



Barcelona ISKO 2019

Actas del IV Congreso ISKO España-Portugal

Jesús Tramullas
Piedad Garrido-Picazo
Gonzalo Marco-Cuenca
(editores)



ISKO España-Portugal
International Society for Knowledge Organization

Actas del IV Congreso ISKO España-Portugal 2019 XIV Congreso ISKO España

Barcelona, 11 y 12 de julio de 2019

Jesús Tramullas, Piedad Garrido-Picazo
y Gonzalo Marco-Cuenca
(editores)

Sociedad Internacional para la Organización del Conocimiento
(ISKO) – Capítulo Ibérico

Zaragoza, 2020

Primera edición
Zaragoza, 14 de abril de 2020

© De la edición: Sociedad Internacional para la Organización del Conocimiento (ISKO) –
Capítulo Ibérico

© De los textos individuales: los autores

ISBN: 978-84-09-20065-8

Depósito Legal: en trámite

DOI <https://doi.org/10.5281/zenodo.3758131>



Esta obra está bajo una licencia de Creative Commons
Atribución/Reconocimiento - 4.0 Internacional.

Organización del congreso.....	vii
--------------------------------	-----

Prólogo.....	ix
--------------	----

Jesús Gascón

Keynote

Si los nombres denominan a las cosas, esto ya no es lo que era. II: La organización del conocimiento.....	1
---	---

José Antonio Moreiro-González

Recuperación de información

Técnicas y usos en la clasificación automática de imágenes.....	11
---	----

Isidoro Gil Leiva, Pedro Díaz Ortuño, José Vicente Rodríguez Muñoz

Uso do modelo POLE para recuperação da informação investigativa policial por meio de grafos.....	27
--	----

Manoel Camilo de Sousa Netto, Adilson Luiz Pinto, Audilio Gonzales Aguilar

Generación automática de palabras clave para monitorizar dominios en redes sociales.....	37
--	----

Miguel Ángel Rozalén Soriano, Álvaro Aranda Jiménez

Editora UFPB: uma história fundamentada em dados.....	45
---	----

Geisa Cavalcante, Guilherme Dias

An approach to the Contextual Design methodology in the context of Information Science.....	57
---	----

Luís Miguel Oliveira Machado, André Pacheco

Repositório de dados como forma de organização do conhecimento e desenvolvimento científico.....	69
--	----

Elizabete Cristina de Souza De Aguiar Monteiro, Priscila Machado Borges Sena, Ricardo César Gonçalves Sant'ana, Ursula Blattmann

Estudio de la opinión del profesorado de la Universitat de Barcelona de la nueva herramienta de descubrimiento del CRAI, Cercabib.....	79
--	----

Andreu Sulé, Juanjo Boté, Marina Salse

Archivística

A representação arquivística na tradição canadense: subsídios para elaboração de um modelo ideal de ensino por meio da semântica textual.....	91
---	----

Glenda da Rocha Monteiro, Thiago Henrique Bragato Barros

e-EMGDE, RIC, NEDA y las normas de codificación: más allá de un perfil de aplicación en archivos.....	105
<i>María José Baños-Moreno, Francisco José Valentín Ruiz, Antonio Blázquez Martín de las Mulas</i>	

Instrumentos de classificação em arquivos: uma análise comparada na administração pública brasileira.....	119
<i>Paola Rodrigues Bittencourt</i>	

Archivos y organizaciones

Memória socioeconômica do ecossistema de startups de Florianópolis.....	129
<i>Priscila Machado Borges Sena, Ursula Blattmann, José Antonio Moreiro-González, Jéssica Bedin</i>	

Archivos personales y familiares

Estudo de tipologia documental em arquivos pessoais: contribuições para a organização do conhecimento arquivístico.....	137
<i>Mabel Meira Mota, Francisco José Aragão Pedroza Cunha</i>	

Propuesta teórica para la categorización y el etiquetado semántico de los archivos familiares a través del Fondo Kati.....	143
<i>Joseba Martínez de Lahidalga Santillana</i>	

Aproximación teórica y práctica a una propuesta de gestión conjunta: el archivo y la biblioteca de la familia Belmonte-Chico de Guzmán.....	153
<i>Camino Sánchez Oliveira</i>	

Organização do conhecimento e identificação tipológica: aportes para projetos de pesquisa em arquivos pessoais.....	173
<i>Lucia Maria Velloso De Oliveira, Camilla Campoi de Sobral, Bianca Panisset, José Antonio da Silva</i>	

Patrimonio

Colecciones digitales patrimoniales especializadas: estudio de la Red ARACNE....	185
<i>María Luisa Alvite Díez, Nieves Pena Sueiro</i>	

A dimensão social do catálogo Arca do Gosto no Brasil enquanto sistema de informação e sua relação com a organização do conhecimento.....	197
<i>Gabrieli Aparecida da Fonseca, Sonia Troitiño</i>	

A “fala que narra” e a “fala que demonstra”: os Arquivos de História Oral e o seu valor na construção das memórias e das identidades sociais.....	205
<i>Maria Cristina Vieira de Freitas</i>	

A promoção da experiência dos utilizadores como um estímulo à aprendizagem em museus: um estudo exploratório.....	219
<i>André Pacheco</i>	

Estratégias de reconstrução do património cultural pós-desastre, com o uso de tecnologias digitais: estudo de caso do Museu Nacional – Brasil.....	231
<i>Fernanda Miranda de Vasconcellos Motta, Ronaldo André Rodrigues da Silva</i>	

Imagen fotográfica

A utilização e tratamento de documentos fotográficos em projetos de investigação: o caso do Generation.Mobi.....	241
<i>Joana Sousa Rodrigues, Carla Teixeira Lopes</i>	

Análisis de las etiquetas de contenido de un banco de imágenes: Agefotostock.....	253
<i>Crispulo Travieso Rodríguez, Blanca Rodríguez Bravo</i>	

Análisis de tipos de descripción de fotografía antigua en Wikimedia Commons.....	267
<i>Jesús Tramullas, Ana I. Sánchez Casabón, Piedad Garrido-Picazo</i>	

La organización social de fotografías personales en Flickr: una mirada desde la gestión de información personal.....	275
<i>Jorge Franganillo</i>	

Contributos da indexação social para tratamento temático de documentos fotográficos provenientes de arquivos públicos.....	293
<i>Susana Sofia Cunha, Maria da Graça Melo Simões, Luciana de Souza Gracioso</i>	

Televisión y otros recursos audiovisuales

Análisis de los servicios 'a la carta' de las televisiones autonómicas españolas.....	303
<i>Laura Anton, Javier Guallar</i>	

Tecnologías del habla: nuevas oportunidades para los archivos de televisión.....	323
<i>Virginia Bazán Gil, Eduardo Lleida Solano, Manuel Gómez, Carmen Pérez, Alberto de Prada</i>	

Audio-Visual Semantics: propuesta de una ontología para la descripción de secuencias audiovisuales.....	337
<i>Juan Antonio Pastor-Sánchez, Tomás Saorín, Virginia Bazán Gil, Manuel Escribano, María José Baños Moreno</i>	

Organización del conocimiento

Organização do conhecimento durante o processo de investigação: utilização do ATLAS.ti em duas teses de Doutorado.....	349
<i>Luis Corujo, Jorge Revez, Carlos Guardado da Silva</i>	

A interdisciplinaridade como elemento de convergência no contexto patrimonial entre o Mundo da vida e o Mundo dos sistemas.....	363
<i>Marco Donizete Paulino da Silva, Luciana de Souza Gracioso, Maria da Graça Melo Simões</i>	

Filosofia da linguagem pragmática como aporte à representação da memória coletiva.....	373
<i>Luciana de Souza Gracioso, Maria da Graça Melo Simões, Samanta do Prado</i>	

Panorama teórico-analítico-sintético sobre a adoção de facetas no contexto da organização do conhecimento.....	383
<i>Márcio Bezerra da Silva, Zeny Duarte de Miranda</i>	

Emanuele Tesouro nos trópicos: caminhos de uma tradução e crítica teórico-histórica em Organização do Conhecimento.....	391
<i>Gustavo Silva Saldanha, Tatiana de Almeida, Maria Helena Teixeira da Silva</i>	

Sistemas de organización del conocimiento

Análise dos tipos de relações do modelo SKOS: perspectivas de representação de recursos audiovisuais.....	399
<i>Rogério Aparecido Sá Ramalho, Brígida Maria Nogueira Cervantes</i>	

Metacategorias semânticas para aplicação em SKOS.....	409
<i>Rita do Carmo Ferreira Laipelt, Luciana Monteiro Krebs, Rene Faustino Gabriel Júnior</i>	

Una terminología de teatro para organizar el patrimonio cultural universitario.....	419
<i>Catalina Naumis Peña, Ariel Alejandro Rodríguez García, Hugo Alberto Guadarrama Sánchez</i>	

