. Newbury Park, CA: Sage, 1992. 53-68.

Wallis, S (2010). Toward a science of metatheory. // *Integral Review: A Transdisciplinary and Transcultural Journal for New Thought, Research, and Praxis*. 6:3 (2010).

Zacconi, L (1592) *Prattica di musica*. Venice: Bartolomeo Carampello, 1592.

Enviado: 2021-05-17. Segunda versión: 2021-07-19.
Aceptado: 2021-10-05.

A utilização de linguagens de indexação por bibliotecas universitárias: mapeamento sistemático da literatura

El uso de lenguajes de indización en las bibliotecas universitarias: mapeo sistemático de la literatura

The use of indexing languages by university libraries: systematic literature mapping

Jessica Beatriz TOLARE, Mariângela Spotti Lopes FUJITA

Universidade Estadual Paulista – Unesp, Campus de Marília, Dep. de Ciência da Informação, Av. Hygino Muzzi Filho, 737, Campus Universitário, 17525-900 – Marília, SP (Brasil). [jessica.tolare | mariangela.fujita] @unesp.br

Resumen

Se analizan los estudios sobre lenguajes de indización en las bibliotecas universitarias para verificar las similitudes y diferencias, los métodos, las propuestas, las contribuciones y las recomendaciones, a través del mapeo sistemático de la literatura en las bases BRAPCI, LISTA, WoS y Scopus entre 2002 y 2022. Las variables que pueden afectar el proceso de indización son: uso de vocabulario controlado e inapropiado de otras áreas del conocimiento; atribución de la indización por parte del autor; el profesional indizador; falta de política de indización y lenguajes y la consonancia entre ellos; falta de método en la construcción del lenguaje de indización; indisponibilidad del lenguaje de indización por parte de las bibliotecas para los usuarios; falta de un manual de indización; comprensión de los profesionales que realizan la mediación del lenguaje de indización y su complejidad. Cada estudio analizado presenta, al menos en algunos aspectos, similitudes y diferencias entre sí, en métodos, propuestas, contribuciones y recomendaciones. Utilizar solamente las directrices de la política de indización y el manual no es suficiente para resolver todos los problemas que involucran el proceso de indización, es necesario realizar una combinación de multifactores: el indizador, la calidad del lenguaje de indización en su construcción, el instrumento utilizado, los sistemas y el contexto de la biblioteca universitaria.

Palabras clave: Indización. Lenguajes de indización. Bibliotecas universitarias.

1. Introdução

Existem diferentes tipos de bibliotecas como públicas, especializadas, escolares, particulares, comunitárias e universitárias. Cada biblioteca possui sua própria especificidade e seu público-alvo, desenvolvendo sua identidade com produtos e serviços que irá oferecer.

Quando se trata de uma biblioteca universitária, seu objetivo consiste em atender à instituição ao qual ela está vinculada, especificamente, os alunos, professores, funcionários, pesquisadores e, às vezes, a comunidade local. Por isso, oferece

Abstract

The studies on indexing languages in university libraries are analyzed to verify similarities and differences, methods, propositions, contributions and recommendations, through the systematic mapping of literature in the databases BRAPCI, LISA, WoS and Scopus, in the period 2002-2022. The variables that can affect the indexing process are: use of controlled and inappropriate vocabulary from other areas of knowledge; attribution of indexing by the author; the professional indexer; lack of indexing policy and languages and the consonance between them; lack of method in the construction of the indexing language; unavailability of the indexing language by libraries to users; lack of an indexing manual; understanding of the professionals who perform the mediation of the indexing language and its complexity. These variables showed how the indexing process can be affected, involving aspects of the indexer, the system, the elaborated guidelines and how the indexing language construction occurs. We concluded that each study analyzed presented, at least in some aspects, similarities and differences between each other, methods, propositions, contributions and recommendations. Using only the indexing policy guidelines and the manual is not enough to solve all the problems involving the indexing process, it is necessary to perform a combination of multifactors, which are: the indexer, the quality of the indexing language in its construction, the instrument used, the systems and the context of the university library.

Keywords: Indexing languages. Indexing. Academic libraries.

um tipo de serviço de cunho específico e, conseqüentemente, precisa possuir uma organização adequada e própria para tornar a informação acessível, tanto em seu acervo online quanto físico.

A biblioteca universitária tem necessidade de sistematizar processos e condutas de indexação e, nessa tarefa, terá que definir uma política de indexação cujo desenvolvimento depende de planejamento, elaboração e implantação dos procedimentos, técnicas, normas e manual com o obje-

tivo de organizar e recuperar a informação inserida nesse ambiente (Fujita, Favato e Zaniboni & Fagundes, 2016, p. 23). Nesse sentido, a indexação em uma biblioteca universitária possui particularidades em seu processo envolvendo aspectos como acervo especializado e público-alvo específico. O processo de indexação nesse ambiente requer que os termos da linguagem de indexação atribuídos, estejam de acordo com a temática no qual o acervo abarca, pois caso isso não ocorra, pode causar discrepâncias em sua recuperação e prejudicar o usuário. Logo, esse processo se difere de outros tipos de bibliotecas, como públicas e comunitárias, pois, devido a sua especificidade, é necessário envolver especialistas para realizar esse procedimento.

O uso inadequado de uma política de indexação, para Fujita e Gil Leiva (2010), reflete diretamente no uso de linguagens de indexação pelo bibliotecário, da mesma forma que as linguagens de indexação exercem influência na política de indexação tendo em vista os resultados da recuperação.

Ao longo dos anos, pesquisas foram desenvolvidas sobre o uso de linguagens de indexação em bibliotecas, especialmente, em universitárias, caracterizando a problemática desta pesquisa: como a literatura publicada aborda as linguagens de indexação em bibliotecas universitárias?

Para isso, esta investigação propõe analisar estudos sobre linguagens de indexação em bibliotecas universitárias mediante mapeamento sistemático de literatura para realizar pesquisa bibliográfica em base de dados com o objetivo de analisar semelhanças e diferenças, métodos, proposições, contribuições e recomendações.

Espera-se que a análise da literatura publicada sobre o uso de linguagens de indexação em bibliotecas universitárias possa constatar como a variável linguagem de indexação influencia nesse processo de indexação, pois são determinantes para a representação no processo de indexação e podem interferir diretamente na recuperação de recursos informacionais dentro desse ambiente (Arboit, 2017).

2. Linguagem de indexação

A indexação é definida pela norma ISO 25.964-1 (International Standardization for Organization, 2011, p. 5) como uma “[...] análise intelectual do assunto de um documento, com o objetivo de identificar os conceitos representados, a fim de possibilitar a recuperação da informação.”

Neste estudo, a indexação é vista como um processo composto por fases e etapas para sua execução. Cada corrente teórica segue uma quantidade de fases e mesmo havendo divergência,

suas fases podem ser resumidas em: *exame do documento e estabelecimento do assunto de seu conteúdo*, etapa do exame do documento por meio da leitura documentária, para conhecer o assunto (Fujita, 2017); *identificação e seleção dos conceitos presentes no assunto*, fase que possui o intuito de extrair os conceitos que traduzam a essência do documento (Dias & Naves, 2007); e *tradução dos conceitos*, em que os conceitos são traduzidos por termos definidos pela linguagem de indexação.

A linguagem de indexação é definida por Silva e Brito (2018, p. 96) como um “conjunto de termos padronizados, que representam um conceito da linguagem natural, com vistas a uma padronização dos sistemas de indexação”.

A norma internacional *American National Standards Institute/National Information Standards* (ANSI/NISO, 2010) define linguagem de indexação como

Um 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 [objeto de conteúdo] para um armazenamento de informações e de um sistema de recuperação.

A linguagem de indexação é considerada um componente-chave para os sistemas de informação, sobretudo quando a informação é profissional ou acadêmica, devido a ser um contexto associado com informações complexas e de alta qualidade. Por isso, ela precisa ser extensa, para conseguir abordar adequadamente qualquer tipo de conhecimento, principalmente, para as ciências humanas, que pode possuir uma variedade de perspectivas diferentes (Vállez, Pedraza-Jiménez, Condia, Blanco & Rovira, 2015; Golub Tyrkkö, Hansson & Ahlström, 2020).

O uso da linguagem de indexação, dentro de um sistema de informação, ocorre em dois momentos: o primeiro consiste na entrada, momento em que se insere no sistema, no qual a representação da informação é realizada por intermédio do controle definido pela linguagem de indexação; o segundo momento corresponde à saída, quando o usuário realiza a busca pela informação, expressando sua necessidade e utilizando termos do vocabulário controlado (Cintra, Tálamo, Lara & Kobashi, 2002).

Para Gil Urdiciain (2004) y Cruz, Santos e Fujita (2016), a linguagem de indexação é construída por um conjunto de regras preestabelecidas em nível sintático, semântico e pragmático, baseando-se em teorias e metodologias oriundas da terminologia, da linguística textual, da semântica

estrutural e da sintaxe, tendo como princípio o de realizar o controle terminológico e aperfeiçoar o acesso à informação.

Moura, Silva e Amorim (2002), consideram que existe uma complexidade no desenvolvimento de linguagens de indexação, pois ela requer uma compreensão de variáveis que podem influenciar no processo. Essa complexidade se deve ao fato de que é tomado como princípio a garantia literária, a qual ocorre quando um vocabulário controlado é capaz de representar conceitos que realmente se encontram na literatura de assunto. Por isso, esse princípio tende a inibir algumas interfaces da busca entre os usuários e os sistemas de informação, como, por exemplo, o usuário iniciante, que não possui domínio geral dos assuntos e das categorias em que eles estão; e os profissionais, pois as categorias de assunto não foram inseridas no documento.

3. Procedimentos metodológicos

A pesquisa se caracteriza como exploratória e descritiva, de cunho qualitativo. O método utilizado para o seu desenvolvimento foi o mapeamento sistemático da literatura, pois possibilitou investigar os problemas relatados em pesquisas anteriores sobre uma determinada temática.

O mapeamento sistemático garante uma confiabilidade no trabalho por conter etapas bem definidas e tratar-se de um estudo menos aprofundado, mas que ocorre de forma mais abrangente. Esse método é composto por três etapas: planejamento da revisão para identificação da real necessidade da pesquisa; definição do protocolo da revisão; e escrita dos resultados. O protocolo especifica questões da pesquisa e estratégias utilizadas (Kitchenham & Charters, 2007; Ramires, 2022). Os protocolos são compostos por diferentes passos, e, para observar o seu uso, adaptou-se a tabela de Ramires (2022).

De todos os trabalhos apresentados nos resultados (152) nas buscas das bases de dados nacionais e internacionais, foram selecionados 45 estudos para o desenvolvimento da pesquisa. Os critérios estabelecidos para se chegar aos resultados dessa seleção consistiram na premissa de que os estudos abordassem a linguagem de indexação em bibliotecas universitárias, especificamente, apresentando as temáticas de pesquisa em pelo menos um dos locais do texto: no título, nas palavras-chave (tabela 2) ou o seu desenvolvimento nos resumos dos trabalhos selecionados.

Depois de selecionado os 45 estudos, foram realizadas leituras parciais do título, resumo, pala-

avras-chave, da introdução, resultados e considerações finais, a fim de observar se os autores realmente abordaram e desenvolveram o uso de linguagem de indexação em bibliotecas universitárias.

| <i>Passos do protocolo</i> | |
|---|---|
| <i>Passo 1: Mapeamento</i> | |
| <i>Objetivo</i> | Analisar artigos sobre linguagens de indexação em bibliotecas universitárias, com objetivo de comparar semelhanças e diferenças, métodos, proposições, contribuições e recomendações. |
| <i>Pergunta norteadora</i> | Como a literatura publicada aborda as linguagens de indexação em bibliotecas universitárias? |
| <i>População</i> | Artigos científicos |
| <i>Intervenção</i> | Publicações científicas em português e inglês |
| <i>Controle</i> | Publicações científicas em texto completo |
| <i>Resultados</i> | Seleção e análise dos dados das publicações científicas recuperadas para identificação de pesquisas, autores, datas das pesquisas, problemas, procedimentos e desenvolvimentos |
| <i>String de busca</i> | Linguagem de indexação; linguagem de indexação AND biblioteca universitária; linguagem de indexação AND biblioteca acadêmica; indexing language; indexing language AND University language; indexing language AND academic library |
| <i>Passo 2: Definição de critérios de seleção de fontes</i> | |
| <i>Critério de busca</i> | Publicações em forma de artigos científicos em texto completo |
| <i>Idioma</i> | Português e inglês |
| <i>Métodos de seleção</i> | Delimitação do espaço temporal no levantamento bibliográfico (2002-2022) Busca bibliográfica em bases de dados do assunto do primeiro item; Análise e seleção de textos da pesquisa bibliográfica com base no assunto determinado; Leitura parcial da introdução, resultado, considerações finais e resumo que abordaram a temática delimitada para seleção final da análise; Leitura aprofundada dos trabalhos selecionados; Fichamentos dos textos selecionados para realização da análise |
| <i>Passo 3: Definição de fontes</i> | |
| <i>Bases de dados</i> | BRAPCI, LISTA, WoS e Scopus. |
| <i>Passo 4: Critérios de seleção de documentos</i> | |
| <i>Critérios de inclusão</i> | Publicações científicas em forma de artigos disponíveis em texto completo que contemplam as strings de buscas citados. A inclusão de textos terá como critério a leitura do título, resumo, palavras-chave, resumo, introdução, resultados e considerações finais. |
| <i>Critérios de exclusão</i> | Ser artigos de periódicos fora do espaço temporal delimitado; ser de outro idioma do selecionado; não ter nenhuma relação com a temática proposta e não ser a tipologia documental definida. |
| <i>Tipos de documentos</i> | Artigos de periódicos |

Tabela 1. Passos do protocolo do mapeamento sistemático

Após essa leitura parcial, dos 45 estudos, foram selecionados 8, nos quais foram feitas leituras

aprofundadas: (Menard, 2010; Tartarotti, 2014; Cruz, Fujita & Santos, 2017; Sunny & Angadi, 2017; Fujita, Moreira, Santos e Cruz & Ribas, 2018; Cruz, 2019; Fujita, Cruz, Patrício & Branco, 2019; Golub, Tyrkkö, Hansson & Ahlström, 2020).

Os critérios de exclusão se deram ao perceber que as bases de dados apresentaram trabalhos com temáticas que eram diferentes das pesquisadas; fora do espaço temporal delimitado; não estarem no idioma ou na tipologia definida. Houve trabalhos em que foi necessário excluir da seleção, pois tinham como objetivo desenvolver as seguintes temáticas: “elementos da linguística em indexação”, “aplicabilidade do tesouro em biblioteca digital”, “pensamento dialógico da indexação social”, “teoria do conceito”, “indexação jurídica”, “indexação automática”, “ontologias”, “tradução intersemiótica dos mecanismos de buscas” entre outros assuntos diferentes. Nesses estudos, a linguagem de indexação foi apenas citada brevemente, por isso, a exclusão ocorreu por não terem relação de fato com os objetivos propostos.

4. Análise dos resultados

As buscas em bases de dados possibilitaram encontrar em seus resultados trabalhos que abordaram linguagem de indexação em bibliotecas universitárias. Por isso, a análise de resultados foi subdividida em: (I) busca e recuperação de estudos em bases de dados e (II) análise de estudos sobre linguagem de indexação em bibliotecas universitárias.

4.1. Busca e recuperação de estudos em bases de dados

No Gráfico 1 estão os resultados da busca feita a partir do termo composto “linguagem de indexação” em bases de dados nacionais e internacionais.

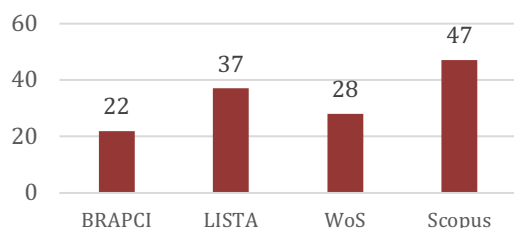


Gráfico 1. Resultados a partir dos termos “linguagem de indexação” em bases de dados nacionais e internacionais

O Gráfico apresenta a quantidade de trabalhos encontrados nos resultados em bases de dados nacionais e internacionais, a partir da busca pelo

termo composto “Linguagem de indexação”. Pode-se observar que a base com maior número de trabalhos foi a Scopus, seguido pela LISTA, WoS e BRAPCI.

A Scopus apresentou 47 pesquisas, no qual 2 abordavam o tema do estudo. Na LISTA foram encontrados 37 trabalhos, no qual apenas 1 possui a temática. A WoS teve 28 estudos em seus resultados, no qual 2 apresentaram relevância em seu assunto. Por último, a BRAPCI apresentou 22 trabalhos, no qual 5 foram considerados relevantes. A partir desses dados, pode-se analisar que houve uma quantidade muito pequena de trabalhos que de fato abordaram o uso de linguagens de indexação em bibliotecas universitárias, principalmente, quando comparado com a quantidade de resultados apresentados.

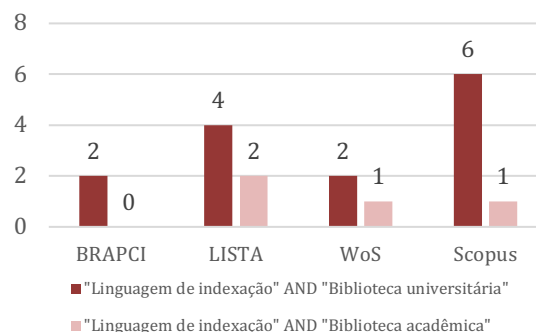


Gráfico 2. Resultados a partir dos termos “linguagem de indexação” AND “biblioteca universitária” em bases de dados nacionais e internacionais

No Gráfico 2 são apresentados os resultados da busca combinando dois termos compostos: “linguagem de indexação” AND “biblioteca universitária” e “linguagem de indexação” AND “biblioteca acadêmica”. A base de dados com a maior quantidade de estudos apresentados em seus resultados foi a Scopus, com 6, sendo 5 relevantes; seguido pela LISTA, com 4 pesquisas, no qual apenas 1 aborda a temática. A WoS e a BRAPCI apresentaram 2 trabalhos em cada, em que, na primeira base foi encontrado apenas 1 estudo e na segunda base os dois trabalhos foram considerados relevantes. Para um trabalho ser considerado relevante, eles tinham que abordar o uso de linguagem de indexação em bibliotecas universitárias. Pode-se observar nele uma quantidade muito pequena de estudos encontrados, no qual sua variação ocorreu entre nenhum trabalho, na base da BRAPCI, até 2 estudos, na base da LISTA. As bases WoS e Scopus apresentaram um trabalho em cada.

Por fim, cada base de dados apresentou uma quantidade diferente de trabalhos em seus resul-

tados; mesmo assim, notou-se que houve trabalhos que se repetiram em mais de uma base de dados. A base de dados que mais apresentou resultados foi Scopus, com 54 pesquisas; seguido pela base LISTA com 43 estudos e pela WoS com 31 trabalhos. A base de dados da BRAPCI ficou em último lugar com 24 pesquisas. Observou-se que o uso de estratégias de buscas como refinamento, operadores booleanos e aspas otimizaram as buscas e deixaram os resultados mais selecionados e consistentes com a temática pesquisada.

Ao analisar as bases de dados, pode-se perceber falhas ao recuperar os trabalhos, podendo gerar questionamento acerca da qualidade da indexação dos documentos, devido aos resultados apresentados após às buscas. As informações disponíveis sobre o processo de indexação nas bases de dados mostraram que os documentos são implementados por meio de indexadores e pelos próprios autores, não deixando claro se há uso da indexação automática ou vocabulário controlado na atribuição de termos. Nesse sentido, utilizar estratégias na hora de realizar a busca por publicações em bases de dados é algo imprescindível e uma forma de otimizar a recuperação nos resultados. Na análise foi possível observar como os resultados se alteraram com o uso de diferentes tipos de estratégias e como uma pesquisa pode ser afetada em sua qualidade dependendo dos tipos de buscas realizadas.

Essas falhas podem ocorrer por diferentes variáveis como, problemas na indexação dos sistemas, erros na atribuição de termos representativos pelos próprios autores ou pelo indexador e falta de conhecimento sobre o uso da linguagem de indexação para padronizar o processo.

4.2. Análise de estudos sobre linguagem de indexação em bibliotecas universitárias

Os 8 trabalhos selecionados para compor a análise estão relacionados na Tabela 1. A temática abordada neles coincide e se relaciona com o tema desenvolvido nesta pesquisa. Essa relação ocorre pelos trabalhos abordarem o uso de linguagem de indexação em bibliotecas universitárias, através da representação e recuperação da informação em repositórios, uso de vocabulários controlados e avaliação da indexação e do profissional indexador. Por outro lado, analisar uma quantidade menor de trabalhos possibilitou realizar análises mais aprofundadas em estudos mais pertinentes ao tema de investigação.

A análise terá início com a observação das palavras-chave atribuídas aos trabalhos selecionados e, na sequência, cada trabalho será individu-

almente apresentado a partir do objetivo, metodologia, resultados e principais conclusões. Ao final, a análise conjunta e comparada dos resultados dos trabalhos.

A tabela II apresenta as palavras-chave dos trabalhos que abordaram a linguagem de indexação. A necessidade em analisá-las ocorreu devido a importância em observar como as diferentes pesquisas atribuíram as palavras-chave para representar seus próprios estudos. Essa análise é pertinente e essencial para o trabalho, pois mostra uma perspectiva que pode interferir na representação e, conseqüentemente, na recuperação, bem como na seleção dos trabalhos para a análise. Pode-se observar uma grande diversidade de palavras-chave, que variam entre assuntos mais gerais até os mais específicos.

| Autor(es) | Palavras-chave |
|--|--|
| Menard (2010) | Não apresenta palavras-chave no texto |
| Tartarotti (2014) | "Tratamento Temático da Informação"; "Bibliotecas universitárias"; "Protocolo Verbal"; "Avaliação da Indexação"; "Ciência, Tecnologia e Sociedade" |
| Cruz, Fujita e Santos (2017) | "Linguagem de indexação"; "Política de indexação"; "Indexação"; "Bibliotecas universitárias" |
| Sunny e Angadi (2017) | "Systematic literature review"; "Evaluation studies"; "Thesaurus"; "Digital information retrieval systems"; "Information retrieval effectiveness"; "Usability studies" |
| Fujita, Moreira, Santos, Cruz e Ribas (2018) | "Indexing languages"; "Hierarchical"; "Engineering"; "Subject heading lists" |
| Cruz (2019) | "Linguagem de indexação"; "Controle de vocabulário"; "Bibliotecas universitárias" |
| Fujita, Cruz, Patrício e Branco (2019) | "Organização do conhecimento"; "Indexação"; "Linguagens de indexação"; "Política de indexação"; "Bibliotecas universitárias" |
| Golub, Tyrkkö, Hansson, Ahlström (2020) | "Digital libraries"; "Digital humanities"; "Institutional repositories"; "Humanities"; "Knowledge organization"; "Bibliographic databases"; "Subject indexing" |

Tabela II. Palavras-chave dos 8 trabalhos selecionados

Observa-se que em quatro estudos as palavras-chave com maior frequência foram: "bibliotecas universitárias" e "linguagens de indexação" / "indexing languages".

Houve trabalhos que apresentaram palavras-chave como, humanidades, engenharia, estudo de usabilidade, estudos de avaliação entre outros. Esses termos, por si só, não são adequados para representar os trabalhos e os que poderiam representar em algum nível não possuem relação direta com a temática abordada.

Percebeu-se, no geral, trabalhos que atribuíram palavras-chave adequadamente, mas também

trabalhos que utilizaram palavras-chave sem relação com a linguagem de indexação. Isso afeta a recuperação, pois mesmo quando o estudo é recuperado, a falta de representação pelas palavras-chave pode afetar a leitura do usuário, fazendo-o questionar se o texto será utilizado. Nessa condição, o texto, poderá ser descartado antes mesmo da leitura, pois, com o tempo limitado e a grande quantidade de texto para ser lido, o usuário apenas lê as informações principais como título, resumo e palavras-chave.

Os trabalhos selecionados serão analisados e discutidos pela seguinte ordem cronológica abaixo: Ménard (2010), Tartarotti (2014), Cruz, Fujita e Santos (2017), Sunny e Angadi (2017), Fujita, Moreira, Santos, Cruz e Ribas. (2018), Fujita, Cruz, Patrício e Branco (2019), Cruz (2019), Golub, Tyrkkö, Hansson, Ahlström (2020). Nesses trabalhos, foram analisados como os autores desenvolveram o uso da linguagem de indexação voltado para a biblioteca universitária, observando os resultados e conclusões apontados em seus estudos acerca dessa temática.

Ménard (2010) abordou em seu estudo a comparação entre dois vocabulários de indexação na recuperação de imagens. A autora concluiu que a análise dos termos de indexação demonstrou que ambas as abordagens de indexação, com o vocabulário controlado e não controlado, são inerentemente diferentes uma da outra. No entanto, em termos de linguagem de indexação, a análise dos termos confirma que as características das abordagens são bastantes semelhantes, dentro da mesma indexação, independentemente da linguagem adotada. Ainda de acordo com a autora, se existirem diferenças, elas são manifestadas mais na indexação utilizando vocabulário não controlado.

Dependendo das circunstâncias, as duas abordagens podem coexistir, ressaltando-se que a linguagem não controlada não deve necessariamente ser considerada como uma alternativa ou uma solução para substituir a indexação que emprega a linguagem controlada. Os resultados obtidos sugeriram que os sistemas de informação podem, em um futuro próximo, se beneficiar do uso combinado do vocabulário controlado e não controlado, recorrendo, por exemplo, à indexação colaborativa por parte dos usuários e por parte de pesquisadores no assunto.

Na investigação realizada por Tartarotti (2014) em sua dissertação acerca do profissional no tratamento temático da informação (TTI), foi observado que o catálogo com as bases de dados de áreas científicas especializadas permite a visibilidade informacional das bibliotecas universitárias, ressaltando que a determinação de assunto,

nesses sistemas, envolve um profissional indexador que, geralmente, não é especialista na área científica especializada em que atua. Por isso, é responsabilidade dele determinar de forma precisa o conteúdo do documento. A autora ressalta que esse processo pressupõe um nível elevado de complexidade, porque existem problemas relacionados à terminologia do documento e da própria área e a influência das variáveis leitor-indexador, texto e contexto do processo de leitura documentária dos profissionais.

No trabalho desenvolvido por Cruz, Fujita e Santos (2017, p. 222) sobre linguagem de indexação em bibliotecas universitárias, chegou-se à conclusão de que muitas instituições oferecem uma variedade de cursos e, com isso, a cobertura de assuntos da biblioteca é abrangente. Por isso, foi recomendado pelas autoras no investimento da construção de uma linguagem de indexação, já que a “[...] linguagem em consonância com a política de indexação trabalharia em função do sistema em questão ao buscar uma padronização mais consistente [...]”. Mesmo assim, as autoras constataram a necessidade em realizar a manutenção periódica da linguagem, devido à constante evolução das ciências, e para que a representação de assuntos possa ser adequada e específica aos objetivos e à realidade da biblioteca, desde que observada a capacidade do software de gestão da linguagem em inserir novos termos ou a retirada dos que não serão mais utilizados.

Sunny e Angadi (2017) desenvolveram uma pesquisa cuja observação ocorreu em estudos de avaliação publicados na língua inglesa e indexados nas bases de dados da LISA e Scopus, sobre a eficácia de tesouros na recuperação da informação em sistemas de informação digital. De acordo com os autores, um tesouro pode melhorar a eficácia da recuperação, se os documentos forem pré-indexados e se a busca for processada com os termos do tesouro. O estudo percebeu que a eficácia da recuperação também pode ser melhorada com a aplicação do tesouro por meio de outras técnicas, como modelo vetorial ou medida de similaridade de Jaccard para classificar os resultados.

Fujita, Moreira, Santos, Cruz e Ribas (2018) relataram o trabalho de uma equipe de catalogadores e pesquisadores engajados em construir uma estrutura hierárquica de uma linguagem de indexação para um catálogo online nas áreas do conhecimento de uma rede de bibliotecas universitárias. Para o desenvolvimento da linguagem de indexação, foram utilizadas estruturas hierárquicas, definindo categorias e subcategorias que formaram uma macroestrutura, usando parâmetros da Biblioteca do Congresso e Terminologia

da Biblioteca Nacional e do Vocabulário do Sistemas de Bibliotecas da Universidade de São Paulo. Os resultados possibilitaram a avaliação das estruturas hierárquicas dos idiomas utilizados na organização de termos superordenados e subordinados, o que contribuiu para a sistematização dos procedimentos operacionais contidos em um manual de linguagem de indexação para catálogos online de bibliotecas.

Fujita, Cruz, Patrício e Branco (2019) realizaram estudo analítico sobre linguagens de indexação em bibliotecas universitárias por meio da aplicação de questionários. Os resultados obtidos revelaram novos cenários, com avanços, problemas e lacunas no uso da linguagem pela biblioteca universitária. Um dos desafios está relacionado ao problema de falta de compreensão pelos profissionais que efetuam a mediação da linguagem de indexação na representação e recuperação da informação. Desse modo, foram realizadas buscas na interface de 24 bibliotecas, na tentativa de verificar a especificidade na representação, porém, não obtiveram o alcance da representatividade na estratégia de busca pelo usuário na recuperação, mostrando que nenhuma das bibliotecas analisadas, nesse aspecto, disponibilizam algum tipo de linguagem para consulta dos usuários. Foi levantada a hipótese da complexidade que seria, para o usuário, consultar mais de uma linguagem para elaborar a sua estratégia de busca.

Ainda no trabalho desenvolvido por Fujita, Cruz, Patrício e Branco (2019, p. 220), notou-se que houve também um avanço considerável em 20 bibliotecas (um terço do total de 60 bibliotecas da amostra), as quais desenvolveram linguagem própria, enquanto 10 bibliotecas declararam pretensão em construir. As autoras consideraram esse dado relevante, pois demonstra uma “[...] tendência da gestão do tratamento temático com política de indexação que define ações equilibradas entre a representação e a recuperação da informação [...]”. Já a lacuna evidenciada pelas autoras está relacionada a esse avanço, porque diz respeito ao desconhecimento de softwares de gestão e manutenção de linguagens de indexação e à possibilidade de construção compartilhada de linguagens de indexação pela interoperabilidade. O único software relatado foi o Tema-Tres por apenas uma biblioteca, o qual dispõe de dispositivos que facilitam o seu uso articulado aos softwares de gestão de bibliotecas.

A pesquisa desenvolvida por Cruz (2019), teve como proposta investigar a aplicação da linguagem de indexação em bibliotecas universitárias brasileiras, através da perspectiva do uso de ferramentas de controle do vocabulário. Como método a autora aplicou questionários, através do

Google Form, em bibliotecas universitárias brasileiras e comparou os resultados com os do estudo de Fujita, Cruz, Patrício e Branco (2019), no qual os autores fizeram um estudo com temática semelhante e abrangente. A autora verificou que todas as bibliotecas universitárias possuem profissionais dedicados à tarefa de realizar a indexação e que 84,8% utilizam a linguagem de indexação no TTI, mas, quando não encontram o termo na linguagem empregada, as bibliotecas usam a linguagem natural, podendo acarretar inconsistências no catálogo da biblioteca e disparidades na recuperação da informação. Ela constatou que metade das bibliotecas disponibilizam a linguagem de indexação no catálogo aos usuários e possui uma política de indexação documentada (Cruz, 2019).

Golub Tyrkkö, Hansson e Ahlström (2020, p. 1210, tradução nossa) procederam a um estudo sobre a indexação de assunto em humanidades, fazendo a comparação entre o repositório de uma universidade pública na Suécia e um serviço bibliográfico internacional. O estudo apontou que os objetivos bibliográficos fixados para garantir o acesso do assunto para artigos de periódicos em humanidades não são adequadamente suportados no repositório da universidade e na Scopus, considerada pelos autores como o maior banco de dados de resumos e citação comercial do mundo. Ainda de acordo com os autores, as políticas de indexação nos dois serviços parecem não atender às necessidades dos acadêmicos de humanidades para termos de indexação de assunto. Foi constatado que nenhum vocabulário controlado é utilizado no repositório da universidade pública da Suécia, em virtude, segundo os autores, do fato de que as categorias do repositório se destinam, principalmente, à análise estatística e não à recuperação, enquanto na Scopus, há o impedimento da recuperação eficaz por meio da omissão de seus vocabulários controlados. Os resultados também indicaram que existe falta de mapeamento entre vocabulários que geram duplicações e impedem os usuários de usar termos de um vocabulário em todos os recursos.

O trabalho desenvolvido por Golub Tyrkkö, Hansson e Ahlström (2020, tradução nossa) concluiu que as palavras-chave incluídas pelo autor, no momento de depositar uma dissertação ou tese no repositório, são complementares às palavras-chave do vocabulário controlado, em virtude de os autores não serem treinados para esse processo de indexação e nem receber diretrizes de como implementar o processo. Para sanar esse problema é sugerido o fornecimento de treinamento e orientações para os autores, com o intuito de aprimorar e agilizar o processo.

Em uma síntese dos estudos analisados pode-se observar uma variedade de abordagens do uso de linguagens de indexação por bibliotecas universitárias. No estudo de Ménard (2010), foi abordada a necessidade de coexistência entre a linguagem de indexação e a linguagem natural (ou não controlada). A justificativa ocorreu ao apontar que ela poderá ser utilizada no futuro como uma indexação colaborativa, ocorrendo por partes dos usuários e pesquisadores no assunto. A indexação colaborativa abordada para aproximar o usuário da indexação e possibilitar que a representação da informação receba diferentes perspectivas em seu processo na perspectiva de contribuir substancialmente para a indexação. Ambas as perspectivas poderão ser contempladas, entretanto, esse processo possui muita complexidade na representação dos conteúdos de recursos informacionais o que poderá gerar discrepâncias entre o usuário e o indexador.

Na dissertação de Tartarotti (2014), percebeu-se a existência de circunstâncias e variáveis de leitura documentária que influenciam no processo de indexação e no uso de linguagem de indexação. Essas variáveis foram comentadas por Fujita, Cruz, Patrício e Branco (2019) e Golub, Tyrkkö, Hansson e Ahlström (2020), no relato de problemas, lacunas e avanços da linguagem de indexação no âmbito da biblioteca universitária, bem como a falta do uso de linguagem de indexação ou o uso inadequado para controlar o vocabulário. Para Tartarotti (2014), o indexador pode influenciar no processo de indexação. A autora apontou que existem aspectos do profissional que podem interferir nesse processo como suas experiências, pré-conceitos e preconceitos, formação acadêmica e conhecimento em diferentes níveis do mundo. Enquanto, para Fujita, Cruz, Patrício e Branco (2019), também existem outros aspectos que podem interferir no processo de indexação como os próprios sistemas e o ambiente em que se inserem; a elaboração de diretrizes de política de indexação; e o uso da linguagem de indexação para controle de vocabulário. Ambos os trabalhos, em conjunto, demonstraram diferentes aspectos e em diferentes níveis que podem interferir no processo de indexação.

5. Considerações finais

Através do aprofundamento teórico, da observação das bases de dados e da análise dos estudos selecionados pode-se chegar a algumas considerações sobre o uso de linguagem de indexação em bibliotecas universitárias.

O desenvolvimento da fundamentação teórica, com base no levantamento da literatura e análise dos resultados permitiram retomar o objetivo de

analisar estudos sobre linguagens de indexação, por meio da pesquisa bibliográfica, com o intuito de analisar semelhanças e diferenças, métodos, proposições, contribuições e recomendações.