Sistemas de organización del conocimiento: dominios concretos

Análises de palavras-chave como fonte de dados para obtenção de conhecimento sobre a evolução da ciência.....	431
<i>Jether Oliveira Gomes, Thiago Magela Rodrigues Dias, Gray Farias Moita, Adilson Luiz Pinto</i>	

A música na ciência da informação: uma análise de domínio da temática no Brasil no período de 1972 a 2010.....	441
<i>Emanuella Maria Barbosa Lourenço Ezerra, Andrea Carla Melo Marinho, Francisco Arrais Nascimento, Daniel Martínez-Ávila, Fabio Assis Pinho</i>	

Enriquecimiento de entidades de Wikidata mediante un modelo de descomposición y mapeado de categorías de Wikipedia.....	451
<i>Tomás Saorín, Juan Antonio Pastor Sánchez</i>	

A representação do domínio da Ética em Organização do Conhecimento nas bases de dados Scopus e KO Literature.....	467
<i>Isadora Victorino Evangelista, José Augusto Chaves Guimarães, Daniel Martínez-Ávila</i>	

Organización del conocimiento: perspectiva social

Garantía cultural: avance de elementos teóricos y metodológicos para su aplicación.....	475
<i>Mario Barité</i>	

A contribuição da organização de conhecimento no procedimento de classificação e indexação e nos processos crime com perspectiva de genero: interpelações acerca dos feminicídios no Estado de São Paulo – Brasil.....	485
<i>Denise Cristina Belan Fioravanti, Francisco Arrais Nascimento, Deise Maria Antonio Sabbag</i>	

O assassinato de Marielle Franco e os algoritmos racistas: dimensões aplicadas da teoria crítica da organização do conhecimento.....	493
<i>Gustavo Saldanha, Franciéle Carneiro Garcês da Silva, Graziela Santos Lima, Dirnéle Carneiro Garcês, Nathália Lima Romeiro</i>	

IV Congreso ISKO España-Portugal

XIV Congreso ISKO España

(Barcelona, 11-12 Julio de 2019)

Coordinación general:

- Jesús Gascón. Universitat de Barcelona, Facultat de Biblioteconomia i Documentació
- Carmen Agustín. Universidad de Zaragoza, Departamento de Ciencias de la Documentación e Historia de la Ciencia

Comité organizador (formado por miembros de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona):

- Ernest Abadal, Sílvia Argudo, Ángel Borrego, Constança Espelt, Jesús Gascón García, Javier Guallar, Núria Jornet, Mario Pérez-Montoro, Carina Rey, Josep Manuel Rodríguez Gairín, Concepción Rodríguez Parada, Andreu Sulé Duesa, Cristóbal Urbano, Aurora Vall Casas.

Comité científico. Comisión del programa científico

- Carmen Agustín. Universidad de Zaragoza, Departamento de Ciencias de la Documentación e Historia de la Ciencia
- Josep Àngel Borràs. Universitat de Barcelona, Facultat de Biblioteconomia i Documentació
- Alicia Conesa. Documentalista audiovisual
- Constança Espelt. Universitat de Barcelona, Facultat de Biblioteconomia i Documentació
- Mario Pérez-Montoro. Universitat de Barcelona, Facultat de Biblioteconomia i Documentació
- Jordi Serchs. Arxiu Fotogràfic de Barcelona
- Eugènia Serra. Biblioteca de Catalunya
- Jesús Tramullas. Universidad de Zaragoza, Departamento de Ciencias de la Documentación e Historia de la Ciencia
- Jesús Gascón, coordinador. Universitat de Barcelona, Facultat de Biblioteconomia i Documentació

Comité científico. Revisores

- Blanca Rodríguez Bravo. Universidad de León. España
- Carlos Guardado da Silva. Universidade de Lisboa. Portugal
- Ciro Lluca. Biblioteca de la Universitat Oberta de Catalunya. España
- Crispulo Travieso Rodríguez. Universidad de Salamanca. España
- Concha Soler. Institut Valencià de Cultura. España
- Fernanda Freita Costa de Torres. Instituto Federal de Brasília. Brasil
- Fernanda Ribeiro. Universidade do Porto. Portugal
- Ferran Burguilllos. Biblioteques Municipals de Sabadell. España
- Inés Nieto. Arxiu Històric de la Ciutat de Barcelona. España
- Iris López de Solís. Radiotelevisión Española. España
- Isabel Juncosa. Biblioteca de l'Il·lustre Col·legi de l'Advocacia de Barcelona. España
- Isidoro Gil Leiva. Universidad de Murcia. España
- Joan Chaparro. Betevé, Informació i Comunicació de Barcelona. España
- Jordi Andreu. Universitat de Barcelona. España
- Jorge Revez. Universidade de Lisboa. Portugal
- José Antonio Frías. Universidad de Salamanca. España
- Juan José Boté Vericad. Universitat de Barcelona. España
- Laia Foix. Institut d'Estudis Fotogràfics de Catalunya. España
- Luis Corujo. Universidade de Lisboa. Portugal
- Mabel Giménez Rayo. Accenture Outsourcing Services, para Antena 3. España
- Margarida Ullate. Unitat de Sonors i Audiovisuals de la Biblioteca de Catalunya. España
- Mari Váñez. Universitat Pompeu Fabra. España
- Maria Cristina Vieira de Freitas. Universidade de Coimbra. Portugal
- Maria Graça Melo Simões. Universidade de Coimbra. Portugal
- María Luisa Alvite Díez. Universidad de León. España
- Marta Masafret. Museu Nacional d'Art de Catalunya. España
- Mònica Bechini. Arxiu Audiovisual de Televisió de Catalunya. España
- Neus Conte. Museu Nacional d'Art de Catalunya. España
- Núria Jornet. Universitat de Barcelona. España
- Nuria Lloret Romero. Universitat Politècnica de València. España
- Pilar Ortego. Universidad de Extremadura. España
- Rosa San Segundo Manuel. Universidad Carlos III de Madrid. España
- Sergi Montes. Ateneu Barcelonès. España

Prólogo

A lo largo de sus encuentros bienales, la comunidad científica de ISKO en España y, desde hace unos años y con la constitución de ISKO España-Portugal, también la portuguesa, ha trabajado en multitud de temáticas y aspectos de la organización del conocimiento. Los congresos, como suele y ha de ser, han sido lugares de intercambio y debate en torno a estos aspectos, además de hacer posible la relación interpersonal.

En 2019, la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona acogió el IV Congreso ISKO España-Portugal, decimocuarto considerando los de ISKO España. Quedaba lejos el que, en 2005, había acogido la misma facultad. En esta nueva ocasión, tomó el relevo de Coimbra para ser el marco del encuentro, ahora con la colaboración del Área de Biblioteconomía y Documentación de la Universidad de Zaragoza, que se ha hecho responsable de la publicación de sus actas. El congreso tuvo lugar conjuntamente con el IX Encuentro Ibérico de EDICIC y las III Jornadas de Estudiantes de Ciencias de la Documentación: estos tres congresos conjuntos fueron los últimos actos acogidos por la más que centenaria facultad (sucesora de la Escola de Bibliotecàries fundada en 1915) antes de cambiar su nombre al actual de Facultat d'Informació i Mitjans Audiovisuals (Facultad de Información y Medios Audiovisuales).

La incorporación de estudios sobre comunicación audiovisual en la institución influyó en la elección del tema central del congreso. En la historia de los encuentros de ISKO, en pocas ocasiones se había tratado el tratamiento de los recursos audiovisuales y, en un momento en que su presencia es capital, se quiso profundizar en su tratamiento. La abundancia de proyectos de recuperación y difusión del patrimonio audiovisual (en televisiones, filmotecas, etc.) llevó a ampliar el foco temático a los recursos patrimoniales de cualquier tipo, otro de los campos donde se encuentran más iniciativas (desde programas de digitalización a titulaciones de postgrado y máster especializadas) y que, quizás por ese enfoque hacia el pasado, no recibe tanta atención en la bibliografía profesional.

Así, el congreso se configuró en torno a estos dos ejes temáticos y con el lema *Organización del conocimiento para la explotación de colecciones patrimoniales y archivos audiovisuales*. Justo es decir que la respuesta fue excelente y que, realmente, las contribuciones recibidas permitieron elaborar un programa predominantemente dedicado a los dos temas, aparte de algunas sesiones dedicadas a cuestiones generales de organización del conocimiento, como suele acontecer.

Vale la pena remarcar el intento de renovar el concepto mismo de encuentro científico. La presión por la publicación de resultados ha llevado a la necesidad de presentar trabajos, de publicar actas y de entender el congreso como el lugar donde darlos a conocer, sin más. Hemos vivido una época de saturación de congresos, seminarios y encuentros de todo tipo en el que, también por una u otra razón, asistimos a presentaciones donde los autores leen o replican el texto escrito que luego se publica en las actas, sin apenas debate ni intercambio de ideas. Hemos convertido

el congreso en una sesión de presentación de trabajos ya acabados, tal como se publican (o incluso, ya publicados en actas). Hemos perdido así la esencia misma del encuentro, el “encontrarse” para debatir e intercambiar y, sólo a partir de este intercambio, llegar a producir un trabajo mejor. Esta idea fue la que dio origen a los primeros congresos científicos modernos (el *Versammlung Deutscher Naturforscher und Ärzte*, en Leipzig en 1822 y, en 1828, en Berlín y convocado por A. von Humboldt, el primer congreso internacional).