As semelhanças encontradas na literatura analisada abordam, principalmente, a influência do autor e do indexador no processo de indexação devido à diferentes interpretações e perspectivas que afetam a representação e recuperação da informação (Ménard, 2010; Tartarotti, 2014; Golub, Tyrjö, Hansson & Ahlström, 2020).

As diferenças observadas na literatura analisada referem-se a variáveis que interferem no processo de indexação tais como as relacionadas ao Indexador. Nesse sentido, sua formação acadêmica, pré-conceitos e preconceitos, conhecimento em diferentes níveis de mundo e suas experiências podem interferir no processo de indexação, pois se trata de um ser humano que possui personalidade e pensamentos que difere do outro ser humano (Tartarotti, 2014). Enquanto, para outros autores, a variável interferente é o sistema, com a falta de elaboração de políticas e manual de indexação (Fujita, Cruz, Patrício & Branco, 2019). Considera-se que ambas as variáveis estão ligadas inerentemente e interferem no processo de indexação.

Como proposição na literatura analisada, destaca-se que a combinação de vocabulário controlado e não controlado pode, em um futuro próximo, ser benéfica para o processo de indexação, caracterizando, dessa maneira, em um sistema de indexação colaborativa (Ménard, 2010). Ainda nesse sentido, o uso de tesouro, por exemplo, pode melhorar a eficácia na recuperação da informação, se os documentos forem pré-indexados e se a busca for processada com os termos do tesouro, mas nesse ponto envolve o conhecimento do usuário acerca do uso de tesouro no sistema (Sunny & Angadi, 2017).

Uma das contribuições relevantes da literatura analisada consistiu na sistematização dos procedimentos contidos em um manual de linguagem de indexação para catálogos online em uma rede de bibliotecas (Cruz, 2019). O método para a construção da linguagem de indexação, apresenta-se inovador porque consistiu no uso de estruturas hierárquicas para definir categorias e subcategorias, que formaram uma macroestrutura, utilizando parâmetros da *Library of Congress Subject Headings*, Terminologia da Biblioteca Nacional e o Vocabulário do Sistemas de Bibliotecas da USP (Fujita, Moreira, Santos, Cruz & Ribas, 2018). O método utilizado para a construção da linguagem de indexação interfere diretamente no seu funcionamento e, conseqüentemente, em seus resultados. Por isso, necessita

de cuidado e atenção, principalmente, nos detalhes de sua construção.

Com a observação da literatura analisada ressalta-se, também, a recomendação de treinamento e orientações para autores e indexadores sobre o processo de indexação de forma mais eficiente. A importância do treinamento decorre do impacto que a falta dele pode gerar na atuação do bibliotecário. A falta de atualização acerca de como realizar o processo de indexação, especialmente, no uso da linguagem de indexação, atribuição de termos, controle de vocabulário e do próprio sistema pode causar impacto na qualidade da indexação, prejudicando todo o procedimento e a recuperação da informação pelo usuário. Por isso, a necessidade de manter o indexador sempre atualizado com treinamentos e cursos (Ménard, 2010; Tartarotti, 2014; Golub; Tyrkø; Hansson e Ahlström, 2020); e, por outro lado, em investimentos na construção, manutenção e avaliação periódica da linguagem de indexação (Cruz, Fujita e Santos, 2017).

As ideias apresentadas pelos autores envolveram indexação compartilhada, necessidade de treinamento de autores e indexadores para adequada atribuição de termos com consulta da linguagem de indexação e a avaliação e manutenção constante da linguagem.

A partir do desenvolvimento dos trabalhos observou-se que a política de indexação precisa incluir a combinação de multifatores, de forma a garantir uma representação e recuperação da informação adequada ao sistema e seus diferentes atores. Os multifatores e as variáveis considerados que podem interferir nesse processo são: o indexador, a qualidade da linguagem de indexação, o instrumento utilizado, sistemas e o próprio contexto da biblioteca universitária.

Esse contexto requer um sistema de representação e recuperação da informação eficaz e, em função do emprego intensivo de informações com características comuns e como forma de assegurar esse processo, devem ser feitas a elaboração e a implantação da linguagem de indexação. A linguagem de indexação atua nos sistemas de informação, orientando o catalogador sobre quais os melhores termos para a representação de um assunto, da mesma maneira como auxilia os usuários na elaboração de estratégias de buscas por informações no sistema. Em sua ausência pode ocorrer problemas relacionados ao controle de vocabulário no sistema, afetando diretamente a representação e recuperação como resultados não condizentes com a busca realizada devido aos excessos de materiais de outras temáticas e falta dos documentos no assunto pesquisado. Enquanto a política de indexação é

um guia para tomada de decisões e o estabelecimento de diretrizes como: linguagem de indexação utilizada, como será feito o processo de indexação e recursos materiais e financeiros, atendimento das necessidades dos usuários entre outros procedimentos necessários (Moura, Silva & Amorim, 2002; White, 2013; Tejada-Lorent et al., 2014; Vázquez; Pedraza-Jiménez, Condia e Blanco & Rovira 2015).

Recomenda-se que haja continuidade em trabalhos sobre linguagens de indexação, devido ao desenvolvimento constante da tecnologia e da ciência, para que possibilite se aperfeiçoar, acompanhar e evoluir na temática, sem deixar se tornar obsoleto e causar perdas dos dados inseridos nos sistemas utilizados.

Referencias

- American National Standards Institute/National Information Standards Organization (ANSI/NISO). Z39.19-2005 (R2010). Guidelines for the construction format, and management of monolingual controlled vocabularies. Bethesda, Maryland: NISO Press, 2010.
- Arboit, A. E. (2017). É possível uma organização e representação do conhecimento sem partidos?. // P2P & Inov. Rio de Janeiro. 3:1 (set./mar. 2017).
- Cintra, A. M. M.; Tálamo, M. de F. G. M.; Lara, M. L. G.; Kobashi, N. Y. (2002). Para entender as linguagens documentárias. Rio de Janeiro: Polis, 2002.
- Cruz, M. C. A. (2019). Linguagens de indexação em bibliotecas universitárias: estudo analítico em território nacional. 2019. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2019.
- Cruz, M. C. A.; Santos, L. B. P.; Fujita, M. S. L. Linguagens de indexação em bibliotecas universitárias brasileiras: diagnóstico preliminar das regiões sul e sudeste. // X Asociación de Educación e Investigación en Ciencia de la Información de Iberoamérica y el Caribe (X EDICIC), Belo Horizonte, 2016.
- Cruz, M. C. A.; Fujita, M. S. L.; Santos, L. B. P. (2017). Linguagem de indexação no contexto da política de indexação: estudo em bibliotecas universitárias. // Pinho, F. A.; Guimarães, J. A. C. (Orgs.). Memória, tecnologia e cultura na organização do conhecimento. Recife: UFPE, 2017, 217-224.
- Dias, E. W.; Naves, M. M. L. (2007). Análise de assunto: teoria e prática. Brasília: Thesaurus, 2007.
- Fujita, M. S. L. (2017). Manual de política de indexação para bibliotecas universitárias da Unesp. // São Paulo: Unesp (2017) 36.
- Fujita, M. S. L.; Gil Leiva, I. (2010). As linguagens de indexação em bibliotecas nacionais, arquivos e sistemas de informação na América-latina. // Seminário Nacional de Bibliotecas Universitárias. 2. 2010.
- Fujita, M. S. L.; Moreira, W.; Santos, L. B. P.; Cruz, M. C. A.; Ribas, R. R. de B. (2018). Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University (Unesp). // Knowledge Organization. 45:3. 2018.
- Fujita, M. S. L.; Cruz, M. C. A.; Patrício, B. O. M.; Branco, L. B. P. R. (2019). Linguagens de indexação em bibliotecas universitárias: estudo analítico. // Informação & Informação. 24:1 (jan./abr. 2019) 190-225.

- Fujita, M. S. L.; Favato, V. A. M.; Zaniboni, M. M.; Fagundes, S. A. (2016). A elaboração da política de indexação para bibliotecas universitárias da Unesp. // Fujita, M. S. L. (Org.). Política de indexação para bibliotecas: elaboração, avaliação e implantação. Marília: Cultura Acadêmica, 2016.
- Fujita, M. S. L.; Santos, L. B. P. (2016). A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias. // Scire. 22:2 (jul./dez. 2016) 37-46
- Gil Urdiciain, B. (2004). Manual de lenguajes documentales. // Gijón: Trea. 2004.
- Golub, K; Tyrkkö, J; Hansson, J; Ahlström, I. (2020). Subject indexing in humanities: a comparison between a local university repository and international bibliographic service. // Journal of Documentation. 76:6 (2020) 1193-1214.
- International Organization for Standardization (ISO) 25.964-1. (2011). Information and documentation – thesauri and interoperability with other vocabularies – part 1: thesauri for information retrieval. // Genebra. 2011.
- Kitchham, B; Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. // Version 2.3. Keele: Keele University/School of Computer Science and Mathematics/Software Engineering Group. 65 p.
- Menard, E. (2010). Ordinary image retrieval in multilingual context: a comparison of two indexing vocabularies. // Aslib Proceedings: New Information Perspective. 62: 4/5 (2010) 428-437.
- Moura, M. A; Silva, A. P; Amorim, V. R. (2002). A concepção e o uso das linguagens de indexação face às contribuições da semiótica e da semiologia. // Informação & Sociedade: estudos. 12:1. 2002.
- Ramires, A. Q. (2022). Estratégias de leitura no ensino fundamental para formação de leitores profissionais em indexação. 2022. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2019.
- Silva, R. C; Brito, J. F. (2018). Proposta de um manual de indexação para bibliotecas universitárias. // Informação@Profissões. 7:1 (jan./jun. 2018) 92-113.
- Sunny, S. K; Angadi, M. (2017). Evaluation the effectiveness of thesauri In digital information retrieval systems. // The electronic library. 36:1 (2018) 55-70.
- Tartarotti, R. C. D. E. (2014). Atuação bibliotecária no tratamento temático da informação em unidades de informação: estudo comparativo qualitativo. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos – UFSCar, 2014.
- Tejeda-Lorente, A.; Porcel, C; Peis, E; Sanz, R; Herrera-Viedma, E. (2014). A quality-based recommender system to disseminate information in a university digital library. Information Science. :261. 52-69.
- Vállez, M; Pedraza-Jiménez, R; Condia, L; Blanco, S; Rovira, C. (2015). Updating controlled vocabularies by analyzing query logs. // Online Information Review. 39:7 (2015) 870-884.

Enviado: 2022-04-10. Segunda versão: 2022-06-08.
Aceptado: 2022-06-10.

Desarrollo de una taxonomía interoperable a partir del mapeo de taxonomías empíricas: un estudio de caso con los sitios web de las asociaciones de amigos del Camino de Santiago

Developing an interoperable taxonomy by mapping empiric taxonomies: a case study on the websites of the associations of friends of the Way of St James

Francisco Javier GARCÍA MARCO (1), Joaquín LÓPEZ DEL RAMO (2)

(1) Instituto de Investigación en Patrimonio y Humanidades, Universidad de Zaragoza (España), jgarcia@unizar.es (2) Departamento de Periodismo y Comunicación Corporativa, Universidad Juan Carlos I de Madrid (España), joaquin.lopezdelramo@urjc.es

Resumen

Se explora la construcción de taxonomías web interoperables mediante el mapeo de las categorías utilizadas en sus menús de navegación a partir del estudio de un caso real. Se analiza el concepto de taxonomía, y se conceptualizan los menús y mapas de sitio como taxonomías web empíricas. Se realiza un desarrollo con los 34 sitios web operativos de las asociaciones afiliadas a la Federación Española de Asociaciones de Amigos del Camino de Santiago. Para ello, se buscaron los menús y mapas de sitio, se extrajeron sus etiquetas y se codificó su orden. Luego, se mapearon las categorías equivalentes eligiendo la etiqueta más específica o asignando una en caso de ambigüedad. Se reconstruyeron las relaciones jerárquicas con un enfoque inductivo (de abajo a arriba) respetando la inclusividad natural, y, para completarlas, se añadieron algunas clases genéricas. La ordenación sistemática final se inspiró en los modelos de Ranganathan, Laswell, las siete preguntas y el de análisis de dominios. Además, se desarrollaron tres conjuntos de etiquetas susceptibles de servir como título, categoría de menú o descriptor conforme a ISO 25964, de manera que la taxonomía pueda ser útil para un amplio conjunto de funciones de navegación y búsqueda. Como resultado, fue posible mapear la totalidad de las categorías de las asociaciones en una taxonomía de enlace (hub), que puede además utilizarse para generar un tesaurus altamente clasificado, formar el esqueleto de una ontología, soportar el diseño de arquitecturas web y funcionalidades de integración y agregación de páginas web. También permite realizar un mapa de la ecología informacional, y servir de base para futuras colaboraciones y normalizaciones entre sus agentes. Finalmente, las taxonomías web se conceptualizan como un tipo clave de sistemas de organización del conocimiento que deben ser integrados en la teoría, práctica y estrategia de la disciplina; y se ofrecen recomendaciones para la investigación futura.

Palabras clave: Taxonomías. Interoperabilidad semántica. Sitios web. Camino de Santiago. Organizaciones no gubernamentales. Organización del conocimiento.

Abstract

The construction of interoperable web taxonomies by mapping the navigation menu categories of a set of websites, and their subsequent re-hierarchization, is explored. A state of the art reviewing the concept of taxonomy, and specifically the definition and role of web taxonomies is discussed: menus and site maps are conceptualized as empirical web taxonomies. A specific methodology was developed working on the 34 operating websites of the associations affiliated to the Spanish Federation of Associations of Friends of the St. James' Way. First, the menus and site maps were analysed, their tags extracted and their order encoded. The equivalent categories were mapped by choosing the most specific label or by assigning one in case of ambiguity. The hierarchical relationships were reconstructed using an inductive approach (bottom-up). To complete the taxonomy, some generic classes were added. The final options of systematic ordering were inspired by the models of Ranganathan, Laswell, seven-questions and domain analysis. Three sets of alternative tags were created to serve as titles, menu categories or ISO 25964 descriptors, so that the hub taxonomy may support a wide range of navigation and search functions. It was possible to map all the associations' menu categories to the hub taxonomy, which may also be used to generate a highly classified thesaurus, form the skeleton of an ontology, support the design of web architectures, and assist in the integration and aggregation of external information in web pages. The taxonomy can be also useful to create a map of the St James' Way informational ecology, and serve as a basis for future standardization and collaborations among its agents. As a final conclusion, web taxonomies are conceptualized as a key type of knowledge organization system that must be integrated into the theory, practice and strategy of the discipline. Finally, recommendations for future research are suggested.

Keywords: Taxonomies. Semantic interoperability. Websites. St. James' Way. Non-governmental organizations. Knowledge organization.

1. Introducción

El objetivo de este artículo es presentar el desarrollo de una taxonomía interoperable para los sitios web de las asociaciones de amigos del Camino de Santiago que se construyó a partir de las categorías utilizadas en sus menús de navegación, jerarquizados según la inclusividad de sus conceptos y la aplicación de varios modelos de alto nivel. Para enmarcar este desarrollo concreto, se discute el concepto de taxonomía y su aplicación en el entorno digital en el apartado de estado de la cuestión.

El trabajo forma parte de un proyecto más amplio en el que se está estudiando la ecología informacional del Camino de Santiago como ejemplo de las posibilidades de integración de la información cultural procedente de bibliotecas, archivos, museos y otras instituciones GLAM en las aplicaciones y sitios web de información turística (García Marco, 2021). Se trata de comprobar cómo diferentes agentes interesados en el Camino de Santiago proporcionan información sobre un mismo dominio desde diferentes puntos de vista y cómo la ecología que forman se puede optimizar a partir del conocimiento obtenido.

Hasta el momento, se han realizado estudios sobre las administraciones municipales (López del Ramo y García Marco, 2017) y las autonómicas (López del Ramo y García Marco, 2018) y están en curso para las asociaciones y empresas. Estas comparaciones, basadas en una matriz de análisis diseñada por Joaquín López del Ramo (2017) y compuesta de seis dimensiones, estudian: sus datos de localización; estructura organizativa; red hipertextual; interactividad con el usuario; encuadres, temas y recursos informativos; y grafismo.

En este trabajo se aprovecharon los datos obtenidos en el análisis de la segunda dimensión, la estructural, que se ocupa de estudiar en qué lugar del sitio web se sitúa la información sobre el Camino, cómo estructuran las páginas en menús y mapas, y cuántas páginas contiene dicha estructura. Sobre estos datos, se realizó un análisis en detalle de la organización de la información presentada usuario a través de menús y mapas de navegación, conceptualizados como taxonomías. Los resultados preliminares sobre la organización empírica de esos sitios web han sido ya publicados en un trabajo anterior (López del Ramo y García Marco, 2022). En ese mismo trabajo se aborda también una clarificación conceptual y teórica de su terminología dentro del campo de la organización del conocimiento. Menús y mapas de sitio se identifican como los sistemas de organización conceptual jerárquica de los sitios web,

dentro del enfoque que propone la complementariedad entre sistemas alfabéticos y sistemáticos en el acceso a la información; y que está en la base de pares de sistemas tan diferentes como las enciclopedias sistemáticas y alfabéticas, la conjunción de clasificaciones y materias en las bibliotecas, o los tesauros con su doble presentación (García Marco, 2016).

Sobre este análisis, se da un nuevo paso utilizando las taxonomías de los sitios web de las asociaciones para construir una taxonomía de enlace que las relacione a todas y, de esta forma, permita interoperar los sitios a nivel de página y de los conceptos individualizados por los nodos. Y también para que pueda integrar automáticamente recursos externos cuyos vocabularios hayan sido debidamente mapeados con dicha taxonomía de enlace.

En definitiva, este trabajo se inserta dentro del objetivo más general del proyecto de optimizar las ecologías informacionales favoreciendo su interoperabilidad mediante sistemas de organización del conocimiento de enlace (García-Marco, Galindo, Lasala y López del Ramo (2020). En concreto, se pretende facilitar avances en cuatro direcciones: 1) mejorar el acceso por parte de los usuarios ampliando las equivalencias terminológicas y a las asociaciones entre las etiquetas; 2) facilitar la compartición de información a través de la normalización de las taxonomías o alternativamente su referencia a través de un mapeo; 3) estudiar hasta qué punto puede ser posible proponer un modelo canónico de organización del conocimiento para un sector de sitios web del mismo tipo; y 4) contribuir al análisis empírico de ecologías informacionales al facilitar una taxonomía modelo que permita observaciones precisas de los temas tratados por cada tipo de agente, las comparaciones entre las taxonomías empíricas de los diferentes sitios web, y la toma de ulteriores decisiones políticas sobre cooperación y especialización entre los agentes.

En este artículo se pretende avanzar en esos cuatro frentes de investigación respondiendo a tres preguntas de investigación concretas, que se ordenan de la más específica a la más general:

1. ¿Es posible mapear entre sí las categorías de los menús y mapas de los sitios web de un determinado dominio de interés? Y secundariamente: ¿con qué nivel de éxito? ¿qué problemas emergen? ¿con qué soluciones se pueden abordar?
2. ¿Es posible desarrollar un sistema de organización del conocimiento de enlace que permita organizar dicho mapeo de una forma inteligible utilizando los principios de la organización del conocimiento? Y secundariamente:

¿con qué nivel de éxito? ¿qué problemas emergen? ¿con qué soluciones se pueden abordar?

3. ¿Tiene sentido considerar los menús y mapas de sitio como sistemas de organización del conocimiento, en particular como taxonomías, y por tanto integrarlos en el dominio de la disciplina?

En las páginas siguientes, se intenta responder a estas preguntas, comenzando por un encuadre teórico dentro de los campos de la gestión del conocimiento, las ecologías informacionales y la organización del conocimiento; siguiendo con la presentación de la metodología, los resultados —en particular la taxonomía de enlace conseguida—, y las conclusiones y recomendaciones.

2. Estado de la cuestión

2.1. Entre la gestión de la información y la del conocimiento: las ecologías informacionales y su organización

Desde un punto de vista teórico, el proyecto se inserta dentro de la línea de investigación sobre ecologías informacionales que se viene realizando en el marco de la disciplina de la ciencia de la información, pues se apoya en registros objetivos de información (Wilson, 2002), en este caso las páginas web de las asociaciones. Sin embargo, no pierde de vista la perspectiva más amplia de la gestión del conocimiento, en el sentido de que —según se constata en los diferentes ámbitos de práctica— la información documentada es sólo una parte de la información utilizada por las personas y las comunidades, que buena parte de esta se presenta bajo la forma de conocimiento tácito (Nonaka, 1994), y que la correcta interacción entre información tácita, explícita y registrada es clave para el funcionamiento de cualquier organización que aprende. En este sentido, los sitios webs de las asociaciones se comprenden como núcleos de captura, almacenamiento y difusión del conocimiento de sus miembros asociados. Se trata de sistemas integrados o mixtos, pues combinan un aspecto interno, en ocasiones casi archivístico —como los registros textuales y gráficos de sus actividades o las actas de las reuniones—, con otro decididamente externo —o comunicativo en el sentido de las ciencias de la comunicación— por el que trasladan a diferentes audiencias sus mensajes, esto es, constituyendo un medio de comunicación.

Otra idea clave de este proyecto es que estos sitios web no pueden entenderse de forma aislada —ni ninguno, en realidad—, sino como parte de una ecología informacional, un concepto intuitivo pero que conviene intentar precisar. De forma

sencilla, una ecología informacional puede definirse como un sistema —un conjunto de elementos en interacción dinámica en función de un objetivo (Rosnay, 1979)— en el que diversos agentes y sus recursos de difusión (elementos) colaboran, compiten, coexisten, etc., (interacción dinámica) dentro de un nicho configurado por las necesidades informativas de un conjunto de personas y sus diversas agrupaciones e interacciones (objetivo) (García Marco, 2017). En computación e información, el concepto de sistema de información se utiliza restringidamente para los sistemas diseñados; mientras que el concepto de ecología informacional, también sistémico, sirve precisamente para los sistemas de información surgidos de procesos sociales no diseñados, si se quiere “naturales”.

Así, por ejemplo, en el caso del Camino es posible estudiar cómo sus diferentes agentes se relacionan; e incluso es posible clasificarlos en grandes grupos de acuerdo a sus motivaciones y a la información que ofrecen: gubernamentales —y dentro de ellas internacionales, nacionales, regionales, comarcales o municipales—; no gubernamentales sin ánimo de lucro —notablemente asociaciones, iglesias y personas individuales—; y empresas (López del Ramo y García Marco, 2018, p. 706-707).

Como ocurre con el concepto que está en su origen —la idea de sistema—, el concepto de ecología informacional se ajusta por niveles de forma parecida a un microscopio o un telescopio, o, más precisamente en acertada metáfora de Joel de Rosnay (1979), como un macroscopio. Quiere decir esto que una ecología informacional se puede enfocar según el dominio de análisis: desde las secciones de un sitio web, hasta los sitios web de una comarca, región, nación o especialidad. La necesidad de operar una reducción para poder acometer el objetivo de investigación impedirá a su vez desentrañar la complejidad a otros niveles, por lo que será interesante enfocar la herramienta de diversas maneras para alcanzar una perspectiva más global y acertada.

Para comprender cualquier sistema de información hay que partir de la propia naturaleza humana y específicamente de su cognición. Un aspecto básico del conocimiento y la comunicación humana que se experimenta a todos los niveles es su carácter estructurado (García Marco y Esteban, 1993). Para abordar la complejidad de lo real, el ser humano elige centros de atención, y dichos centros de atención son clasificados en organizaciones más grandes para permitir su recuerdo en la memoria a largo plazo y su manejo en la memoria a corto plazo, de capacidad limitada. Por otra parte, en la medida en que el co-

nocimiento va siendo compartido, también las estructuras de conocimiento se convierten en objetos compartidos y cobran también un aspecto comunicacional. Esta realidad es el fundamento de la organización del conocimiento y de su modalidad investigada aquí, las taxonomías.

Los sitios web —como instrumentos para compartir conocimiento— necesitan de un mapa inteligible por su comunidad de usuarios que permita llegar de lo sabido a lo que se quiere conocer con arreglo a un marco en gran medida compartido, predecible. Por lo tanto, en ellos se intenta llegar a un equilibrio entre las características idiosincrásicas de la materia del sitio web y de su orientación comunicativa, de una parte; y la necesidad de recurrir a marcos de referencia compartidos que permitan la inferencia comunicativa, en este caso, acceder a la información que el usuario —muchas veces desconocido en una página web— desea localizar, de la otra. Ciertamente, cualquier mensaje o conjunto de mensajes requiere una organización que permita un acceso y comprensión lo más eficaz y eficiente, es decir, requiere un sistema de organización del conocimiento; y, la Internet como espacio de recuperación en general y los sitios web en particular tienen que tener sus propios sistemas de organización del conocimiento, e incluso dar origen a nuevos tipos (Hjørland, 2016).

2.2. Taxonomías, taxonomías corporativas y taxonomías web

En los sitios web esa función la desempeñan dos tipos principales de herramientas: los menús de navegación, prácticamente omnipresentes con mayor o peor calidad, y los mapas de sitio, que no siempre están disponibles y que en la mayor parte de los casos —salvando problemas de actualización o recortes de niveles para limitar su tamaño— suelen corresponderse, al menos en gran parte, con los menús de navegación. Aunque esto puede no ser cierto en grandes sitios corporativos compuestos de muchos sitios web con diferentes responsables de su desarrollo y mantenimiento y diversos estados de mejora y actualización; sí se cumple en los sitios web de pequeño y mediano tamaño, como es el caso de los que estamos estudiando.

En esta investigación se ha adoptado el término taxonomía para referirnos a estos menús y tablas de contenido de acuerdo a los usos más recientes en el campo de la informática web y corporativa. No obstante, dicho sentido del término constituye un uso restringido y técnico de un concepto que ha sido usado de forma distinta en diferentes disciplinas y tradiciones, generando un término notablemente polisémico, a veces contradictorio en sus acepciones en aspectos clave (Gilchrist,

2003a; Garshol, 2004). Como ha señalado Gnoli (2020), el término taxonomía no es nuevo en la organización del conocimiento, sino absolutamente central. Su idea, sin esa denominación, fue conceptualizada hacia el 343 a. C. por Aristóteles (1992, ed. en esp.), que estableció la relación género-especie como el criterio fundamental para clasificar los seres vivos. El término taxonomía fue acuñado por su epígono Linneo en su magna obra *Sistema Naturae* de 1735; y de allí, a través de la cladística, se ha convertido en una de las inspiraciones y referencias fundamentales de la teoría de la clasificación bibliográfica, que constituye el núcleo de la organización del conocimiento (Dahlberg, 2006, 2014). La nota central del concepto clásico de taxonomía es la estricta y exclusiva relación de pertenencia entre las clases, que deben ser necesariamente disjuntas según los principios de la lógica aristotélica. También interesa resaltar que las taxonomías —y dentro de ellas las clasificaciones bibliográficas— permiten hacer abarcable y deducible un conjunto de elementos por medio de una estructura lógica, basada en principios claros y accionables —en el sentido de la organización del conocimiento como formación de teorías formulado por Jaenecke (1997)—. Esta posibilidad de encontrar el camino entre distintos puntos de una estructura conceptual es también central a la experiencia de uso de la Internet, en dónde, como es conocido, se denomina navegación.

En este sentido, el término ha sido cooptado desde los años noventa del siglo pasado para nombrar a los diferentes esquemas que se han inventado para clasificar los contenidos de los sitios web, bases de datos y otros nuevos sistemas surgidos con la revolución digital (Gilchrist & Kibby, 2000; Gilchrist, 2001, 2003b). En parte, creemos que esto ha sucedido porque el término como tal no estaba siendo utilizado dentro de la organización del conocimiento por entonces consolidada (clasificaciones bibliográficas, encabezamientos de materia, tesauros, etc.). Por ello, podía ser utilizado para destacar la novedad de estas nuevas aproximaciones, novedad que surge tanto de la tecnología subyacente como del hecho de que están centrados más en el contenido organizacional que en el disciplinar.

Su uso se ha visto beneficiado por su confluencia con la gestión del conocimiento y la gestión de la información en las organizaciones, fundamentalmente bajo la forma de taxonomías corporativas (Gilchrist & Kibby, 2000; Gilchrist, 2001, 2003b; Zhonghong, Chaudhry and Khoo, 2006). Las taxonomías corporativas sustentan el acceso a la información de una entidad, a diferencia de los sistemas de organización del conocimiento de bibliotecas y centros de documentación que están

orientados a un campo de estudio o las clasificaciones archivísticas tradicionales centradas en el principio de procedencia, aunque están cercanas a los enfoques de la clasificación funcional de archivos. Desde ese punto de vista, las taxonomías corporativas se pueden definir como sistemas de organización del conocimiento complejos que son capaces de organizar la totalidad de los recursos de información de una organización (Gilchrist, 2003a), y se trata de un campo consolidado de práctica hasta el punto de que han surgido compañías especializadas en proveer estos servicios como Wordmap (2019).

La existencia de un uso técnico del concepto de taxonomía como taxonomía corporativa no significa que en buena parte de la literatura el concepto taxonomía no siga utilizándose de forma muy laxa, como sinónimo de cualquier estructura que contenga una relación jerárquica —desde clasificaciones, menús y tablas sistemáticas, hasta tesauros y encabezamientos de materia—. Esta confusión de fondo ha inspirado diversos trabajos de clarificación terminológica (e.g., Gilchrist, 2003b; Garshol, 2004) que, a pesar de su gran valor, tampoco han permitido alcanzar un consenso entre expertos, ni mucho menos normar el uso entre la amplísima comunidad multidisciplinar que usa el término en un sentido amplio.

Por otra parte, este uso generalizado tiene también aspectos positivos. En el contexto más amplio de la consolidación de las ontologías —y de la modelización ontológica de otros sistemas de organización del conocimiento anteriores—, el término taxonomía se está generalizando para denominar a su parte jerárquica —el esqueleto en torno al cual construir las y desarrollarlas con eficacia (Costin, Eastman e Issa, 2017)—, contribuyendo así a resituar la clasificación en el corazón de los esfuerzos más punteros por organizar la Internet (Dahlberg, 2006, 2014).

Por otra parte, y en un ámbito mucho más concreto, el término taxonomía fue rápidamente adoptado por los creadores de software para gestión de contenidos (Content Management Systems, CMS), que empezaron a utilizar con gran éxito las organizaciones y personas que deseaban publicar su información en formato digital de una forma más dinámica sin el concurso continuo de profesionales de la edición web, frente a la llamada publicación estática, en ficheros html, y a la programación web.

Conforme la web se ha transformado en el entorno de información de referencia para la mayoría de las actividades de nuestras sociedades avanzadas, los CMS se han convertido en una herramienta generalizada y casi imprescindible. Los sitios web de las asociaciones de amigos del

Camino de Santiago efectivamente hacen un uso extensivo de CMS: 50 % WordPress, Blogger (11,76 %), Bootstrap (5,88 %), CM4all (2,94 %), ADODB/ASP developments (2,94 %), DotNetNuke (2,94 %), Google Sites (2,94 %) y Joomla (2,94 %). Solo no usan un CMS el 17,65 %, aunque sí poseen menús.

En los sitios web, la herramienta de navegación por excelencia es en general el hipertexto entre un anclaje y un destino. Sin embargo, más concretamente, actúan como sistemas de referencia ciertos conjuntos o concentraciones más o menos organizadas de ellos: a saber, los menús, los mapas de sitio y las palabras clave asignadas, conocidas como etiquetas.

Aunque existen menús e índices de un solo nivel ordenados en una dimensión no conceptual (alfabética, cronológica), en general los menús y mapas de sitio respetan los principios de la cládica: todos los ítems están clasificados (sistematicidad) y la relación entre las categorías es monojerárquica. Este último criterio se cumple en menor medida, pues el hipertexto permite crear fácilmente relaciones a una sola página desde diferentes categorías del menú o del mapa, aunque esto es menos frecuente. Sin embargo, dicho esto, menús y mapas muestran en la mayoría de los casos una disposición taxonómica a efectos de la experiencia de usuario, pues se le ofrece una presentación clasificada en la que elige un único destino.

En sitios orientados a la recuperación de catálogo —por ejemplo, una biblioteca o una tienda virtual—, y que no son los que estudiamos aquí, los menús adoptan una disposición facetada, que permite elegir las dimensiones de búsqueda de manera independiente; y luego o bien el sistema va filtrando o el usuario lanza una búsqueda. En estos sistemas, las facetes son los propios campos de búsqueda o una agregación de ellos, y son realmente opciones alternativas de clasificar el dominio en cuestión.

Los CMS entraron pronto a ofrecer herramientas para facilitar la gestión de este aspecto fundamental de la usabilidad de los sitios web. Ya en 2002 *Drupal*, en su versión 4.0.0, incluyó un módulo de taxonomías para gestionar sus herramientas de navegación (menús y etiquetas), que se transformó en multilingüe en su versión 8.0.0 de 2015 (Drupal, 2020). En 2007, *WordPress*, su competidor y líder en el entorno *blogger*, incluyó también un módulo de taxonomías en su versión 2.3 (WordPress, 2007). A partir de estos populares programas, las taxonomías se han abierto un espacio propio no solo en el mundo de la empresa y la informática corporativa, sino en el seno de la cultura popular contemporánea, mediada por las

redes sociales y los sitios web. La capacidad que proporcionan de integrar menús y etiquetado social ha contribuido también a aflorar el valor de la organización del conocimiento en entornos mucho más amplios. Es legítimo afirmar que, junto con el etiquetado social, las taxonomías forman actualmente el corazón de la organización del conocimiento del nuevo medio: la Internet.

2.3. Interoperabilidad, integración y compatibilidad entre taxonomías

Como se ha señalado en la introducción, esta investigación se inserta precisamente en un proyecto más amplio que busca analizar estos instrumentos de organización del conocimiento digital a través de casos concretos, en esta ocasión las asociaciones de amigos del Camino de Santiago. La idea básica es estudiar cómo las asociaciones organizan su dominio de conocimiento —el Camino, la vida asociativa y la peregrinación— y presentan esa organización a los usuarios de sus sitios web, que son a su vez estudiados como un subsistema de la ecología informacional en torno al Camino de Santiago.

Pero además, se pretende dar otro paso más: a saber, comprobar si dichas taxonomías se pueden compatibilizar para facilitar la interoperabilidad de la información entre ellas y los servicios de agregación: por ejemplo, mostrando todas las páginas sobre los servicios de albergue soportados por todas las asociaciones, monumentos destacados o leyendas típicas; o integrando en una página otra información complementaria procedente de otra fuente externa. En un paso más allá, se busca crear una metataxonomía que pueda eventualmente sistematizarlas y servir para la definición de un marco común de trabajo y búsqueda, de una organización del conocimiento compartida que ahorre el tiempo de los gestores de los sitios web y de los usuarios que los consulten; o pueda al menos servir de referencia para futuros trabajos en dicha dirección.