Por ello, innovando (paradójicamente, volviendo a los orígenes), se propuso que en el congreso, las presentaciones por parte de los autores no fueran lecturas del trabajo, sino que, después de un máximo de diez minutos de presentación de las líneas y resultados principales, se dedicaran diez o quince al debate con los asistentes. Para ello, los autores pudieron distribuir, con antelación al congreso, los trabajos en plataformas en línea y el web del congreso ofrecía los enlaces a cada uno de los trabajos. Así, cualquier asistente podía leer y valorar los trabajos que le interesaban días antes de su presentación.

En las sesiones del congreso, con el público ya conocedor de la propuesta, se podía pasar directamente al debate y al intercambio y, a partir de las ideas que podían surgir en la sesión, el autor podía modificar su trabajo y mejorarlo. La prepublicación que había ofrecido en la plataforma se enriquece con las observaciones de los colegas que han intervenido en la sesión. Igualmente, se habilitaron foros en línea para cada sesión, para que cualquier persona pudiera exponer sus puntos de vista y contribuir, así, a la mejora de la propuesta.

Son estos trabajos, modificados después del congreso por sus autores y enriquecidos por las aportaciones del público en este “sistema de revisión en abierto”, los que se publican ahora en las actas.

La propuesta de este modelo de funcionamiento de congreso, que en este ámbito disciplinario no había sido ensayada, se llevó a cabo en este encuentro de ISKO España-Portugal y podemos valorarla como satisfactoria. Algunas sesiones se convirtieron, realmente, en foros de debate en torno al tema central de cada una de ellas y tuvo más peso el intercambio en el lugar que la exposición del trabajo en sí, que, al fin y al cabo, se podrá consultar ahora en estas actas. Se puede mejorar el modelo en aspectos como un mayor uso de los foros virtuales, poco utilizados, pero el primer paso hacía unos encuentros que realmente tenga sentido llevar a cabo se ha dado con la propuesta. Obviamente, requiere una mayor participación de los coordinadores de cada sesión, regulando las intervenciones y propiciando el debate, una mayor implicación de quienes presentan los trabajos y, sobre todo, una implicación absoluta del público, que ha de tener un papel activo no solo antes, sino también durante el congreso. Si queremos que el congreso sea un foco activo de producción e intercambio, todos los participantes, de una u otra manera, debemos de implicarnos para que así sea.

Otra particularidad del congreso fue la (reiterada en cada sesión) adhesión a la Iniciativa Helsinki sobre Multilingüismo en la Comunicación Científica, por lo que

con el objetivo de “asegurar que se proporciona un acceso igualitario al conocimiento científico en distintos idiomas” cada persona podía usar en sus intervenciones la lengua que prefiriese, sin traducción. Jugamos con la ventaja que las tres lenguas oficiales del congreso —español, portugués y catalán— son románicas y, con poco esfuerzo y poniendo facilidades de cada parte, fácilmente inteligibles. Así, el congreso se desarrolló indistintamente en estas tres lenguas, sin problemas graves de comprensión, al menos aparentemente. No solo se trata de ahorrar costes de organización sino de reivindicar la validez de todas las lenguas en la comunicación y la publicación científicas; más en momentos en que la tecnología permite superar muchas de las barreras lingüísticas.

Aun habiendo recibido algo más de 150 propuestas, no quisimos hacer un programa demasiado cargado para los asistentes: concentramos el encuentro en dos jornadas con no más de tres sesiones paralelas, combinando presentaciones de comunicaciones con mesas redondas y talleres. Fueron seleccionadas setenta propuestas de las que, finalmente, se presentaron sesenta, junto con tres talleres, una mesa redonda y trece pósteres. Se publican ahora aquellas que se han recibido ya revisadas y corregidas por sus autores.

En general, el tratamiento de los asuntos nucleares del congreso ha quedado equilibrado con el de los temas generales habituales, relacionados con la organización del conocimiento y la recuperación de la información. Ha habido más sobre organización en recursos patrimoniales y cabe destacar la presencia de un número considerable relacionadas con recursos archivísticos y, por ello, la presencia de profesionales de la archivística, poco habitual en otros congresos previos. En lo referente a los medios audiovisuales, la fotografía y la imagen fija han tenido algo más de atención que la imagen en movimiento, centrada casi siempre en los recursos televisivos. Entre los asuntos generales, al margen de los dos ejes temáticos, predominaron los trabajos sobre sistemas de organización del conocimiento, su aplicación en ámbitos particulares y sobre vocabularios; en cambio, se redujo la presencia de trabajos puramente teóricos.

Por ofrecer algunos datos estadísticos, casi la mitad de las contribuciones presentadas provienen del Brasil (en cantidad, son mayoría las de universidades de los estados de São Paulo y, en segundo lugar, de Rio de Janeiro). Algo más de la cuarta parte provienen de España y nueve de las comunicaciones de Portugal. Aunque la diferencia es grande en términos absolutos, no lo es tanto si relacionamos el número con el peso demográfico de cada estado y la cantidad de instituciones que trabajan en el campo de la Documentación en cada uno de ellos. Otras comunicaciones son obra de autores de México, Colombia, Cuba, Uruguay y, fuera de Iberoamérica, de Francia. Habría que sumar la presencia del belga Roger Roberts, ponente invitado y el neerlandés Eppe van Nispen en una mesa redonda.

Y como datos que pueden dar una idea de la forma en que trabajamos, solo once de las contribuciones son obra de un único autor: la colaboración es la norma general, con una media de entre dos y tres autores por trabajo, en la mayoría de casos,

compañeros en la misma institución. Para acabar, del total de 158 autores, 81 son mujeres y 77, hombres. Extraer conclusiones de estos datos, sin más contexto, sería aventurado y lo dejamos para los especialistas.

Tampoco podemos evaluar sin parcialidad el grado de calidad del evento ni la satisfacción de los asistentes: tenemos indicadores objetivos (como encuestas de valoración) y, sobre todo, subjetivos, pero esta valoración corresponde más a cada una de las personas que vivió el encuentro y que, a partir de su experiencia, habrá adquirido algo y guardará un recuerdo del conjunto. Siempre hay aspectos mejorables que los mismos organizadores detectan cuando se encuentran y que “apuntan” mentalmente para que no vuelvan a repetirse en un futuro, pero que ya es tarde para corregir sobre la marcha. En todo caso, parece que estos fueron menos que los aspectos positivos de los que todos podemos felicitarnos.

Acabamos con los obligados y merecidos agradecimientos a los agentes que han apoyado y colaborado para llevar a buen puerto este acontecimiento. Desde entidades que han financiado parte del equipamiento necesario (MDPI, Ores, Springer Nature, Libnova, el Departament de Cultura de la Generalitat de Catalunya o el Centre de Recerca en Informació, Comunicació i Cultura de la Universitat de Barcelona) a ISKO España-Portugal por confiar su encuentro bienal a nuestras capacidades y a todas las personas que han participado directamente en la organización y desarrollo: estudiantes voluntarios, profesores y personal administrativo y de apoyo de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona, miembros de los comités de organización, científico y de revisión, coordinadores de sesiones, público asistente... A todos, en fin, y muy especialmente al equipo de la Universidad de Zaragoza, coorganizadora y responsable de la publicación de estas actas, al de la actual Facultat d'Informació i Mitjans Audiovisuals de la Universitat de Barcelona y al equipo de coordinación de los tres congresos conjuntos que se hicieron en dicha facultad. Unos y otros pusimos de nuestra parte para conseguir un resultado que, en estas actas, podrá valorarse parcialmente; otras partes del congreso, igualmente importantes, no podrán recuperarse y quedarán en el recuerdo de cada uno. Quizás acabe siendo esto lo más relevante de un encuentro de este tipo...

Jesús Gascón García
Coordinador general del IV Congreso ISKO España-Portugal
Departament de Biblioteconomia, Documentació i Comunicació
Audiovisual
Facultat d'Informació i Mitjans Audiovisuals
Universitat de Barcelona

Si los nombres denominan a las cosas, esto ya no es lo que era. II: La organización del conocimiento

José-Antonio Moreiro-González

ORCID [0000-0002-8827-158X](https://orcid.org/0000-0002-8827-158X). Departamento de Biblioteconomía y Documentación,
Universidad Carlos III de Madrid, España.
jamore@bib.uc3m.es

Resumen. Segunda parte de la conferencia de clausura de EDICIC 2019 y de apertura de ISKO 2019, impartida en la Universidad de Barcelona el 11/07/2019. Esta parte atendió a las consecuencias de la transformación digital sobre los vocabularios semánticos. Se diferencian los lenguajes documentales consecutivos a la ciencia positivista de los lenguajes coordinados para la recuperación de la información propios de la Information Science, hasta plantear la respuesta de los sistemas de organización del conocimiento a las necesidades del entorno digital, en especial desde los de carácter semántico, taxonomías y ontologías, pero sin olvidar las folksonomías..