Este programa es importante tanto por razones teóricas como prácticas. Los sistemas de organización del conocimiento nacen siempre de comunidades de uso y de la iniciativa de líderes que, a su escala, son capaces de aportar el trabajo necesario. En el caso de los sitios web y sus taxonomías, este carácter calidoscópico se incrementa porque son especialmente cercanas a los usuarios (Gilchrist, 2003b; Zhonghong, Chaudhry y Khoo, 2006). Y sin embargo, los procesos de integración inherentes a la eficiencia de las economías de escala, el entorno web y el propio carácter universalista de la organización del conocimiento han puesto siempre la compatibilidad e integración de los sistemas de organización del

conocimiento en el centro del debate científico en la disciplina. Por ello, la interoperabilidad ha sido siempre uno de los frentes de investigación especialmente atendidos en organización del conocimiento (Zweng, 2019), cuyo interés se ha incrementado si cabe con la necesidad de asegurar la interoperabilidad en los sistemas de información digital, cuyo medio es global por definición.

Por ello, la ventaja de la especificidad que las taxonomías en Internet ofrecen a sus usuarios debe compensarse con mecanismos de interoperabilidad que aseguren el funcionamiento de los agentes automáticos —cada vez más frecuentes—, y que proporcionen un lenguaje cada vez más común a los usuarios, aunque sea de forma subyacente a través de mapeos y estructuras de referencia.

En el caso específico que nos ocupa, la segunda parte de la nueva norma de tesauros ISO 25964-2 (ISO, 2013; Dextre Clarke, 2011, 2012; Dextre Clarke y Zeng, 2012) aborda específicamente la interoperabilidad de los tesauros con las taxonomías, y tanto en sus aspectos generales (modelos de interconexión) como específicos (objetivos, tipos, componentes y relaciones) constituye una guía de trabajo fundamental. Su modelo de *hub*, troncal o de enlace ha servido de inspiración directa para esta investigación.

No hemos podido documentar otras experiencias o propuestas concretas de mapeo que involucren taxonomías más allá de las relatadas en la norma, tras realizar la búsqueda “*taxonom* AND 25964*” en *Title and All fields* en *Web of Science*. Sin embargo, se pueden encontrar numerosas experiencias de mapeo entre taxonomías para favorecer la recuperación federada o agregada (Avesani, Giunchiglia y Yatskevich, 2005; Aanen, Vandic y Frasinca, 2015; Lin, Wang, Jia, Xiong, Zhang y Cheng, 2015; Nederstigt, Vandic y Frasinca, 2012, 2016), especialmente si se considera el término taxonomía en sentido amplio. Más cercanas a los objetivos de esta investigación, se encuentran también referencias sobre la construcción de una taxonomía que sirva de *hub* a las utilizadas en las fuentes (Jung, 2006, 2008; Mourino-García et al., 2018).

Tampoco se han podido encontrar referencias a investigaciones sobre el mapeo de menús entre sí o a una taxonomía de enlace, para lo que se usó la búsqueda “*menu* AND (interoper* OR map*)*” —tanto en *Title* (10 ítems, todos no relevantes) como en *All fields* (564 ítems, todos no relevantes)— en *Web of Science* el 27 de mayo de 2022 por última vez. La búsqueda se hizo deliberadamente con trucados muy amplios para evitar silencios, dado que búsquedas anteriores más estrictas ya habían dado resultados infructuosos.

Casi seguro, un silencio tan clamoroso se debe a razones de agenda de investigación disciplinar. Puesto que el mapeo por su misma naturaleza trasciende el marco de la arquitectura web para referirse al problema de recuperación en espacios de información, la investigación sobre menús y mapas queda restringida al campo de la arquitectura de los sitios web y su usabilidad; mientras que resulta redundante como frente de investigación dentro del campo de la recuperación de la información, ya que la búsqueda de páginas se considera resuelta con buscadores y, por lo tanto, no tiene sentido proponer y realizar mapeos intermedios complicando un sistema que funciona, aunque sus grafos puedan aprovechar mapeos ya realizados como los que publica *Wikidata*.

Otro problema añadido es que, por su complejidad, los mapeos conceptuales necesarios para la construcción de taxonomías de enlace tienen que ser realizados por juicio experto, y no pueden competir en su relación eficacia/eficiencia con otros enfoques automatizados, que suelen limitarse a mapear los sistemas entre sí o con relación a uno que se toma como central, aunque los resultados sean parciales y contengan ambigüedad.

Este hecho, no obstante, no excluye que el mapeo de menús y mapas de sitio no pueda ser interesante para otros objetivos en campos de investigación ajenos a la recuperación de la información en la web. En este trabajo se defiende su uso como instrumento de estudio de ecologías informacionales específicas y sus comunidades de conocimiento de referencia, para optimizar y potenciar sus procesos de gestión del conocimiento (captura, almacenamiento, difusión). En este caso, se plantea el problema de organización del conocimiento desde el interés más amplio de la gestión del conocimiento, sin limitarse, por lo ya explicado, en exclusiva a la recuperación de la información con listas sistemáticas.

Precisamente, en este artículo se intenta contribuir a esta línea de trabajo resolviendo un caso concreto de mapeo del conjunto de las taxonomías de los sitios web de las asociaciones federadas del Camino, que sirva, en un segundo paso, para proponer una taxonomía común, que, a su vez, permita dar pasos más avanzados en la integración e interoperabilidad de sus contenidos de cara a su comunidad de usuarios, tanto potenciales como efectivos. En las secciones siguientes se presentan la metodología, los resultados, la discusión, conclusiones y recomendaciones de dicho esfuerzo.

3. Metodología y resultados

En general, el mapeo y la construcción de la taxonomía de enlace se desarrolló mediante un

procedimiento inductivo —en la terminología de ISO 25964, ascendente (*bottom-up*, en inglés)—, puesto que se pretendía incorporar a la nueva taxonomía la totalidad de las categorías existentes en los menús de los sitios web, es decir, hacer un mapeo completo. Se intentó respetar estrictamente el principio taxonómico de que cada categoría tuviera un único sitio en el árbol, y también que el mapeo de las categorías previas fuera una aplicación o función sobreyectiva. Los principios de inclusividad y sobreyección ya fuerzan en buena parte el resultado final, pero efectivamente en los niveles superiores caben grados de libertad, que debían ser reducidos mediante un procedimiento deductivo —en el lenguaje de la norma, descendente (*top-down*, en inglés)—, que se inspiraron en diversos modelos teóricos y decisiones prácticas que se relatan después.

El camino desde las taxonomías de los sitios web hasta la propuesta de una taxonomía de enlace y sus posibles proyecciones se realizó en las siete etapas que se describen en las subsecciones siguientes. La metodología y los resultados se van presentando acompasadamente en cada una de las fases por su carácter secuencial.

3.1. Determinación del conjunto de análisis

Como se ha señalado en la introducción, el dominio estudiado está constituido por el universo de los 35 sitios web de las asociaciones de amigos del Camino pertenecientes a la Federación Española de Asociaciones de Amigos del Camino de Santiago entre septiembre de 2020 y mayo de 2021, según estaban publicadas en <https://www.caminosantiago.org/cpperegrino/asociaciones/asociaciones.asp>. Se incluyó también la página de la propia Federación, pues constituye una asociación. Se pudieron analizar 34 porque una de ellas se encontraba caída permanentemente.

La toma de datos inicial fue realizada en el marco de la encuesta más amplia sobre dichos sitios web que se realizó en otoño de 2020. En ese primer estudio se analizó únicamente el primer nivel de categorías, permitiendo identificar el tipo de taxonomía según su criterio de ordenación y ciertas características cuantitativas muy interesantes, notablemente el número de secciones y de niveles, cuyos resultados se describen en García Marco y López del Ramo (2022).

3.2. Identificación de los menús y determinación de su alcance taxonómico

Una vez determinado el universo de estudio —los menús de los 34 sitios web accesibles de las asociaciones—, la siguiente cuestión importante fue determinar el conjunto de partida para el desarrollo de la taxonomía de enlace.

Se constató que ninguno de los sitios contaba con un mapa dirigido al usuario, por lo que se eligió como único objeto de análisis a efectos de esta investigación el conjunto de los menús de los sitios que aparecieran de forma exenta y en todas las páginas. Se desconsideraron, por consiguiente, las meras listas de tópicos u otros elementos destacados muy variables, pues se trataba de analizar los instrumentos globales de organización del conocimiento de los sitios, esto es, para el conjunto de cada sitio web, de todas sus páginas, no de sus secciones o de las páginas aisladas.

Por tanto, sólo se atendió la organización sistemática y orientada a los usuarios de los sitios — y, por tanto, visible o “front-end”—, es decir, los menús principales, entendidos como jerarquías de categorías que describen el sitio completo. No se ha investigado la estructura “back-end”, la revelada por la organización interna en carpetas o

por las bases de datos utilizadas por los gestores de contenidos, aunque es conocido que suele coincidir con los menús visibles. Por ello, la taxonomía aquí presentada se refiere a las taxonomías navegacionales públicas de las asociaciones, tal y cómo aparecen en sus sitios web. Es cierto que, por la naturaleza de dichos sitios, dirigidos primeramente a los asociados y en segundo lugar a los potenciales peregrinos (y potenciales futuros asociados), la taxonomía pública y la taxonomía corporativa o interna pueden coincidir en gran manera. En muchos casos, el sitio web es la propia herramienta de comunicación de la asociación por lo que éste constituye el instrumento fundamental de gestión del conocimiento de la asociación. Por consiguiente, puede considerarse que su taxonomía pública también responde a su forma idiosincrásica de organizar la información, es decir, a su taxonomía corporativa.

Menús de las asociaciones de amigos del Camino de Santiago: Madrid

URL: <http://www.demadridalcamino.org/>

Lugar: Madrid
Provincia: Madrid

| Término libre | Nivel | Dep. | Nº Obsoleto | Código sistemático | Observaciones |
|--|-------|------|--------------------------|--------------------|---------------|
| Inicio | 1 | 0 | <input type="checkbox"/> | 00 | |
| Día a día | 1 | 1 | <input type="checkbox"/> | 01 | |
| El Camino de Santiago | 1 | 2 | <input type="checkbox"/> | 02 | |
| ¿Qué camino escoger? | 2 | 2 | <input type="checkbox"/> | 02.01 | |
| Compostela | 2 | 2 | <input type="checkbox"/> | 02.02 | |
| Consejos prácticos | 2 | 2 | <input type="checkbox"/> | 02.03 | |
| En bicicleta | 2 | 2 | <input type="checkbox"/> | 02.04 | |
| Albergues | 2 | 2 | <input type="checkbox"/> | 02.05 | |
| Nuestro albergue en Tardajos | 2 | 2 | <input type="checkbox"/> | 02.06 | |
| Remedios | 2 | 2 | <input type="checkbox"/> | 02.07 | |
| Filosofía | 2 | 2 | <input type="checkbox"/> | 02.08 | |
| Simbolos | 2 | 2 | <input type="checkbox"/> | 02.09 | |
| Credencial | 1 | 3 | <input type="checkbox"/> | 03 | |
| Camino de Madrid | 1 | 4 | <input type="checkbox"/> | 04 | |
| Albergues | 2 | 4 | <input type="checkbox"/> | 04.01 | |
| Recorrido | 2 | 4 | <input type="checkbox"/> | 04.02 | |
| 1. Madrid-Colmenar Viejo | 2 | 4 | <input type="checkbox"/> | 04.03 | |
| 2. Colmenar Viejo-Cercedilla | 2 | 4 | <input type="checkbox"/> | 04.04 | |
| 3. Cercedilla-Segovia | 2 | 4 | <input type="checkbox"/> | 04.05 | |
| 4. Segovia-Sta. María la Real de Nieva | 2 | 4 | <input type="checkbox"/> | 04.06 | |
| 5. Sta. María la Real de Nieva-Alcazarén | 2 | 4 | <input type="checkbox"/> | 04.07 | |
| 6. Alcazarén-Simancas | 2 | 4 | <input type="checkbox"/> | 04.08 | |
| 7. Simancas-Medina de Rioseco | 2 | 4 | <input type="checkbox"/> | 04.09 | |
| 8. Medina de Rioseco-Sahagún | 2 | 4 | <input type="checkbox"/> | 04.10 | |
| Noticias | 2 | 4 | <input type="checkbox"/> | 04.11 | |
| Plan Director 2012-2021 | 2 | 4 | <input type="checkbox"/> | 04.12 | |
| Estadísticas de peregrinos | 2 | 4 | <input type="checkbox"/> | 04.13 | |
| Publicaciones | 1 | 5 | <input type="checkbox"/> | 05 | |
| Más sobre el Camino | 1 | 6 | <input type="checkbox"/> | 06 | |
| Enlaces | 2 | 6 | <input type="checkbox"/> | 06.01 | |
| Transporte | 2 | 6 | <input type="checkbox"/> | 06.02 | |
| Tracks GPS | 2 | 6 | <input type="checkbox"/> | 06.03 | |
| Ecología | 2 | 6 | <input type="checkbox"/> | 06.04 | |
| Biblioteca básica | 2 | 6 | <input type="checkbox"/> | 06.05 | |
| ¿Preguntas? | 2 | 6 | <input type="checkbox"/> | 06.06 | |
| Después del Camino | 2 | 6 | <input type="checkbox"/> | 06.07 | |
| El PAU de Las Tablas y el Camino de Santiago | 2 | 6 | <input type="checkbox"/> | 06.08 | |
| Quiénes somos | 1 | 7 | <input type="checkbox"/> | 07 | |
| Actividades | 2 | 7 | <input type="checkbox"/> | 07.01 | |
| Hacerse socio | 2 | 7 | <input type="checkbox"/> | 07.02 | |
| Contactar | 1 | 8 | <input type="checkbox"/> | 08 | |

Niveles de menú: 2 CMS Google Sites

Observaciones:

Figura 5. Vista de la base de datos con el menú del sitio web de la asociación de Madrid

3.3. Generación de la base de datos de datos brutos ligadas al url del sitio y carga completa de los menús

Tras cargar en FileMaker Pro 18 la hoja de cálculo con los datos del análisis de noviembre de 2020,

se implementó una base de datos relacional a la que se añadió una nueva tabla con campos para las etiquetas, número de nivel jerárquico y número de orden de la categoría en cada nivel. El url del sitio se tomó como campo clave, al ser único. (Si

se desea que la base de datos tenga utilidad diacrónica, se debería generar una clave administrada automáticamente por el gestor de bases de datos en previsión de que en el futuro puedan cambiar las URL.) En la Figura 5, en la página anterior, se puede apreciar una vista de la ficha de datos para la Asociación de Amigos del Camino de Santiago de Madrid a la derecha. A la izquierda se ve el resultado del análisis de noviembre de 2020, limitado al primer nivel y con formato de tabla plana.

| <i>Término</i> | <i>Ocurrencias</i> |
|-----------------------|--------------------|
| Inicio | 21 |
| Contacto | 17 |
| Noticias | 15 |
| Enlaces | 14 |
| Actividades | 12 |
| Asociación | 11 |
| Información | 11 |
| Etapas | 10 |
| Hospitalidad | 9 |
| Mapas | 9 |
| Relatos | 9 |
| Fotos/Videos | 8 |
| La Asociación | 8 |
| La Credencial | 8 |
| Albergues | 7 |
| Consejos | 7 |
| Quiénes Somos | 7 |
| Publicaciones | 6 |
| El Camino de Santiago | 5 |
| Hazte socio | 5 |
| Historia | 5 |
| Caminos | 4 |
| El Camino | 4 |
| Enlaces de interés | 4 |
| Estatutos | 4 |
| Fotos | 4 |
| Galería de fotos | 4 |
| Historia del Camino | 4 |
| Junta directiva | 4 |
| Actualidad | 3 |
| Biblioteca | 3 |
| Concursos | 3 |
| Consejos prácticos | 3 |
| Contactar | 3 |
| Estadísticas | 3 |
| Revista | 3 |

Tabla 1. Etiquetas de categoría más frecuentes

En febrero de 2021 se analizaron completamente los menús en todos sus niveles. A partir del nivel

y el número de orden se generó un código sistemático que permitía ordenar jerárquicamente los menús como una tabla de contenidos o un índice sistemático. Se incluyó un campo de categoría obsoleta, porque en este nuevo análisis se comprobó que habían desaparecido categorías anteriores por reorganización de los sitios web en cuatro casos: 3 en León y una en Oviedo (4 de 966, el 0,41 %). Una vista para la asociación de Madrid se puede apreciar a la derecha de la Figura 5.

A efectos del análisis, las etiquetas de las categorías se conceptualizaron como palabras clave. Una vez eliminados los duplicados, se convirtieron en una lista de descriptores libres extraídos. Para ello, se realizaron sendas relaciones de autounión sobre la tabla de descriptores libres: la primera para detectar los duplicados, con el campo de clave principal único; y la segunda con la palabra clave que etiqueta cada categoría. Como resultado, se detectaron 706 términos únicos y 260 duplicados. En la Tabla I se presentan las etiquetas de categoría más frecuentes.

3.4. Normalización de los conceptos (categorías) y selección de los términos preferentes (etiquetas)

Incluso sin revisar la lista completa, en la Tabla I se pueden apreciar los típicos problemas de control de vocabulario que comprometen la eficacia de la búsqueda y la navegación con silencios y ruido (Lancaster, 1986). En el caso de la sinonimia, coocurren etiquetas como “Asociación” y “La asociación” o “Consejos” y “Consejos prácticos” como sinónimos; o “Fotos” y “Galería de fotos” como cuasisinónimos. Por el contrario, aparecen términos polisémicos, entre los que destaca el conjunto de términos referidos al Camino que pueden denotar, de lo más general a lo más concreto: el Camino como experiencia, los Caminos de Santiago en su conjunto, el Camino Francés o cualquiera de los otros Caminos, sus distintas variantes, o la zona e incluso etapas en la que se centra cada asociación (véase la sección 1 del Apéndice 1). Para hacerse una idea del problema, la cadena “[C]c]amino” ocurre en 106 de los 964 términos, esto es, en el 10,96 %.

Por ello, se procedió a controlar el vocabulario, reduciendo los términos —etiquetas, en la terminología de los menús y las taxonomías web— a los conceptos —categorías, en dicha terminología—, y seleccionando uno de sus términos como preferente. Para ello, se crearon dos nuevas tablas en la base de datos: una para los futuros términos controlados, y otra para servir de enlace entre la tabla de términos libres y los controlados, con sendos campos clave y uno de observaciones. Como resultado se obtuvieron las 242

categorías que se pueden apreciar, ya sistematizadas, en el Apéndice 1. Sin duda, se trata de una fase laboriosa, que requiere un cuidadoso análisis intelectual por un experto o un conjunto de expertos humanos que tengan la voluntad de alcanzar un resultado sólido y exhaustivo.

Una cuestión importante que surgió en esta etapa fue la necesidad de seleccionar términos preferentes diferentes para objetivos distintos. Por un lado, el control de vocabulario exige términos preferentes únicos, que solo puedan referirse a un determinado concepto, de manera que no haya ambigüedad en su interpretación, ni, por tanto, en su uso en la indización y la búsqueda. Por el otro, la usabilidad web complica esta perspectiva: por un lado, cada nodo debe ser perfectamente distinguible, de manera semejante a lo que ocurre en

el control de vocabulario; por el otro, el reducido espacio del que se suele disponer para menús obliga a buscar expresiones cortas y muy sintéticas que tienden a ser polisémicas.

La solución (Figura 6) ha sido planificar tres tipos de “términos preferentes” o etiquetas de concepto: uno que sirve para describir inequívocamente el concepto, incluso a efectos de crear un título para una página web, denominado “Término controlado local”; otro “Simplificado” para su uso en menús; y un tercero, que intenta ser conforme con las recomendaciones de la norma ISO 25964 en cuanto a la forma de los términos, al que se ha llamado “Término ISO”. En todo momento, se mantiene una columna que indica cuántas ocurrencias de ese concepto se pueden encontrar en los menús de los sitios web estudiados.

| Nivel | Clave | Clase | Término controlado local | Término simplificado | Término ISO | Observaciones | Nº sitios | Término controlado común |
|-------|----------|-------|--|----------------------|-------------------------|--|-----------|--|
| 1 | 7C35EC50 | 0 | [Instrumentales] | | | | | 0 [Instrumentales] |
| 2 | 7637DEE0 | 0.1 | Búsquedas | | | Buscar | 4 | 0.1 Búsquedas |
| 2 | 7AF4EEC | 0.2 | Idiomas | | | | 6 | 0.2 Idiomas |
| 1 | A46F0EB5 | 1 | Camino de Santiago | El Camino | Camino de Santiago | En general. El Camino de Santiago, El Camino | 13 | 1 Camino de Santiago |
| 2 | 25872139 | 1.0 | Camino de Santiago (otros) | | | Página dedicada a los otros caminos de Santiago, salvo los que cuida esa | 1 | 1.0 Caminos de Santiago (otros) |
| 2 | 9F6855A7 | 1.1 | Camino de Santiago en España | Caminos | Camino de Santiago | Enfoque general que resalta su diversidad y alternativas. Sin: Caminos | 8 | 1.1 Caminos de Santiago en España |
| 2 | 6DF558C8 | 1.2 | Camino de Santiago europeos | | | | 2 | 1.2 Caminos de Santiago europeos |
| 2 | F5B26AA4 | 1.3 | [Contexto cultural del Camino de Santiago] | | | Scriptorium | | 1.3 [Contexto cultural del Camino de Santiago] |
| 2 | 421FCF6C | 1.4 | [Puntos de interés turístico] | | | Visitar en, [¿qué ver?] | | 1.4 [Puntos de interés turístico] |
| 2 | 3DEFDC7 | 1.5 | Instituciones jacobeanas | | | | 1 | 1.5 Instituciones jacobeanas |
| 2 | CC68935E | 1.6 | Personalidades jacobeanas | | | TR Testimonios | 1 | 1.6 Personalidades jacobeanas |
| 1 | 1FDD47 | 2 | Asociaciones jacobeanas | Asociación | información asociativa | Información sobre la asociación (asociación, La, nuestra), ¿quiénes somos? (≈6) | 38 | 2 Asociaciones jacobeanas |
| 2 | B37BC602 | 2.1 | Datos de contacto | Contacto | datos de contacto | Contactar UF sede, sedes, localización, situación. Hay que contar uno más, el | 29 | 2.1 Datos de contacto |
| 2 | 8ACCF9F | 2.2 | [Régimen interno] | | | [Funcionamiento interno] | | 2.2 [Régimen interno] |
| 2 | 020709B7 | 2.3 | [Relaciones entre asociaciones] | | | | | 2.3 [Relaciones entre asociaciones] |
| 2 | 096A7B63 | 2.4 | Servicios | Actividades | actividades asociativas | Actividades de la asociación. En dos ocasiones sinónimo eventos (cádiz), UF | 30 | 2.4 Servicios |
| 1 | A0596047 | 3 | Peregrinación | | | Manual básico del peregrino en http://deguadalajarasantiago.blogspot.com/ | 4 | 3 Peregrinación |
| 2 | B0E44650 | 3.0 | Consejos | Consejos | consejos de | consejos prácticos, sugerencias, consejos útiles. Manual básico del peregrino en | 22 | 3.0 Consejos |
| 2 | D8A15CC | 3.1 | Transporte | | <peregrinación según | [Modalidades de peregrinar] | 1 | 3.1 Transporte |
| 2 | 57131131 | 3.2 | Equipamiento | | | material y equipo TR bicicleta | 2 | 3.2 Equipamiento |
| 2 | 1D2105DE | 3.3 | Salud | | | medicina peregrina, remedios | 1 | 3.3 Salud |
| 2 | 006ABDD | 3.4 | Acreditación del peregrino | Credencial | | La Credencial y la Compostela, en la de León | 1 | 3.4 Acreditación del peregrino |
| 2 | B10D9D53 | 3.5 | Alojamiento | Alojamiento | alojamiento | | 3 | 3.5 Alojamiento |
| 2 | E8E0B8D0 | 3.6 | [Fases de la peregrinación] | | | Prepara tu camino de Santiago (Jaen), <Peregrinación según momentos> | 1 | 3.6 [Fases de la peregrinación] |
| 2 | FC2AD3C | 3.7 | [Espiritualidad] | | | Filosofía | 7 | 3.7 [Espiritualidad] |
| 2 | B8B7C6B | 3.8 | Ecología | | | | 1 | 3.8 Ecología |
| 2 | D7A9716B | 3.9 | [Documentación jacobea] | | | Scriptorium en caminosantiago.com TR biblioteca | 1 | 3.9 [Documentación jacobea] |

Figura 6. Vista de la tabla de términos controlados y taxonomía

Para crear estos términos preferentes ha sido necesario en algunas ocasiones asignar términos, porque los extraídos de las páginas resultaban ambiguos en el conjunto de la taxonomía de enlace, aunque fueran fácilmente interpretables en su contexto local, esto es, en su sitio web concreto. Estos términos se han denotado incluyéndolos entre corchetes en el campo de “Término controlado local”. Han sido 18 casos completos de 242 (un 7,44 %): uno de primer nivel (0 [Instrumentales]), 7 de segundo nivel, 8 de tercero y

4 de cuarto. También se han completado tres categorías ambiguas añadiéndoles modificadores: “Etapas [del Camino del Norte]”, “Mapas [del Camino del Norte]” y “Esculturas [jacobeanas]”, todos ellos de nivel inferior. Como consecuencia, se puede comprobar que las etiquetas originales son muy específicas en la gran mayoría de los casos (un 92,66 %, el conjunto complementario) (Tabla II, en la página siguiente).

Resulta notable por su enorme variabilidad terminológica el caso del concepto que hemos denominado “Espiritualidad”. Para referirse a él se emplean siete expresiones tan diversas como “Miscelánea”, “Filosofía”, “Lo antiguo es nuevo”, “La clave del Camino”, “Requisito inicial”, “Naturaleza en el Camino” o “Silencios para pensar”. Muchas personas que hacen el Camino o se comprometen con él no lo hacen desde una espiritualidad cristiana, y estas expresiones sirven para integrar esa experiencia espiritual más amplia.

| <i>Términos asignados</i> | <i>N</i> | <i>Causa</i> |
|--|----------|---|
| 0 [Instrumentales] | 0 | Clasificar herramientas |
| 1.3 [Contexto cultural del Camino de Santiago] | 0 | Generar clase genérica |
| 1.3.3 [Celebraciones religiosas] | 0 | Generar clase genérica |
| 1.4 [Puntos de interés turístico] | 0 | Generar clase genérica |
| 1.4.1 [Poblaciones] | 1 | Generar clase genérica |
| 1.4.1.3 [Carrión de los Condes] | 0 | Generar clase genérica |
| 1.4.1.4 [Madrid] | 0 | Generar clase genérica |
| 2.1.0 [Identidad gráfica] | 2 | |
| 2.2 [Régimen interno] | 0 | Generar clase genérica |
| 2.3 [Relaciones entre asociaciones] | 0 | Generar clase genérica |
| 2.6.3 [Actividades culturales] | 0 | Generar clase genérica |
| 2.6.4 [Promoción del Camino] | 2 | Sustituir términos muy específicos: Homologación GR-239, SOS Camino |
| 2.6.5 [Celebraciones] | 0 | Generar clase genérica |
| 3.6 [Fases de la peregrinación] | 1 | Sustituir expresión verbal: “Prepara tu camino de Santiago (Jaen)” |
| 3.6.2 [Durante el camino] | 1 | Completar “Antes de partir” |
| 3.7 [Espiritualidad] | 7 | Gran variabilidad. Véase la explicación en texto |
| 3.9 [Documentación jacobea] | 1 | Sustituir término muy idiosincrásico: “Scriptorium” |
| 3.9.4 [Información cartográfica] | 0 | Generar clase genérica |

Tabla II. Términos asignados completamente

Otro problema interesante desde el punto de vista de la norma UNE-ISO 25964 (2014, 2016) lo constituye la prescripción de la forma canónica

en singular para incontables y conceptos abstractos, y plural para los concretos. Debido a la naturaleza de estas taxonomías, que se utilizan tanto para materias como autoridades, muchas etiquetas —que luego podrán funcionar también como descriptores tras una tesaurización—, se utilizan en singular y en plural, pues responden tanto a entidades únicas como concretas (exclusivas de esa asociación o de un lugar específico): por ejemplo, Camino/Caminos/caminos, Etapa/etapas, Hemeroteca/hemerotecas, Albergue/albergues, etc. En realidad, esta casuística se puede abordar en el marco de la norma poniendo las entidades singulares en mayúsculas y en su forma completa, estableciéndolas claramente como nombres propios, y relacionándolas con su concepto común contable en plural.

3.5. Reclasificación (taxonomización) del vocabulario controlado

Como se ha señalado al principio de la sección, aunque muchas relaciones jerárquicas se derivan necesariamente de relaciones empíricas entre pares de términos, su simple recurrencia no produce una taxonomización completa necesaria, para la que pueden existir diferentes alternativas.

Para construir una de la forma menos arbitraria posible, se buscó inspiración cinco modelos: las *Septem Circumstantiae* aristotélicas y tomistas (Tomás de Aquino, ed. 1859, Parte I-IIae - Cuestión 7, p. 95 ss.), la teoría general de facetas de Ranganathan (1937) y las preguntas de Laswell (1948, p. 117), con orientación a un dominio específico (Hjørland y Albrechtsen, 1995; Hjørland, 2017), lo que conllevó alguna innovación en las facetas de alto nivel resultantes, en particular la “0 Instrumentales” (Tabla III).

Entre las heurísticas que se han demostrado efectivas destaca la de estudiar en primer lugar las relaciones jerárquicas de los sitios más grandes, pues ofrecen una granularidad y complejidad mayor, que suele incluir a la de los sitios más pequeños. Sin embargo, este no puede ser el único criterio, pues cada sitio enfoca su labor a sus usuarios de forma diferente (Laswell, 1948), presentando ontologías que, aunque comparten muchos aspectos, difieren en otros (García Marco y López del Ramo, 2022).

La versión 1 de la taxonomía consiguiente, que es uno de los principales resultados de la investigación, se presenta completa en el apéndice 1. Para aspectos muy generales dentro de una categoría se ha utilizado el número “0”, que no está presente en todas las categorías, sino solo en los sitios en los que ha resultado conveniente, aunque este aspecto queda pendiente de revisión en

subsiguientes versiones. Cada nivel puede utilizar los dígitos 0 a Z, aunque en pocos casos ha sido necesario utilizar letras, notablemente en la sección 1.1 Caminos de Santiago en España, la más prolija.

La taxonomía se compone de cuatro ramas: Instrumentales, Camino de Santiago, Asociaciones Jacobeas y Peregrinación. En la Tabla III se puede consultar su tamaño y su correspondencia con algunos de los esquemas conceptuales de carácter general que se han utilizado como referencia. Las dos últimas coinciden significativamente en tamaño. Combinando las ramas principales, interpretadas como facetas complejas, resulta una definición de propósito clara: las *Asociaciones* facilitan la *peregrinación* en el *Camino*, publicando en la web su información con las *herramientas* necesarias para su localización y consulta. En esta definición se puede ver claramente cómo se ha aplicado la perspectiva del análisis de dominio; más adelante, en las conclusiones se discute el problema de conciliar la perspectiva facetada dentro de objetos informacionales que se pretenden taxonomías.

La sección instrumental es muy reducida, pero no era posible incluirla con lógica en ninguna de las otras ramas. Ciertamente, pertenece a la categoría del medio de comunicación, la web, que ofrece menús de carácter universal para búsquedas, idiomas y otras cuestiones generales que en los webs estudiados no estaban desarrolladas, más allá de “Contacto” e “Inicio” que solían corresponder a la propia asociación o al camino. Constituye una faceta muy interesante para explorar en otros ámbitos con perspectiva comparativa.

La rama más amplia es la dedicada al Camino propiamente dicho. Se divide en siete secciones.

En la 0, de carácter instrumental, están las menciones desde una asociación a otros Caminos que no son de su inmediata preocupación. La sección más grande, con 76 categorías, es la dedicada a los Caminos de Santiago en España, con sus tramos y etapas. Salvando la pequeña sección sin subclases de los Caminos europeos, el resto está dedicado a los aspectos contextuales —si se quiere diacrónicos o históricos— del Camino: historia, puntos de interés, instituciones y personas. Ciertamente, hubiera sido posible separar esta sección en una faceta o rama temporal, pero no se consideró adecuado ponerla al mismo nivel jerárquico que las otras dos ramas —las asociaciones y la peregrinación— a efectos de la definición de propósito resultante. Esta cuestión queda, no obstante, sujeta a revisión en el futuro.

La rama dedicada a las asociaciones proporciona información sobre sus datos de contacto, régimen interno (registro, estatutos, gobierno, historia, asociarse, etc.), servicios (agenda, atención al peregrino, actividades culturales, promoción del Camino, celebraciones, descuentos, publicaciones y biblioteca).

Por fin, la cuarta rama o faceta está dedicada a la peregrinación en sí misma: consejos generales, transporte equipamiento, salud, acreditación del peregrino, alojamiento, fases de la peregrinación, espiritualidad, ecología y documentación jacobea.

En la Figura 6 de la sección anterior se puede ver cómo, a efectos de completar las jerarquías y definir criterios de inclusión claros para cada taxón, fue necesario insertar 17 categorías asignadas —esto es, nuevas, no extraídas de las páginas web analizadas—. La asignación se ha denotado conjuntamente mediante su inclusión entre corchetes y la ausencia de sitios correspondientes en la columna de número de sitios.

| Rama | Nº | Ranganathan | Septem Circumstantiae | Laswell |
|--------------------|-----|----------------|--|---------------------------|
| Instrumentales | 3 | — | — | In which channel |
| Camino de Santiago | 106 | Space, Time | Ubi, Quid, Quando, (Cur) | Says what |
| Asociaciones | 65 | Personality | Quis | Who, With what effect |
| Peregrinación | 65 | Energy, Matter | Quem ad modum, Quibus adminiculis, Cur | To whom, With what effect |

Tabla III. Ramas de la taxonomía, importancia y justificación teórica

3.6. Tesauroización

Aunque esta fase queda ya fuera del objetivo de esta investigación, que se centra en la construcción de una taxonomía de enlace —categorías, etiquetas y relación jerárquica completa—, durante el proceso de análisis se fueron recogiendo

variantes terminológicas (relaciones de equivalencia o términos equivalentes) y también relaciones no jerárquicas entre términos (relaciones asociativas o términos relacionados). De esta forma, resulta fácil convertir la taxonomía en un tesoro a efectos de expresarla como tal, para

usos como la construcción de sitios web combinando menús y etiquetas, para facilitar la navegación; la representación en lenguajes de web semántica; el reenvío automático desde expresiones sinónimas o cuasisinónimas en menús y buscadores internos; o su uso en bibliotecas y centros de documentación especializados de las asociaciones, etc. En una fase siguiente de esta investigación se espera producir un tesoro completo y poder informar de los problemas encontrados y las soluciones adoptadas, aunque los criterios generales ya han sido establecidos en la subsecciones inmediatamente anteriores (control de vocabulario y jerarquización).

3.7. Ontologización: transformación en una ontología

Este paso no se ha implementado, pero se describe a efectos de mostrar todo el camino de diseño de una metodología general de elaboración de taxonomías de enlace que tenga cabida dentro de una perspectiva sólida y evolutiva de los sistemas de organización del conocimiento. Una vez que se ha derivado una taxonomía a partir de conceptos con soporte informacional real (esto es, fijados en documentos, bases de datos, etc.) es posible pasar a la fase más avanzada (Costin, Eastman e Issa, 2017) de descripción formal de propiedades, restricciones y relaciones no jerárquicas cuidadosamente especificadas que constituye el aspecto nuclear de la definición de ontología (Grueber, 1993). Además, dicha ontologización puede favorecer mapeos que soporten procesos inferenciales más avanzados (Guarino, 1997), incluso para abordar el nivel de datos con todas las posibilidades de enlace que ofrece, no sólo el de sitios y páginas web. Como es sabido, la adición de propiedades, restricciones y relaciones avanzadas requiere la formalización de la ontología con ayuda de un lenguaje de representación de ontologías, entre los que cabe destacar OWL (W3C, 2012) por su carácter abierto y su uso en la web semántica.