Palabras clave: Transformación digital; Sistemas de organización del conocimiento; Vocabularios semánticos; Información y Documentación

Abstract. It is part of the EDICIC 2019 closing conference and ISKO 2019 opening conference gived at University of Barcelona, 11/07/2019. This part dealt with the consequences of the digital transformation on semantic vocabularies. Documentary languages consecutive to positivist science are differentiated from coordinated languages for the information retrieval - Information Science, until the response of knowledge organization systems to the needs of the digital environment, especially from semantic taxonomies and ontologies, but without forgetting folksonomies.

Keywords: Digital transformation; Knowledge organization systems; Semantic vocabularies; Library and information science.

1 Introducción

Acudamos a Google que se ha propuesto eliminar los impedimentos físicos al acceso. Lo que ha tenido efectos notorios en la práctica de la comunicación académica. Así lo demuestra la manera en que se indizan los documentos científicos en Google Scholar o en su afán por situarse como una de las bases en el proceso de transición hacia la

Ciencia abierta para compartir toda la producción intelectual. Incluso el año pasado se estrenó Google Dataset Search, un motor de búsqueda que facilita el acceso universal a los conjuntos de datos ubicados en los repositorios de internet. Dos caminos, que se reflejan en la responsabilidad de los profesionales a la hora de suministrar recursos a los investigadores y a los estudiantes, bien sea desde las bibliotecas o desde los laboratorios e institutos que generan los datos de investigación.

Tanto Google, en cuanto arquetipo, como otras herramientas posibilitadas por la Web han multiplicado de tal manera la cantidad de información disponible y sus formas de acceso, que han acelerado y multiplicado las posibilidades de publicación y distribución de la información. Por más que sea un problema vivido antes, pues sucedió a finales de la revolución industrial y, luego, con las primeras aplicaciones informáticas en los años 60. Momentos que impulsaron el crecimiento de la ciencia, de su comunicación y su tratamiento avanzado. Y que supusieron cambios en los paradigmas de organización y representación del conocimiento. Hoy se produce y comparte información más que nunca y, con ello, como cada vez que hay una acumulación ingente de datos, aumenta su descontrol y desorganización. De repente, hemos vuelto a tropezarnos con una inesperada cosecha cuya cantidad de frutos somos incapaces de aprovechar.

2 Lenguajes documentales y ciencia positivista

Llegados aquí, hemos de ir hacia los nombres como reflejo de los momentos. Los lenguajes documentales, de base positivista, se materializaban como instrumentos simbólicos y conceptuales con el mismo nivel que lo hacían el resto de las coordenadas y unidades de medida universales. Su empleo pretendía, por encima de cualquier otra consideración, establecer la causa de cualquier fenómeno por medio de leyes generales y universales que, en este caso, perseguía una comunicación inequívoca.

Para asegurarla, dentro de cada dominio del conocimiento, se establecieron unos lenguajes muy formalizados, lógicamente coherentes y de intercambio desambiguado en los que el léxico se fijaba como nomenclatura. Cuando no se representaba por códigos alfanuméricos, pues de esta manera no dependía de ninguna de las lenguas habladas y tenía la ventaja de ser internacionalmente comprensible. Pese a que su mayor dificultad de empleo consistía precisamente en que el lenguaje formal de los códigos no se consideraba entre los propósitos de la comunicación humana asentada sobre el lenguaje natural. En su nombre va que, cuando los lenguajes terminológicos se formalizaban como vocabularios, el sustantivo era la forma de representación indispensable. Los sustantivos, los nombres manifestación tantas veces de conceptos estáticos, alejados de las acciones que cambian los estados de cosas. Entonces, para comprender el contexto de cada término, se le situaba dentro de unos esquemas apriorísticos de representación justificados a partir de las categorías generales de cada una de las ciencias y sus tecnologías. Se legitimaba así la representación científica de

una idea, tanto por el sustantivo individual que la significa, como colectivamente al ubicarla en su lugar correcto dentro del conocimiento de un dominio determinado. Por lo que su estructuración se establecía desde relaciones jerárquicas.



Figura 1. *Ciencia positivista-Documentación.* Sistemas clasificatorios. Los nombres.

3 **Lenguajes coordinados e *Information Science***

Muchas de las características positivistas de la representación de información pervivieron en la norma ISO 2788 de 1986, que todavía se justificaba desde el empleo normalizado de los términos. Si bien la sintaxis que los combinaba y relacionaba mejoró los atributos semánticos. Desde luego, la función de las representaciones es comunicar ideas, pero su manifestación por especialidades se estableció desde grupos de términos controlados, los descriptores, de significado estable por su univocidad nominal. En el tesoro cada descriptor se situaba dentro de su clase correspondiente, según la organización taxonómica concerniente a los esquemas jerárquicos de origen positivista. Por otra parte, esta pauta fue siempre la seguida en cualquier intento de organizar una colección desde aquellas bibliotecas de la edad antigua, por lo que se establece como uno de los fundamentos y de los métodos básicos de nuestro hacer.

Si los tesauros se convirtieron en el arquetipo de los vocabularios terminológicos fue porque garantizaban, desde la regulación y el control, unas indizaciones y recuperaciones adaptadas a las bases de datos. Para lo que eran idóneos por facilitar el uso combinado de descriptores en las búsquedas. Pero, más que nada, por contar con relaciones de asociación que permitían navegar por la lista de descriptores hasta dar con el más conveniente. Por mucho que su empleo fuese demasiado rígido en la descripción, poco exacto al fijar el tipo de relaciones e inadecuado respecto al potencial asociativo de sus términos.

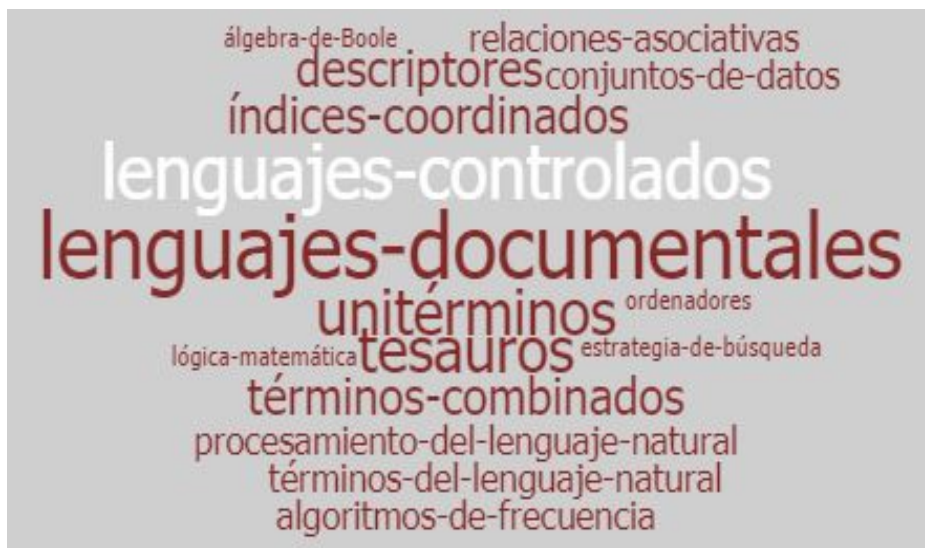


Figura 2. *Information Retrieval - Information Science.* Lenguajes documentales. Los nombres.

El empleo de la Inteligencia Artificial ha venido a ser otra de las causas del progresivo abandono de los lenguajes formales para organizar y recuperar el conocimiento. Una de sus ramas más conocidas, la del Procesamiento del lenguaje natural (PLN) lleva varias décadas realizando notorios avances en el modelaje de las lenguas por lo que su importancia ha sido definitiva en la generación, adaptación y empleo de los vocabularios de indización semántica y, desde luego, en su caracterización ontológica.

4 Transición digital y SKOS

Los lenguajes documentales tuvieron que renovarse para responder a los requisitos de los objetos digitales. Hasta la llegada de los SKOS, hace diez años, el contenido de los esquemas de ideas que estructuran los Sistemas de organización del conocimiento presentaba muchos problemas para mostrar con precisión las relaciones existentes entre los términos y para compartir información en la Web. Lo que no impedía que una parte de los anteriores métodos de recuperación de información siguieran usándose, como los algoritmos de localización o los cálculos de frecuencia terminológica, pero no bastaban para solucionar el enredo que causó la llegada del ciberespacio.

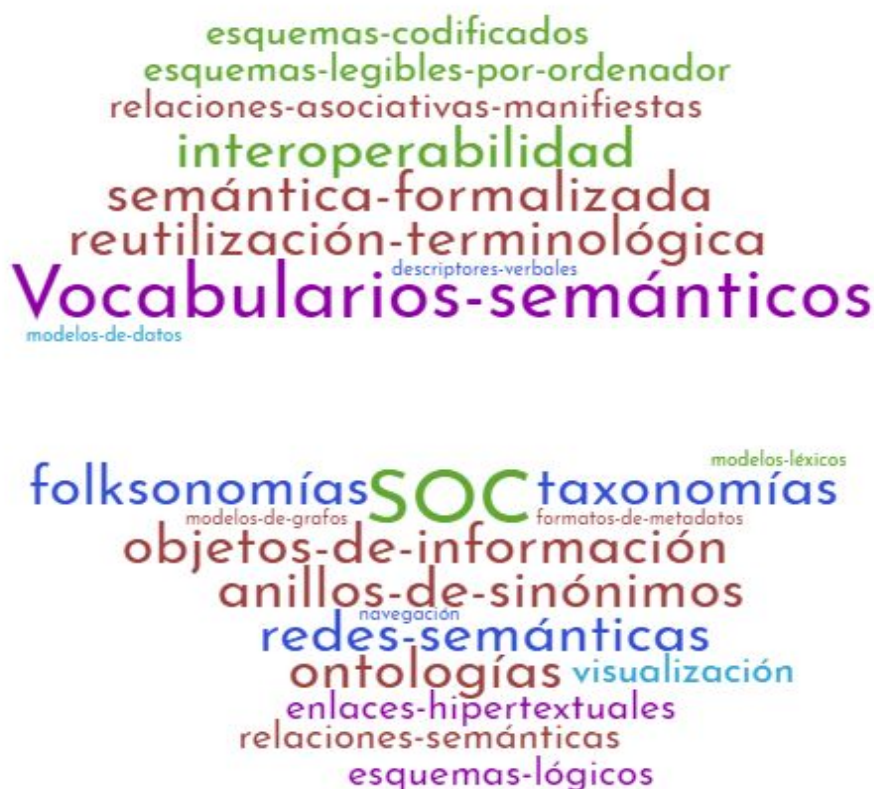


Figura 3. El entorno digital. Organización del conocimiento-SOC. Los nombres

Había que disponer de una semántica más formalizada que asegurase su comprensión por los ordenadores en el momento de agrupar y manejar la información con precisión. El progreso fue tan notorio que pronto hubo que establecer otras normas de indudable transcendencia para fundamentar y elaborar unos vocabularios de carácter semántico.

Y entonces, ¿cómo se organiza al conocimiento? Volvamos al atractivo tan espontáneo que ejercen los tesauros. De ser naufragos terminológicos en la Isla del tesoro hemos pasado a habitar el continente de los conceptos al que llegamos por una navegación ontológica que los vocabularios coordinados ya ofrecían, pero solo en la mente de sus usuarios, pues sus enlaces no estaban formalizados. Incluso a veces nos empeñamos en seguir aislados como si nada hubiese cambiado pues, por mucho apego que a esas peculiaridades tengamos y aunque estuviésemos convencidos de que el acierto las inspira, continuamos refiriéndonos y hasta enseñando los lenguajes controlados como si no existiesen las nuevas normas de 2005, 2007 y 2011. Incluso debemos de plantearnos si es funcional y efectivo seguir formando a los futuros

profesionales solo en métodos basados en jerarquías, como se hacía a principios del siglo XX, cuando ahora gran parte de los préstamos y colecciones tienen carácter electrónico y la perspectiva de los usuarios es ajena a cualquier cinturón que estreche sus búsquedas. Por más que aquellos métodos sigan mostrando validez en organizaciones clasificadas. Pelea constante de contradicciones entre la aspiración y la utilidad, lo tradicional y lo reciente.

5 Taxonomías y ontologías. Tan viejas y recién llegadas

Pues en este panorama inestable y discordante, las taxonomías y las ontologías aparecieron en el espectro de los sistemas de organización del conocimiento. Lo recién llegado, como respuesta a las carencias recuperadoras, que se designó con un nombre clásico en filosofía, el primero, o en las ciencias de la vida, el segundo. La ontología que desde el siglo XVII nombraba a la parte de la metafísica que trataba del ser en general y de sus propiedades trascendentales. Por consiguiente, de las cosas, de los objetos y de sus atributos y características. Parece que estoy aún cursando Ontología en la primera filosofía, cuando todos estos nombres recorrían el estudio del ser. Ahora los empleamos para identificar, definir y relacionar los objetos de información, también seres, cosas o clases. Son sinónimos de clasificación, ordenación, división, catálogo, inventario o sistemática que describen el modelo conceptual de un dominio. De forma habitual actúan como vocabularios particulares de una organización que no disponga de tesoro ni de clasificación que, además de organizar los contenidos propios de una institución, se aplica a sus servicios, productos y recursos humanos entre los que se establece jerarquía y se transfieren las propiedades semánticas desde los términos hiperónimos-genéricos a sus hipónimos-específicos.

Como además hacen las Taxonomías, uno de los grandes nombres de la Información y Documentación, pues se emplea en cuanto agrupación de los principios, métodos y fines de organizar. Tomó su apelativo, precisamente, de las clasificaciones que se establecían en la ciencia positivista para ordenar, jerarquizar y nombrar a los seres vivos, dentro de la biología. Ahora lo usamos como sinónimo de clasificación, ordenación, división, catálogo, inventario o sistemática. En definitiva, algo que hicimos siempre, que permanece en la base organizativa de los vocabularios semánticos como antes lo fue de los lenguajes documentales. El empleo del término taxonomía se ha resignificado por iniciativa de quienes investigan con las organizaciones informáticas de objetos (palabras incluidas) y cuya mayor aportación son los enlaces entre conceptos y la formalización de estos que les concede, precisamente, características ontológicas.

La conocida reutilización ecológica de las aplicaciones documentales, que no desprecia los argumentos ni los procesos, aunque se hayan superado, ha impulsado la persistencia de la superestructura jerárquica y clasificatoria que organizaba los lenguajes documentales. De forma que, ni en lo recién llegado se ha dejado de

plantear la necesidad de estructurar los conceptos, términos o palabras en categorías que podemos llamar con un nombre tan positivista como el de taxones que, al tiempo, ha recobrado plena actualidad. Lo que no es excusa para seguir utilizando un denominador tan sobrepasado como el de Lenguajes documentales. Las nuevas normas atienden a todos los tipos de SOC: taxonomías, tesauros, ontologías, encabezamientos de materia, esquemas de clasificación y redes semánticas. Por más que las taxonomías, puestas a organizar el conocimiento, actúen enlazadas y la responsabilidad de su elaboración, actualización y empleo haya pasado a manos de especialistas del sector, empresa o institución en la que se aplican. Como sucede en las taxonomías clínicas, contables, jurídicas y hasta en las aplicadas a pequeñas redes de PIMES, que se desarrollan por completo como aplicación directa y ventaja competitiva de sus entidades correspondientes.

Hace también treinta años que la red mundial de ordenadores posibilitó que los científicos compartiesen la información de sus archivos. Su transcendencia fue más allá, hasta alcanzar a modificar los patrones y modos de vida. Cuando se quiso distribuir esa información con carácter semántico hubo que actuar desde significados determinados con precisión y vinculados mediante enlaces. La web semántica se estableció en el cruce de los métodos seguidos en la reutilización informática del conocimiento, el tratamiento lógico-lingüístico y la Ciencia de la información. Ámbitos que intervienen para determinar y describir el significado de la información y cómo se debe de entender, además del propio dominio al que la ontología se aplica.

Por consiguiente, la transcendencia conceptual y operativa se ha dado al pasar de unos vocabularios terminológicos a otros conceptuales cuya vocación ontológica necesitaba unas normas específicas para fijar la recuperación de la información. Aparecieron, primero, la ANSI/NISO Z39.19:2005 y la británica BS 8723 (2005 y 2007), más tarde la ISO 25964-1:2011. Todas reconocen, incluso desde el mismo título, la interoperabilidad entre lenguas, vocabularios y sistemas como marco de enlace digital de los objetos para recuperar la información. Hay que llamar la atención respecto al empleo del término objetos de información o de contenido, en lugar de documentos, para denominar las entidades que contienen información. La causa está en admitir la unificación en el tratamiento de los diferentes recursos, incluso los museísticos, como consecuencia de la fuerza con que han irrumpido los soportes electrónicos y la posibilidad de su enlace hipertextual. Esas normas cubren todo el panorama de los vocabularios documentales. Para dinamizarlos propusieron ampliar el número de relaciones y visualizarlas mediante grafos de navegación entre conceptos. Se pasaron a emplear, además, categorías gramaticales antes inhabilitadas para afinar los matices semánticos y se establecieron nuevas conexiones asociativas entre los recursos de información y los dominios de aplicación. Como consecuencia, aumentaron la precisión y la eficacia, a la par que se mejoraba la representación y la recuperación.