4. Conclusiones y recomendaciones

Además de intentar aportar al problema práctico de la información y documentación sobre el Camino de Santiago y analógicamente otros casos parecidos, durante esta investigación ha sido posible reflexionar en profundidad sobre aspectos teóricos, metodológicos y aplicativos del mundo de las taxonomías, la organización del conocimiento y la usabilidad web, principalmente. Son muchas las lecciones que se han obtenido y que se intentan sintetizar a continuación.

En el plano teórico, se ha intentado avanzar en la incorporación de las taxonomías web al dominio

de la organización del conocimiento, al que pertenecen por tradición intelectual y por su propio objeto y orientación: el uso de la clasificación conceptual como herramienta para mejorar el conocimiento en todas sus fases y dimensiones —comprensión, memoria y comunicación—. La importancia del concepto de taxonomía a lo largo de la historia intelectual y su aparición recurrente en diversos ámbitos de aplicación —y recientemente en el universo digital— es coherente con el papel central de los procesos clasificatorios en cualquier actividad mediada por el pensamiento (García Marco y Esteban, 1993).

De forma más específica, una taxonomía se puede definir como un conjunto sistematizado (sistema) de clases (categorías) designadas con un término (etiqueta) ordenadas secuencialmente, ya sea con un criterio de ordenación reconocible —alfabético, cronológico, conceptual...— o no; y opcionalmente anidadas jerárquicamente en varios niveles, formados por subclases de las clases a las que pertenecen. Se trata de un modelo plena y fácilmente expresable con el álgebra de conjuntos, la lógica de predicados, el álgebra de Boole y la teoría de grafos por lo que resulta muy adecuado no solo para describir el dominio de un sistema informático —y de un sitio web en particular—, sino también para realizar operaciones automatizadas y deterministas de recuperación de información, lo cual se hace en el caso de las taxonomías para usuarios normalmente a través de procedimientos de visualización (browsing) o navegación en mapas y menús.

¿Es adecuado el uso término taxonomía tal y como se está utilizando en el ámbito de la arquitectura de la información web, esto es, para menús, mapas y sistemas de etiquetado social? Cabe concluir que sí, que menús y mapas son taxonomías, pues pretenden clasificar la información de los sitios de forma completa, si se quiere exhaustiva, en general de forma jerárquica y en clases en principio disjuntas. Es cierto que esto no sucede siempre de forma estricta, pues pueden contener reenvíos internos, especialmente en las categorías más específicas, generando polijerarquías; pero eso no perjudica a la experiencia de usuario, sino que la facilita.

Ciertamente, al intentar conciliar la perspectiva taxonómica pura con el análisis de facetas, como se ha hecho en este artículo, surge una situación aporética desde el punto de vista estricto de la teoría de la clasificación. En la cladística sólo cabe la pertenencia a una única clase: no existen perspectivas alternativas de clasificación de los objetos, que es lo que se postula en la teoría de facetas. Sin embargo, en las taxonomías informacionales —muchas de ellas relacionadas no con géneros y especies, sino con conceptos de

muy diverso tipo y ciertamente facetados— la mejor manera de construir los árboles sin que se produzcan incoherencias de pertenencia es precisamente usar el método de facetas. El trabajo presentado aquí puede servir una vez más para mostrar la utilidad de esta estrategia frente a las clasificaciones disciplinares, que, por otra parte, lejos de ser despreciadas, deben ser consideradas una faceta —o varias, si hay diversos paradigmas— en sí mismas.

Por fin, más allá de las variables lógicas y terminológicas, la dimensión teórica de la investigación sobre taxonomías tiene que abordar el problema social de su uso. La variedad de taxonomías encontradas en un conjunto limitado de sitios web ha podido mostrar como su diseño constituye una actividad social. A pesar de que la organización de los sitios se intenta realizar en función de los usuarios —como exige el paradigma de la arquitectura de la información—, investigaciones como ésta muestran que sus sistemas de organización del conocimiento son más bien en buena parte el resultado del consenso de grupos de usuarios cualificados y del liderazgo de las personas responsables de mantener y organizar los sitios web. Para comprender y optimizar la organización de los sitios web es fundamental trabajar con esos tres grupos de agentes: usuarios, grupos implicados, y líderes y administradores de sitios; cualquier desequilibrio puede comprometer la efectividad de los sistemas.

En las dimensiones metodológica y aplicada, se ha desarrollado con éxito un procedimiento para transformar un conjunto de taxonomías web empíricas, en este caso los menús de los sitios web de las asociaciones españolas federadas del Camino de Santiago, en una taxonomía que es capaz de interoperar con todas ellas, susceptible de expresarse en RDF u OWL para soportar servicios de enlace, acceso y búsqueda distribuidos, entre otros. El proceso operativo abarca la posibilidad de generar un tesoro y posteriormente servir de base para una ontología, partiendo del hecho de que muchas veces las taxonomías constituyen su esqueleto y primer paso de desarrollo (Costin, Eastman e Issa, 2017).

Desde el punto de vista operativo, hay que reconocer que este tipo de trabajos requieren mucho tiempo y son difícilmente rentables para el diseño de sitios web fuera de contextos de investigación o cuando existen aplicaciones muy claras, como en el caso de los agregadores con alto valor económico, notablemente los de comercio electrónico o ciertos casos en empresas, administraciones y organizaciones no gubernamentales de gran tamaño. Por ello, no es raro que haya tan poca normalización en los menús y mapas de sitios web con contenidos muy semejantes, como

ha mostrado esta investigación para un caso concreto. En parte este hecho es comprensible porque la diversidad de taxonomías empíricas surge de diferentes enfoques y metáforas organizativas que buscan atender a usuarios con visiones distintas en cuanto a objetos de atención y prioridades. Su dimensión comunicacional incluye también un componente competitivo por la atención de los usuarios, un factor complejo que incluye un difícil y creativo equilibrio entre predictibilidad —para disminuir el gasto energético de los usuarios— y dosis originalidad —para captar y sostener su atención—. Por fin, no se debe olvidar el componente dinámico y evolutivo inherente a las ecologías informacionales, y a cualquier sistema ecológico en general.

Por otra parte, apesar de la importancia del vector originalidad (mayor en unos tipos de sitios que en otros), lo cierto es que las diferencias en organización de la información en los sitios web complican enormemente la navegación a los usuarios, su aprovechamiento y, por consiguiente, comprometen los propios objetivos buscados con la presencia en Internet. Esta situación se incrementa notablemente en el contexto de las nuevas versiones de la Internet, especialmente en el caso de la web semántica, cuyo objetivo es compartir los datos presentes en las páginas web de forma que sean identificables por agentes automáticos. En este escenario, los trabajos de análisis de las ecologías informacionales web para detectar, organizar y normalizar los mapas web de los diferentes sitios en estructuras de mínimo común múltiplo constituye un trabajo absolutamente necesario de cara al futuro, cueste lo que cueste. De hecho, se vislumbra como una nueva actividad en el campo de la organización del conocimiento que creemos tendrá un desarrollo muy importante.

De cara a la continuidad de esta investigación, esperamos avanzar en diferentes líneas de trabajo: intensivas, extensivas y comparativas. En la dimensión intensiva, queda pendiente la tesaurización y ontologización de la taxonomía, y su implementación en soluciones concretas que sirvan a los intereses de los amigos del Camino de Santiago. También queda mucho trabajo por delante comparando y comprendiendo la naturaleza de las organizaciones empíricas de los sitios web en función de numerosos factores —objetivos comunicativos, formación y características de los gestores y usuarios, modelos del dominio, etc.— de las que se puedan derivar lecciones de interés general.

En la dimensión extensiva, cabe aplicar esta metodología a otros campos diferentes, pues se puede utilizar con cualquier dominio en Internet que tenga detrás una comunidad de interés clara

y una ecología de servicios web que colaboren, compitan y coexistan para darle satisfacción. Finalmente, en el aspecto comparativo es necesario seguir trabajando en deslindar los diferentes tipos de taxonomías (científicas, corporativas, web...), así como en profundizar en su relación con otros sistemas de organización del conocimiento.

Además, es necesario contrastar estas técnicas intelectuales —costosas pero con resultados muy sólidos— con otras automáticas basadas en la aplicación de métodos estadísticos avanzados como el análisis de clústeres, la clasificación automática o la coocurrencia de términos y otras —que resultan muy baratas, pero que ofrecen resultados muy fluidos—. Finalmente, como señala Smith (2020) son necesarios estudios que evalúen la eficacia recuperadora y la usabilidad de los mapeos basados o inspirados en ISO 25964-2 para demostrar la manera en que pueden añadir valor a los procesos de búsqueda, especialmente en la búsqueda interdisciplinar; como ya documentaban Wang, Isaac, Schlobach, van der Meij, y Schopman (2012) en el campo que nos ocupa del patrimonio cultural.

Agradecimientos

Este artículo publica resultados de investigación financiados con el proyecto CSO2015-65448-R (MINECO/FEDER). Se agradece sinceramente el apoyo a las instituciones financiadoras sin cuya generosidad esta investigación no hubiera sido posible.

Referencias

- Aanen, S. S.; Vandic, D.; Frasinca, F. (2015). Automated product taxonomy mapping in an e-commerce environment. // *Expert Systems with Applications*. 42:3 (Feb) 1298-1313.
- Aristóteles (1992). *Investigación sobre los animales*. Madrid: Gredos.
- Avesani, P.; Giunchiglia, F.; Yatskevich, M. (2005). A large scale taxonomy mapping evaluation. // Gil, Y.; Motta, E.; Benjamins, V. R.; Musen, M. A. (eds.). *Semantic Web - ISWC 2005, Proceedings*. Berlin, Springer-Verlag Berlin. 3729. 67-81.
- Costin, A. M.; Eastman, C.; Issa, R. R. A. (2017). The Need for Taxonomies in the Ontological Approach for Interoperability of Heterogeneous Information Models. // *ASCE International Workshop on Computing in Civil Engineering (IWCCE)*, Univ Washington. Seattle, WA: Amer. Soc. Civil Engineers.
- Dahlberg, Ingetraut (2006). Knowledge Organization: A New Science?. // *Knowledge Organization*. 33:1, 11-19.
- Dahlberg, Ingetraut (2014). What is Knowledge Organization?. // *Knowledge Organization*. 41:1, 85-91.
- Dextre Clarke, Stella G. (2011). In Pursuit of Interoperability: Can We Standardize Mapping Types?. // Boteram, F.; Goedert, W.; Hubrich, J. (eds.). *Concepts in Context: Cologne Conference on Interoperability and Semantics in Knowledge Organization*; held 19-20 July 2010; Cologne, Germany. Würzburg: Ergon Verlag.
- Dextre Clarke, Stella G. (2012). ISO 25964: A Standard in Support of KOS Interoperability. // Gilchrist, Alan; Vernau, Judi (ed.). *Facets of Knowledge Organization*; 4-5 July 2011; London. London: Emerald. 129-134.
- Dextre Clarke, Stella G.; Zeng, Marcia Lei. (2012). From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. // *Information Standards Quarterly*. 24:1, 20-26.
- Drupal (2020). Legacy Drupal release history. <https://www.drupal.org/docs/understanding-drupal/understanding-drupal-version-numbers/legacy-drupal-release-history>.
- Federación Española de Asociaciones de Amigos del Camino de Santiago (s. f.). Directorio de Asociaciones de Amigos del Camino de Santiago. // *El Camino de Santiago*: página web de la Federación Española. <https://www.caminosantiago.org/cpperegrino/asociaciones/asociaciones.asp>
- García Marco, Francisco Javier (2016). The Interaction between the Systematic and Alphabetical Approaches to Knowledge Organization and Its Subjacent Mechanisms: a Long-term Primary Wave?. // Guimarães, José Augusto Chaves; Milani, Suellen Oliveira; Dodebei, Vera (eds.). *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society: Proceedings of the Fourteenth International ISKO Conference 27-29 September 2016 Rio de Janeiro, Brazil*. Würzburg: Ergon-Verlag GmbH, 2016. p. 105-115. ISBN 978-3-95650-221-7.
- García Marco, Francisco Javier (2015). Apropiación y uso de la información y del conocimiento en el contexto de las ecologías informacionales. // González de Gómez, María Nélica; Rabello, Rodrigo (org.). *Informação: agentes e intermediação*. Brasília: IBICT, 2015. 313-353. <http://livroaberto.ibict.br/handle/123456789/1068>.
- García-Marco, Francisco Javier (2021). La organización del conocimiento al servicio de la interconexión entre las instituciones de la memoria y el turismo cultural. // *Scire*. 27:1, 55-70. doi.org/10.54886/scire.v27i1.4780
- García-Marco, Francisco Javier; Esteban Navarro, Miguel Angel (1993). On some Contributions of the Cognitive Sciences and Epistemology to a Theory of Classification. // *Knowledge Organization* 20:3, 126-132. <https://doi.org/10.5771/0943-7444-1993-3-126>.
- García-Marco, Francisco Javier; Galindo, Fernando; Lasala, Pilar; López del Ramo, Joaquín (2020). Advancing the Interoperability of the GLAM+ and Cultural Tourism Sectors through KOS. // *Advances in Knowledge Organization*. 17, 151-160. doi.org/10.5771/9783956507762-151.
- García-Marco, Francisco Javier; López del Ramo, Joaquín (2022). Interoperability and Standardization of Web Navigation Taxonomies: Clarifying some Operational Concepts and Developing a Tentative Basic Vocabulary. // Lykke, Marianne; Svarre, Tanja; Haynes, David; Skov, Mette; Thellefsen, Martin; Martínez-Ávila, Daniel (eds.). *Knowledge Organization across Disciplines, Domains, Services and Technologies: Proceedings of the Seventeenth International ISKO Conference 6-8 July 2022 Aalborg, Denmark*. 2022. 73-86. eISBN 978-3-95650-956-8. Baden-Baden: Nomos. <https://doi.org/10.5771/9783956509568>
- Garshol, Lars Marius (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. // *Journal of Information Science*. 30:4 378- 391.
- Gilchrist, Alan (2001). Corporate taxonomies: report on a survey of current practice. // *Online information review*. 25, 94-103. doi.org: // 10.1108/14684520110390033
- Gilchrist, Alan (2003a). Taxonomies and Information Architecture. // *Scire: representación y organización del conocimiento*. 9:1, 37-46. <https://www.iberid.eu/ojs/index.php/scire/article/view/1453>
- Gilchrist, Alan (2003b). Thesauri, taxonomies and ontologies: an etymological note. // *Journal of Documentation*, 59:1, 7-18. doi.org/10.1108/00220410310457984.