6 Folksonomías

No buscan esta primera finalidad semántica en sus representaciones las folksonomías, aplicadas para indizar en la web social aprovechando las ventajas de cooperación que ofrece el uso extendido de las tecnologías de la información y la comunicación ("TIC"). Con las folksonomías se ha subvertido la noción inicial de orden en la red y la capacidad de los especialistas para organizarla, al hacer de ellas indizaciones sobre objetos con enlaces URL individuales. Se trata de una práctica de indización cooperativa de los propios usuarios que etiquetan los contenidos con palabras libres, a partir de su conocimiento o desconocimiento del asunto analizado.

En las folksonomías, el etiquetado hecho por cualquiera se comparte con otros usuarios que, a su vez, pueden indizar repetidamente el mismo objeto de información. El resultado ofrece una visión descentralizada y colaborativa, lo que contrasta con los sistemas de información tradicionales que eran jerárquicos y dirigidos. Ahora, cada nueva etiqueta se agrega a modo de "descripción intersubjetiva", creando algo muy diferente de las palabras-clave que los autores asignan a sus propios artículos o de los términos controlados en un sistema de indización profesional. Por esta razón, es tan importante entender el comportamiento de los usuarios, porque la forma que elijan para proceder en su etiquetado será un factor condicionante a la hora de recuperar con éxito la información, en especial de aquella dispuesta en la Web social.

7 Pensamiento final

Por encima del tipo de vocabulario, todos son semánticos (ontológicos, por tanto) con independencia de su estructura terminológica, o no lo son como vemos que sucede en origen a las folksonomías. De forma que una taxonomía o un tesaurus adquieren carácter ontológico si cumplen los requisitos de esquematización, formalización e interoperabilidad, con el añadido de que las relaciones contextuales de los conceptos se pueden representar empleando redes semánticas que mejoran su visualización y facilitan la representación.

La alta confianza que generan las ontologías cuando se gestiona el conocimiento se debe al ajuste con que representan los significados y los dominios y a la facilidad con que se entienden tanto por personas como por agentes de software. En su empleo exitoso influyó que la información se situaba siempre dentro de un contexto y se manejaba estructurada mediante lenguajes de etiquetado para posibilitar su lectura por las máquinas. El conocimiento de una especialidad se determinaba, pues, sin posible ambigüedad mediante la configuración formal de sus términos y relaciones. Lo que suponía, además, que se pudiese compartir y reutilizar.

Sobre este proceso se pasó de la recuperación de información, uno de los fundamentos iniciales y persistentes de la Ciencia de la Información, a la recuperación del conocimiento que se establece como característica principal de la web semántica para obtener respuestas precisas a consultas concretas. No se alcanza esta Web sin

modelar el dominio de conocimiento correspondiente, por tanto, el contexto que lleva a entender el contenido de un objeto de información.

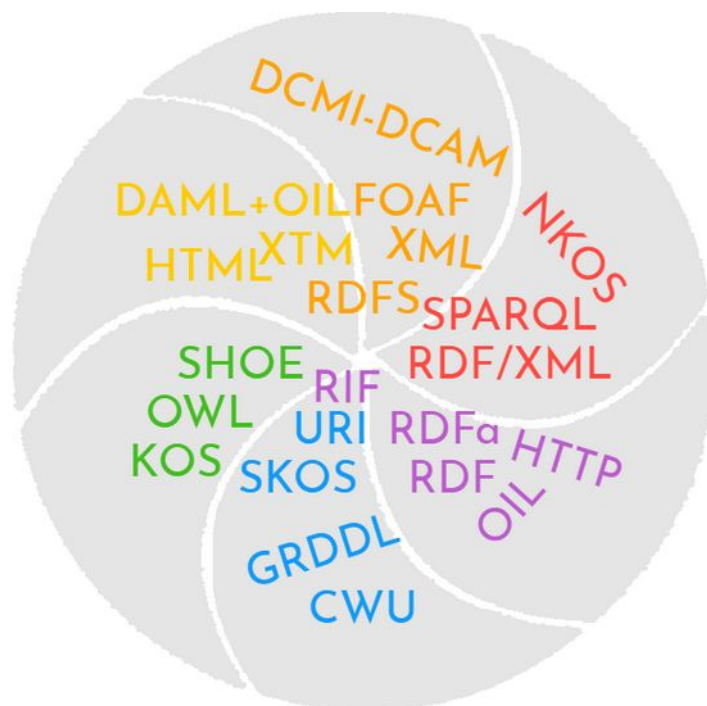


Figura 4. Sopa de letras. Vocabularios, tecnologías y estándares para la Web semántica.

Se obtiene una buena muestra de la interdisciplinariedad investigadora en los SOC al valorar la producción española indizada en WoS y Scopus y comprobar que, de los doce investigadores cuya aportación es superior a cuatro trabajos en los últimos nueve años, cinco de ellos, los que ocupan los primeros puestos, tienen su campo de actividad en Informática. Desde una cercanía que llega a conformar con frecuencia equipos con integrantes de ambas especialidades. E incluso con la presencia de investigadores en Lingüística aplicada.

Es así porque las ontologías favorecen la interoperabilidad, pues especifican cómo se relaciona un concepto con otros. De manera que cualquier aplicación llegará a aprovechar un concepto que no exista en su ontología pero que esté relacionado con sus contenidos. Por otra parte, las ontologías permiten hacer anotaciones semánticas en colecciones de objetos de información no textuales. Por lo que facilitan el comercio y los servicios electrónicos al integrar la descripción de los productos de un sector, objetos de información, al fin y al cabo. Así, los agentes que usen ontologías podrán buscar productos, negociar compras o localizar servicios de interés para los usuarios. De esta forma, los SOC ontológicos han trascendido las aplicaciones científico-técnicas para aplicarse a los ilimitados recursos que ofrece la web.

Referencias

- ANSI/NISO (2005). *Z39.19-2005. Guidelines for the construction, format and management of monolingual controlled Vocabularies*. Bethesda, Maryland: NISO Press.
- BSI Group. (2005 y 2007). *BS 8723/1-4. Structured vocabularies for information retrieval: guide*. London: BSI.
- ISO (2011). *25964-1-2011. Information and Documentation - Thesauri and interoperability with other vocabularies*. Genève: ISO.
- Moreiro-González, J.A. (2018). Adaptación de los vocabularios documentales al ambiente digital en red: léxico, significado y relaciones semánticas. *Informação&Sociedade: Estudos*, 28(1), 35-46.
- Moreiro-González, J.A.; Bolaños-Mejías, C. (2018). Folksonomy Indexing from the Assignment of Free Tags to Setup Subject: A Search Analysis into the Domain of Legal History. *Knowledge Organization*, 28(7), 574-585.

Técnicas y usos en la clasificación automática de imágenes

Isidoro Gil Leiva¹, Pedro Díaz Ortuño² y José Vicente Rodríguez Muñoz³

¹ ORCID [0000-0002-7175-3099](https://orcid.org/0000-0002-7175-3099). Facultad de Comunicación y Documentación, Universidad de Murcia, España
isgil@um.es

² ORCID [0000-0002-2975-766X](https://orcid.org/0000-0002-2975-766X). Facultad de Comunicación y Documentación, Universidad de Murcia, España
diazor@um.es

³ ORCID [0000-0001-6756-9638](https://orcid.org/0000-0001-6756-9638). Facultad de Comunicación y Documentación, Universidad de Murcia, España
jovi@um.es

Resumen. La producción y generación de información visual mediante teléfonos móviles y cámaras es ingente. También y principalmente a través de la teledetección, mediante la obtención de imágenes de la superficie terrestre por medio de aviones, naves espaciales y satélites que captan y sirven datos sobre meteorología, oceanografía, geología, geografía, geolocalización, seguridad, etc. Estos instrumentos de captura de imágenes generan cada día información visual imposible de procesar manualmente, de ahí que se recurra a diversas técnicas y métodos para la extracción automática de conocimientos útiles. Esta revisión bibliográfica, pretende conocer las técnicas y usos de la clasificación automática de imágenes. Para ello, se emplearon las Bases de datos Scopus y WoS para localizar documentos sobre clasificación automática de imágenes publicados entre 2008 y 2018. De los registros resultantes se buscaron los textos completos de los mismos, llevando a cabo un análisis del contenido para averiguar las técnicas más recurrentes y sus aplicaciones. Con todo ello, se hace patente que las tres técnicas más usadas para la clasificación automática de imágenes son los árboles de decisiones, redes neuronales y máquinas de vectores de soporte, siendo la aplicación de la clasificación automática muy variada, con la que se buscan automatizar procesos repetitivos, la inspección y vigilancias complejas, el control y desarrollo urbanístico o el reconocimiento y valoración tras catástrofes naturales, entre otros asuntos.

Palabras clave. Clasificación automática de imágenes; técnicas y aplicaciones; revisión bibliográfica.