- Gilchrist, Alan; Kibby, Peter (2000). *Taxonomies for business: access and connectivity in a wired world*. London: TFPL.
- Gnoli, Claudio (2020). Genealogical classification. // *Encyclopedia of Knowledge Organization*. Version 1.3. <https://www.isko.org/cyclo/genealogical>.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. // *Knowledge Acquisition*. 5:2, 199-220.
- Guarino, Nicola (1997). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. // *International Summer School on Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*. New York: Springer-Verlag. 139-170.
- Hjørland, Birger (2017). Domain analysis. // *Knowledge Organization*. 44: 6, 436-464. https://www.isko.org/cyclo/domain_analysis
- Hjørland, Birger; Albrechtsen, Hanne (1995). Toward a new horizon in information science: Domain analysis. // *Journal of the American Society for Information Science*. 46: 6, 400-425.
- Hjørland, Birger (2016). Knowledge organization. // *Knowledge Organization*. 43:6: 475-84. También en: Hjørland, Birger, ed. *ISKO Encyclopedia of Knowledge Organization*. http://www.isko.org/cyclo/knowledge_organization
- ISO (2013). ISO 25964-2:2013. Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. Geneva: ISO.
- Jaenecke, P. (1997). Knowledge organization due theory to theory formation. // *Organización del Conocimiento en Sistemas de Información y Documentación 2 (1997)*: 39-55.
- Jung, J. J. (2006). Taxonomy alignment for interoperability between heterogeneous digital libraries. // *9th International Conference on Asian Digital Libraries*, Kyoto, Japan. Springer-Verlag: Berlin. 274-282.
- Jung, J. J. (2008). Taxonomy alignment for interoperability between heterogeneous virtual organizations. // *Expert Systems with Applications*. 34:4 (May) 2721-2731.
- Lancaster, Frederick Wilfrid (1986). *Vocabulary control for information retrieval*. 2nd. Ed. Arlington, Va.: Info Resources Press.
- Lasswell, Harold (1948). *The Structure and Function of Communication in Society*. // Bryson, L. (ed.). *The Communication of Ideas*. New York: Institute for Religious and Social Studies.
- Lin, H. L.; Wang, Y. Z.; Jia, Y. T.; Xiong, J. H.; Zhang, P.; Cheng, X. Q. (2015). An Ensemble Matchers Based Rank Aggregation Method for Taxonomy Matching. // Cheng, R.; Cui, B.; Zhang, Z.; Cai, R.; Xu, J. Cham. (eds.). *Web Technologies and Applications*. Springer International Publishing Ag. 9313. 190-202.
- López del Ramo, Joaquín (2017). *Modelo de análisis descriptivo de arquitectura, grafismo y contenidos de sitios web*. Madrid: Registro de la Propiedad Intelectual. Ref. 16/2017/2670.
- López del Ramo, Joaquín; García Marco, Francisco Javier (2017). Información y diseño de los portales web municipales del Camino Francés: propuesta y aplicación de un análisis descriptivo. // Rubio Gil, M. A.; Sanagustín Fons, V. (eds.). *Turismo religioso. Camino europeo del Santo Grial y otras rutas culturales para el desarrollo*. Zaragoza: Universidad de Zaragoza. 62-77.
- López del Ramo, Joaquín; García Marco, Francisco Javier (2018). El Camino de Santiago en los sitios web de las Comunidades Autónomas: análisis del Contenido, orientación y encuadres temáticos predominantes. // *Revista General de Información y Documentación*. 28, 703-26.
- Mourino-García, M.; Pérez-Rodríguez, R.; Anido-Rifón, L.; Fernández-Iglesias, M. J.; Darriba-Bilbao, V. M. (2018). Cross-repository aggregation of educational resources. // *Computers & Education*. 117 (Feb) 31-49.
- Nederstigt, L.; Vandic, D.; Frasinca, F. (2012). An Automated Approach to Product Taxonomy Mapping in E-Commerce. // Casillas, J.; Martínez Lopez, F. J.; Corchado, J. M. (eds.). *Management Intelligent Systems*. Berlin, Springer-Verlag Berlin. 171. 111-120.
- Nederstigt, L.; Vandic, D.; Frasinca, F. (2016). A lexical approach for taxonomy mapping. // *Journal of Web Engineering*. 15:1-2 (Mar) 84-109.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. // *Organization Science*. 5, 14-37. <http://www.jstor.org/stable/2635068>
- Ranganathan, S. R. (1937). *Prolegomena to library classification*. Madras: The Madras library association; London: E. Goldston, Ltd.
- Rosnay, Joel de. (1979). *The Macroscopic*. New York: Harper & Row.
- Smith, Linda C. (2020). Interdisciplinary Searching as a Use Case for Vocabulary Mapping. // Lykke, Marianne; Svarre, Tanja; Skov, Mette; Martínez-Ávila, Daniel (eds.). *Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, 2020 Aalborg, Denmark*. O. Baden-Baden: Ergon-Verlag. 428-435. <https://doi.org/10.5771/9783956507762-428>.
- Tomás de Aquino (1562). *Summae theologiae*. T. III. Soler Fratres.
- UNE/ISO (2014). UNE/ISO 25964-1:2014: Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 1: Tesauros para la recuperación de la información. Madrid: Aenor.
- UNE/ISO (2016). UNE/ISO 25964-2:2016: Información y documentación. Tesauros e interoperabilidad con otros vocabularios. Parte 2: Interoperabilidad con otros vocabularios. Madrid: Aenor.
- W3C (2012). *OWL 2 Web Ontology Language: Document Overview (Second Edition)* / W3C OWL Working Group. W3C Recommendation 11 December 2012. <http://www.w3.org/TR/owl2-overview/>
- Wang, S. H.; Isaac, A.; Schlobach, S.; van der Meij, L.; Schopman, B. (2012). Instance-based Semantic Interoperability in the Cultural Heritage. // *Semantic Web*. 3:1 45-64.
- Wilson, T. D. (2002) The nonsense of 'knowledge management'. // *Information Research*. 8:1, paper no. 144. <http://InformationR.net/ir/8-1/paper144.html>
- Wordmap (2019). <https://www.wordmap.com/wordmap-enterprise-main-menu>
- WordPress (2007). *WordPress 2.3*. // *WordPress Blog*. 24 (September 2007). <https://wordpress.org/news/2007/09/wordpress-23/>.
- Zeng, Marcia Lei (2019). Interoperability. // *Knowledge Organization*. 46:2, 122-146. También en: Hjørland, Birger and Gnoli, Claudio eds. *ISKO Encyclopedia of Knowledge Organization*, <http://www.isko.org/cyclo/interoperability>
- Zhonghong, Wang; Chaudhry, Abdus Sattar; Khoo, Christopher (2006). Potential and Prospects of Taxonomies for Content Organization. // *Knowledge Organization*. 33:3, 160-169.

Apéndice I. Taxonomía de las asociaciones de la Federación Española del Asociaciones del Camino de Santiago

0 [Instrumentales]

0.1 Búsquedas

0.2 Idiomas

1 Camino de Santiago

1.0 Caminos de Santiago (otros)

1.1 Caminos de Santiago en España

1.1.0 Criterios de valoración

1.1.1 Camino Francés

1.1.1.0.1 Etapas del Camino Francés

1.1.1.0.2 Mapas del Camino Francés

1.1.1.1 Camino Francés en Aragón

1.1.1.1.1 Camino Francés en Aragón (Puerto de Palo)

1.1.1.2 Camino de Santiago en Estella

1.1.1.2.1 Etapas del Camino de Santiago en Estella

1.1.1.2.2 Puntos de interés en Estella

1.1.1.3 Camino de Santiago por Burgos

1.1.1.3.1 Camino de Santiago por Burgos (ciudad)

1.1.1.3.1.1 Iconografía jacobea en Burgos (ciudad)

1.1.1.4.1 Etapas del Camino de Santiago en el Bierzo

1.1.1.5 Camino de Santiago en Palencia

1.1.1.5.1 Etapas del Camino de Santiago en Palencia

1.1.1.6 Camino de Santiago por León

1.1.1.7 Camino de Santiago del Bierzo

1.1.1.8 Camino de Invierno

1.1.1.8.1 Etapas del Camino de Invierno

1.1.1.8.2 Mapas del Camino de Invierno

1.1.1.9 Camino de Fisterra-Muxía

1.1.1.9.1 Etapas del Camino de Fisterra-Muxía

1.1.1.9.2 Mapas del Camino de Fisterra-Muxía

1.1.2 Caminos de Santiago del Norte

1.1.2.1 Camino del Norte

1.1.2.1.0.1 Etapas [del Camino del Norte]

1.1.2.1.0.2 Mapas [del Camino del Norte]

1.1.2.1.1 Camino Vasco del Interior

1.1.2.1.2 Camino de Santiago por Álava

1.1.2.1.3 Camino de las Asturias

1.1.2.2 Camino Primitivo

1.1.2.2.1 Etapas del Camino Primitivo

1.1.2.2.2 Mapas del Camino Primitivo

1.1.2.3 Camino del Salvador

1.1.2.3.1 Etapas del Camino del Salvador

1.1.2.4 Camino de la Costa

1.1.2.4.1 Etapas del Camino de la Costa

1.1.3 Camino de la Plata

1.1.3.1 Etapas del Camino de la Plata

1.1.3.2 Mapas del Camino de la Plata

1.1.4 Camino Mozárabe

1.1.4.1 Camino de Málaga a Santiago

1.1.4.2 Camino de Málaga a Córdoba

1.1.4.2.1 Etapas de Málaga a Córdoba

1.1.4.2.1.1 Etapas de Málaga a Santiago

1.1.4.2.2 Etapas de Jaén a Mérida

1.1.4.2.2 Camino de Córdoba a Santiago

1.1.4.2.2.1 Etapas del Camino de Córdoba a Mérida

1.1.4.2.3 Camino Mozárabe en Orense

1.1.4.2.3.1 Etapas del Camino Mozárabe en Orense

1.1.5 Vía Augusta

1.1.5.1 Etapas de la Vía Augusta

1.1.6 Camino del Levante-Sureste

1.1.6.1 Camino del Levante

1.1.6.1.1 Etapas del Camino del Levante

1.1.6.1.1.1 Tracks del Camino del Levante

1.1.6.2 Camino del Sureste

1.1.6.3 Camino del Argar

1.1.6.3.1 Tracks del Camino del Argar

1.1.7 Camino de la Lana

1.1.8 Camino Jacobeo de Madrid

1.1.8.1 Etapas del Camino Jacobeo de Madrid

1.1.8.2 Plan Director del Camino Jacobeo de Madrid

1.1.9 Camino Portugués

1.1.9.1 Etapas del Camino Portugués

1.1.9.2 Mapas del Camino Portugués

1.1.A Camino Jacobeo del Ebro

1.1.B Camino Castellano-Aragonés

1.1.C Camino de Sant Jaume-Montserrat

1.1.D Camino Catalán-S. Juan de la Peña

1.1.E Camino Castellón - Bajo Aragón

1.1.F Camino Inglés

1.1.F.1 Etapas del Camino Inglés

1.1.F.2 Mapas del Camino Inglés

1.1.G Rutas marítimas a Santiago

1.2 Caminos de Santiago europeos

1.3 [Contexto cultural del Camino de Santiago]

1.3.1 Historia del Camino de Santiago

1.3.1.0 Apóstol Santiago

1.3.1.1 Orígenes del Camino de Santiago

1.3.1.1.1 Codex Calixtinus

1.3.1.1.2 Las Siete Partidas

1.3.2 Leyendas

1.3.3 [Celebraciones religiosas]

1.3.3.1 Año Santo Compostelano

1.3.3.2 Jubileo de la Santa Cruz

1.4 [Puntos de interés turístico]

1.4.1 [Poblaciones]

1.4.1.1 Tarrasa Jacobea

1.4.1.2 Oviedo

1.4.1.2.1 Catedral del Salvador

1.4.1.2.1.1 Cámara Santa

1.4.1.2.1.2 El Salvador

1.4.1.2.1.3 El monasterio de San Pelayo

1.4.1.3 [Carrión de los Condes]

1.4.1.3.1 Monasterio de San Zoilo

1.4.1.4 [Madrid]

1.4.1.4.1 Programa de Actuación Urbanística (PAU) de Las Tablas de Madrid

1.4.2 Monumentos

1.4.2.1 Templos jacobeos

- 1.4.2.2 Edificios prerrománicos
 - 1.5 *Instituciones jacobeanas*
 - 1.6 *Personalidades jacobeanas*
 - 1.6.1 Necrológicas
 - 2 Asociaciones jacobeanas**
 - 2.1 *Datos de contacto*
 - 2.1.0 Página de inicio
 - 2.1.1 [Identidad gráfica]
 - 2.2 *[Régimen interno]*
 - 2.2.0 Registro de la asociación
 - 2.2.1 Estatutos
 - 2.2.2 Asambleas
 - 2.2.3 Junta directiva
 - 2.2.4 Códigos éticos
 - 2.2.5 Información legal del sitio web
 - 2.2.6 Asociarse
 - 2.2.7 Patrocinadores
 - 2.2.8 Historia de la asociación
 - 2.2.9 Intranet
 - 2.2.9.1 Fundación de la asociación
 - 2.3 *[Relaciones entre asociaciones]*
 - 2.3.1 Federación Española de Asociaciones de Amigos del Camino de Santiago
 - 2.3.1.1 Directorio de asociaciones adheridas
 - 2.3.2 Federación Andaluza de Asociaciones de Amigos del Camino de Santiago
 - 2.3.3 Asociaciones jacobeanas relacionadas
 - 2.4 *Servicios*
 - 2.4.1 Agenda
 - 2.4.1.1 Actividades por años
 - 2.4.1.2 Memoria de actividades
 - 2.4.2 Atención al peregrino
 - 2.4.2.1 Oficinas del peregrino
 - 2.4.2.1.1 Hospitaleros
 - 2.4.2.1.1.2 Cursos de formación de hospitaleros
 - 2.4.2.1.2 Tienda
 - 2.4.3 [Actividades culturales]
 - 2.4.3.1 Conferencias
 - 2.4.3.2 Jornadas
 - 2.4.3.3 Visitas culturales
 - 2.4.3.4 Excursiones
 - 2.4.3.4.1 Excursiones específicas
 - 2.4.3.5 Concursos
 - 2.4.3.5.0 Bases legales de concursos
 - 2.4.3.5.1 Concursos fotográficos
 - 2.4.3.5.2 Concursos literarios
 - 2.4.3.5.3 Premios
 - 2.4.3.6 Coros
 - 2.4.4 [Promoción del Camino]
 - 2.4.4.1 Homologación de caminos
 - 2.4.4.2 Planes directores del Camino
 - 2.4.4.3 Señalización
 - 2.4.4.4 Esculturas [jacobeanas]
 - 2.4.4.5 Exposiciones jacobeanas
 - 2.4.4.6 Informes
 - 2.4.4.7 Protección del Camino
 - 2.4.5 [Celebraciones]
 - 2.4.5.1 Festividad de Santiago
 - 2.4.5.2 Día de los Caminos del Norte
 - 2.4.5.3 Misa del peregrino
 - 2.4.6 Descuentos para socios
 - 2.4.7 Publicaciones
 - 2.4.7.1 Revistas
 - 2.4.7.1.2 Boletín
 - 2.4.7.1.3 Revista "Peregrino"
 - 2.4.7.2 Guías del Camino de Santiago
 - 2.4.7.3 Noticias
 - 2.4.7.3.1 Blog
 - 2.4.7.3.2 Sala de prensa
 - 2.4.8 Biblioteca
 - 2.4.8.1 Hemeroteca
- 3 Peregrinación**
 - 3.0 *Consejos*
 - 3.0.1 Derechos de los peregrinos
 - 3.0.2 Teléfonos de interés
 - 3.1 *Transporte*
 - 3.1.1 Peregrinación a pie
 - 3.1.2 Peregrinación en bicicleta
 - 3.1.3 Peregrinación a caballo
 - 3.1.4 Transporte de apoyo
 - 3.2 *Equipamiento*
 - 3.2.1 Botiquín y aseo
 - 3.2.2 Documentación y enseres
 - 3.2.2.1 Licencias federativas
 - 3.2.3 Ropa
 - 3.2.4 Mochila
 - 3.2.5 Calzado
 - 3.3 *Salud*
 - 3.3.1 Alimentación
 - 3.3.2 Cuidado de los pies
 - 3.3.3 Fatiga
 - 3.3.4 Cuidados médicos
 - 3.3.4.1 Covid
 - 3.3.5 Seguridad
 - 3.4 *Acreditación del peregrino*
 - 3.4.1 Credencial del peregrino
 - 3.4.1.1 Sellos
 - 3.4.2 Compostela
 - 3.5 *Alojamiento*
 - 3.5.1 Albergues de peregrinos
 - 3.5.1.1 Albergue
 - 3.5.1.1.1 Localización del albergue
 - 3.5.2 Normas de uso (albergues)
 - 3.6 *[Fases de la peregrinación]*
 - 3.6.1 Preparación del camino
 - 3.6.1.1 Selección del camino
 - 3.6.1.2 Épocas para peregrinar
 - 3.6.1.3 Entrenamiento previo
 - 3.6.1.4 Distribución de las etapas

- 3.6.2 [Durante el camino]
 - 3.6.2.1 Técnicas de caminar
- 3.6.3 Después del Camino
- 3.7 [*Espiritualidad*]
 - 3.7.1 Identidad peregrina
 - 3.7.2 Símbolos
 - 3.7.2.1 Bordón
- 3.8 *Ecología*
 - 3.8.1 Life Stars
- 3.9 [*Documentación jacobea*]
 - 3.9.1 Bibliografía jacobea
 - 3.9.2 Testimonios jacobeos
 - 3.9.2.1 Citas jacobneas
 - 3.9.2.2 Libros de visitas
 - 3.9.2.3 Blogs
- 3.9.3 Fotos y vídeos
 - 3.9.3.1 Galerías de fotos
 - 3.9.3.2 Vídeos
 - 3.9.3.3 Películas
- 3.9.4 [Información cartográfica]
 - 3.9.4.1 Apps jacobneas
 - 3.9.4.2 Mapas
 - 3.9.4.3 Tracks del Camino
- 3.9.5 Enlaces
- 3.9.6 Descargas
- 3.9.7 Audioguías
- 3.9.8 Estadísticas
- 3.9.9 Foros

Enviado: 2022-05-01. Segunda versión: 2022-06-14.
Aceptado: 2022-06-15.

Índice de autores

Author index

Almeida, Mauricio, 13
Assunção, Suelene Santana, 23
Baños-Moreno, María-José, 13
Diniz, João Vítor Correa, 23
Felipe, Eduardo R., 13
Fujita, Mariângela Spotti Lopes, 23, 45

García Marco, Francisco Javier, 55
López del Ramo, Joaquín, 55
Pegoraro Novaes, Fernanda Carolina, 33
Piña-Landaburu, Zuriñe, 13
Redigolo, Franciele Marques, 23
Tolare, Beatriz Jessica, 45

Índice de materias en español

Subject index in Spanish

Amazon, 23
Análisis de contenido, 13
Bibliotecas universitarias, 45
Brasil, 23
Camino de Santiago, 55
Clasificación Decimal de Dewey, 33
Crawler by domain, 13
Crawlers, 13
Documentos audiovisuales, 23
Fisiología de la voz, 33
Homeopatía, 13
Indización, 45
Indización temática, 23
Interoperabilidad semántica, 55
Lenguajes de indización, 45

Metateoría, 33
Multimedia, 23
Organización del conocimiento, 55
Organizaciones no gubernamentales, 55
Repositorio Institucional de Geociencias de la
Compañía de Investigación de Recursos Minerales
(Brasil), 23
Repositorio LUME (Brasil), 23
Repositorio UFPA Multimedia (Brasil), 23
Repositorios institucionales, 23
Sitios web, 55
Taxonomías, 55
Tipos de voces, 33
Voces cultivadas, 33
Web, 13

Índice de materias en inglés

Subject index in English

Academic libraries, 45
Amazon, 23
Audiovisual documents, 23
Brazil, 23
Content analysis, 13
Crawlers, 13
Cultivated voices, 33
Dewey Decimal Classification, 33
Homeopathy, 13
Indexing, 45
Indexing languages, 45
Institutional repositories, 23
Institutional Repository of Geosciences of the Mineral
Resources Research Company (Brazil), 23

Knowledge organization, 55
LUME Repository (Brazil), 23
Metatheory, 33
Multimedia, 23
Non-governmental organization, 55
Semantic interoperability, 55
St. James' Way, 55
Taxonomies, 55
Thematic indexing, 23
UFPA Multimídia Repository (Brazil), 23
Voice physiology, 33
Voice types, 33
Web, 13
Websites, 55