Abstract. The production and generation of visual information through mobile phones and cameras is massive. Also and mainly using remote sensing, through

Gil Leiva, I., Díaz Ortuño, P., Rodríguez Muñoz, J.V. (2020). Técnicas y usos en la clasificación automática de imágenes. En J. Tramullas, P. Garrido-Picazo y G. Marco-Cuenca (eds.) *Actas del IV Congreso ISKO España y Portugal 2019* (pp. 11-26).
<https://doi.org/10.5281/zenodo.3733409>

the acquisition of images of the earth's surface by means of planes, spacecraft and satellites that capture and serve data on meteorology, oceanography, geology, geography, geolocation, security, etc. These image capture instruments generate visual information every day that is impossible to process manually, and therefore various techniques and methods are used to automatically extract useful knowledge. This literature review aims to understand the techniques and uses of automatic classification of images. To this goal, the Scopus and WoS databases were used to locate documents on the automatic classification of images published between 2008 and 2018. The resulting records were searched for their full texts, analysing the content to find out the most recurrent techniques and their applications. With all this, it has become clear that the three most commonly used techniques for the automatic classification of images are decision trees, neural networks and support vector machines, with the application of a wide variety of automatic classification, which seeks to automate repetitive processes, complex monitoring and inspection, urban control and development or recognition and assessment after natural disasters, among other matters.

Keywords. Automatic classification of images; techniques and applications; literature review.

1 Introducción

Desde las décadas de 1970-1980 los ingenieros, principalmente, han trabajado intensamente en el desarrollo de técnicas y algoritmos para la categorización e identificación de eventos mediante los datos obtenidos. Inicialmente, fue con el tratamiento de textos con el fin de clasificar automáticamente documentos de acuerdo a los términos de los documentos, para pasar posteriormente al tratamiento de sonido e imágenes en sus diferentes formatos.

En la actualidad, la producción de información no textual visual es enorme. Teléfonos móviles y cámaras (ocio, televisión, drones, etc.); la teledetección que toma imágenes de la superficie terrestre mediante aviones, naves espaciales o satélites (hoy en día más de cinco mil satélites sirven datos sobre meteorología, oceanografía, geología, geografía, geolocalización, seguridad, etc.). Estos numerosos instrumentos de captura de imágenes generan cada día información no textual inconmensurable e imposible de procesar manualmente, de ahí que se recurra a técnicas de selección y minería de datos para la extracción automática de conocimientos útiles.

Las herramientas de extracción de conocimiento tratan de extraer e identificar patrones en los datos para hacerlos comprensibles y útiles, la preparación de estos, la identificación de patrones, la interpretación, la visualización y, por último, la obtención de conocimiento. En las herramientas de extracción de conocimiento, la clasificación automática por medio de clasificadores de diferentes tipos desempeña una función nuclear, y viene empleándose en diferentes ámbitos de la vida real como en el ámbito de la economía (datos para la toma de decisiones), biometría

(reconocimiento de personas), inspección automática de productos y calidad (separación y empaquetado de frutas, carnes, paquetes, etc.), o en el procesamiento de las mencionadas imágenes procedentes de la teledetección.

La información textual y audiovisual ha venido siendo indizada y clasificada manualmente hasta que su volumen era relativamente manejable. En la década de 2000 se comenzó a trabajar en la anotación automática de imágenes para la asignación de los conceptos que aparecen en las imágenes tras el procesamiento del color, la textura, las formas, los objetos o las acciones. De esta manera se detecta y anota automáticamente palabras como: día, noche, interior, exterior, hombre, bicicleta, río, edificio, vegetación, etc. a partir de una colección de imágenes previamente anotadas que sirven de modelo de aprendizaje al sistema. Se trata, por tanto, de una notación física y no semántica de la imagen.

Los procesos seguidos en el análisis de las imágenes podemos dividirlos en tres niveles:

- Procesos de bajo nivel: adquisición de las imágenes y preparación de las imágenes (mejora y realce de contraste o de bordes), identificación del espacio de color a través de modelos que componen cada color mediante una combinación lineal de los tres colores primarios rojo, verde y azul (RGB) y el modelo HSV (Tono, Saturación y Valor).

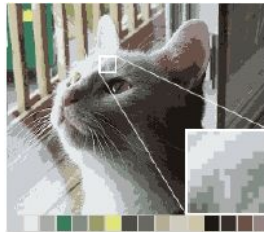


Figura 1. Ejemplo de cuantificación del color. Fuente: Velazco Paredes (2014, p. 16).

- Procesos de nivel medio: segmentación de la imagen por niveles de grises, bordes, o formas, entre otras y, por otro lado, marcado de las partes de la imagen.

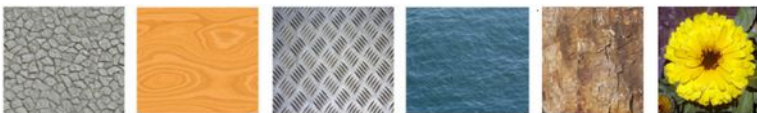

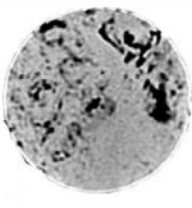


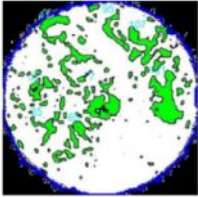
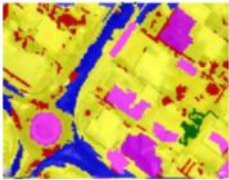


Figura 2. Ejemplos de texturas en una imagen. Piedra, madera, metal, agua, roca y flor.

Tabla 1: Ejemplos de segmentación de imágenes. Fuente: (a) Dong ping Tian (2013, p. 390); (b) Binhui et al. (2017, p. 91); (c) Jabari & Zhang (2013, p. 776).

	(a)	(b)	(c)
Imagen original			
Imagen segmentada			

- Proceso de nivel alto: reconocimiento de patrones e interpretación usando para ello clasificadores basados en redes neuronales, máquinas de soporte vectorial (SVM, en sus siglas en inglés) o árbol de decisión, entre otros.

2 **Objetivos**

El objetivo de esta revisión bibliográfica es triple, por un lado, revisar las técnicas y métodos de clasificación usados en la última década en la clasificación automática de información no textual; describir someramente las técnicas más empleadas; así como averiguar a qué ámbitos se están dirigiendo dichas investigaciones.

3 **Material y método**

Para lograr los objetivos marcados, en primer lugar se seleccionaron las Bases de datos Scopus y WoS para localizar documentos sobre clasificación automática de imágenes publicados entre 2008 y 2018. En la Tabla 3 se detallan las búsquedas empleadas y el número de documentos recuperados.

Tabla 2. Búsquedas y resultados obtenidos.

Base de datos	Ecuación de búsqueda	Publicados 2008-2018
Scopus	Keywords: “automatic indexing” AND Keywords: “image classification” = 44 documentos	32
Scopus	Kw: “automatic indexing” AND Keywords: “image analysis” = 102 doc.	9
WoS	Tema: “automatic indexing” AND Tema: “image analysis” = 12	2
Wos	Tema : “automatic indexing” AND Tema: “image classification” = 11	3
Total		45

De los cuarenta y cinco registros recuperados que cumplían los requisitos de las búsquedas planteadas (temática y período de publicación entre 2008 y 2018 ambos años incluidos), sólo se localizó el texto completo de treinta trabajos. Por tanto, estas treinta publicaciones fueron las manejadas para ejecutar la revisión. Para la recopilación de los datos, durante la lectura, se usó una plantilla donde se consignaron los datos de interés, en este caso, información relativa a las técnicas y métodos de clasificación, el objeto de análisis o el tipo de imágenes manejadas por los autores, entre otros aspectos.

4 Resultados

En el análisis de los resultados obtenidos, dentro de los aspectos a analizar, encontramos las siguientes características

4.1 Técnicas más usadas en la clasificación automática de imágenes.

El análisis de los treinta tres artículos publicados entre 2008 y 2018 nos ha permitido averiguar que la mayor parte de los investigadores recurren a técnicas de clasificación supervisada que cuenta con procesos de entrenamiento y aprendizaje antes de la clasificación real de las imágenes; pero también se observa interés en las no supervisadas que no usan conjuntos de datos de entrenamiento. Las técnicas supervisadas más empleadas han sido el árbol de decisión, las redes neuronales y las máquinas de vectores de soporte.

Tabla 3. Técnicas para la clasificación de alto nivel.

<i>Autor/es</i>	<i>Clasificación de características visuales semánticas de alto nivel</i>
Al-Batah, et al. 2009	Redes neuronales
Amutha y Kavitha, 2011	Máquina de vectores de soporte
Augereau et al. 2014	Máquinas de vectores de soporte
Cintra et al. 2010	Árboles de decisión
Correia et al. 2018	Clasificación K-Medias no supervisada
Dimitrios et al., 2010	Clasificador automático de cuerpos deformados
Ghaffarian, 2014	Clasificación por paralelepípedos mejorada
HaCohen-Kerner et al., 2015	Algoritmo J48, Random Forests (RF) y Sequential Minimal Optimization (SMO)
Hao, Ge y Wang, 2018	Máquinas de vectores de soporte
Hemsley, Mukundan, 2009	Multifractal Spectra. Técnicas de clustering.
Hermosilla, et al. 2011	Árboles de decisión (Clasificador C5.0)
Jabari y Zhang, 2013	Reglas difusas, sistema de inferencia
Jain, 2013	Máquina de vectores de soporte. Redes neuronales
Kupidura, et al. 2016	Modelo “GROUPS”
Li et al., 2017	Máquina de vectores de soporte
Lin, Li, Yu, y Wu, 2018	Algoritmo genético
Ma, et al. 2016	Mejora del método de Lalonde
Minetto et al. 2014	Máquinas de vectores de soporte
Morioka et al., 2016	Clasificador de tabla de decisiones
Murphy, Maggioni, 2018	Aprendizaje de difusión espectral-espacial
Raj y SivaSathya, 2016	Redes neuronales
Rezaeian, 2012	Máquina de vectores de soporte
Senthilnath, et al. 2016	Agrupamiento K-Medias; Algoritmo Genético (GA)
Shen, 2009	Redes neuronales
Yang, et al. 2017	Árboles de decisión
Tian, 2013	Revisión
You, et al. 2018	Algoritmo MBI
Zhang & Zhang, 2011	Redes neuronales. Lógica difusa
Zhang et. Al 2008	Árboles de decisión

4.2 Descripción de las técnicas más usadas en la clasificación automática de imágenes.

A continuación, se describen sucintamente las tres técnicas de clasificación supervisadas más usadas para la clasificación automática de las características visuales semánticas de las imágenes.

1 Árboles de decisión. Esta técnica surge a mitad de la década de 1940 de la teoría de juegos de John Von Neumann y Oskar Morgenstern. Posteriormente, con la Informática se incorporan los conjuntos de entrenamiento. En el ámbito de la inteligencia artificial, son usados los árboles de decisión para enseñar a un algoritmo a comportarse como un humano ante una situación determinada. Un árbol de decisión es una representación en forma jerárquica de una decisión y de todas sus consecuencias. Ayudan a tomar la decisión más acertada ante varias posibles y así alcanzar la resolución de un problema según la mejor probabilidad. Este proceso decisorio se visualiza en un modelo gráfico arbóreo con un sistema de notación compuesto por cuadrados (nodo de decisión que representa el momento puntual de la decisión y se extiende en tantas ramas como alternativas posibles existan); círculos (nodo de incertidumbre y probabilidad que muestra los posibles resultados); y flechas (ramificación de las alternativas donde cada línea indica la probabilidad de un posible resultado. Las probabilidades de dos alternativas suman siempre 1).

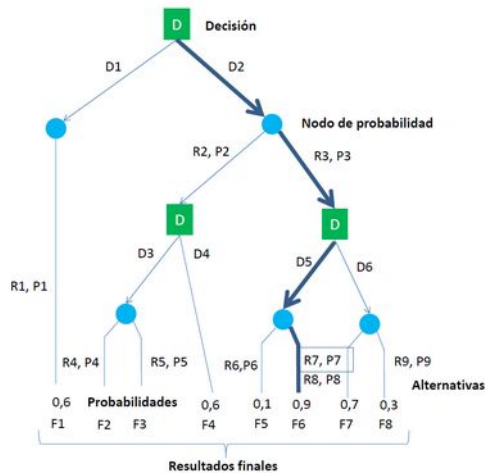


Figura 3. Representación de un árbol de decisión.

2 Redes neuronales. La modelización mediante redes neuronales busca simular capacidades humanas como la de asociación de hechos y, de este modo, aplicar la experiencia acumulada. Se trata de una simulación abstracta de los sistemas nerviosos

biológicos. El neurofísico Warren McCulloch y el matemático Walter Pitts en 1943 propusieron el primer modelo computacional basado en la actividad nerviosa mediante circuitos eléctricos. Una clasificación básica de los modelos neuronales los divide en modelos biológicos (tratan de simular sistemas neuronales biológicos) y modelos artificiales (redes artificiales aplicadas). Los modelos de red neuronal de tipo biológico están modelizados como una regla de propagación de señales con una entrada, una toma de decisiones y una salida.

Los modelos de redes neuronales artificiales tienen la capacidad de aprender a realizar tareas mediante una experiencia inicial o entrenamiento; capacidad para la auto-organización tras la fase de aprendizaje; capacidad de operar en tiempo real debido al uso de un gran número de nodos con un alto grado de interconectividad; y capacidad para realizar un procesamiento no lineal que permite clasificar patrones y aumentar su inmunidad a cometer fallos ante situaciones distorsionadas o datos incompletos. La estructura de la neurona artificial está compuesta por nodos o unidades. En la Figura 4 se muestran sus componentes.

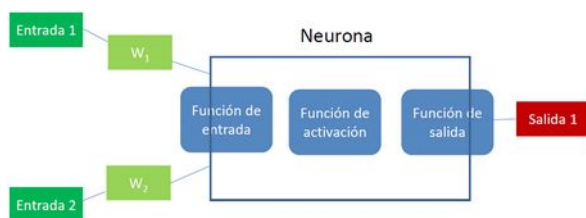


Figura 4. Componentes de una red neuronal artificial.

La función de entrada reduce todos los valores de entrada (W_1 , W_2 , W_3 , etc.) a un único valor, que se denomina entrada global, y se obtiene mediante una función (sumatorio, producto o el valor de entrada mayor), multiplicados por sus correspondientes pesos. La función de activación calcula el valor o estado de actividad de una neurona que puede ser (0, 1) ó (-1, 1) mediante el cómputo de una función, siendo las más usadas la función umbral, lineal, sigmoidea o tangente hiperbólica. El tercer elemento de una red neuronal es la función de salida que proporciona la respuesta o valor de salida que puede ser transmitido a otras neuronas vinculadas. Las redes neuronales se organizan en capas y cada capa contiene un número variable de neuronas. La capa de entrada que recibe directamente la información de fuentes externas a la red; las capas ocultas, que se encuentra entre la de entrada y la salida y no mantienen contacto alguno con el exterior (puede ser de una sola capa o multicapa); y por último, la capa de salida que envía información de la red hacia el exterior.

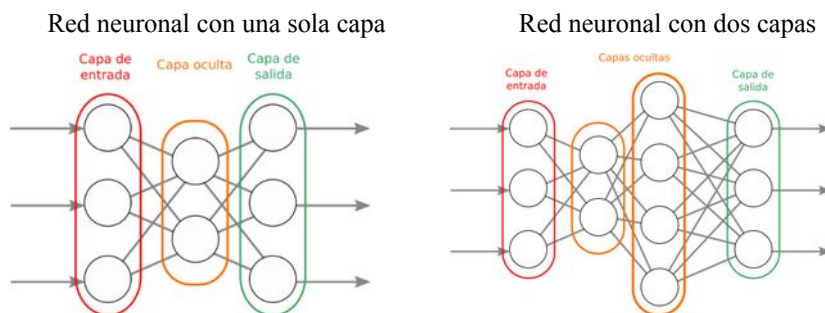


Figura 5. Capas en una red neuronal artificial. Fuente: <https://robologs.net/2017/01/22/tutorial-de-redes-neuronales-con-vrep-c-y-linux/>

El conocimiento en una red neuronal está distribuido por todo el sistema y el aprendizaje en una red implica un cambio constante en los pesos de las conexiones entre neuronas, lo que significa la destrucción, modificación y creación de conexiones entre ellas. Los pesos representan la información usada por la red para resolver un problema, y cuando los pesos permanecen estables significa que el proceso ha finalizado y por tanto, la red ha terminado el aprendizaje.

Hay dos fases en la modelización con redes neuronales: fase de entrenamiento y aprendizaje (se emplean al menos dos conjuntos de datos y patrones con ciertas variaciones para definir un modelo a partir de unos datos de entrada y una salida deseada) y fase de prueba (se persigue minimizar el error entre la salida del modelo y lo deseado).

3 Máquinas de vectores de soporte. Las máquinas de vectores soporte (SVM, en sus siglas en inglés) fue introducida en la década de 1990 por Vladimir Vapnik y sus colegas de laboratorio desde la teoría del aprendizaje estadístico, para resolver problemas de regresión y clasificación binaria lineal o hiperplano usando las funciones kernel. La SVM se ha utilizado en la categorización de textos, el reconocimiento de grafías o la clasificación de imágenes.

La modelización de SVM a partir de unas entradas determinadas busca establecer dos clases separadas por una línea y un hiperplano representadas con puntos en un espacio bidimensional. Una SVM es un algoritmo que a partir del producto escalar de los vectores multidimensionales de las muestras construye un hiperplano o conjunto de hiperplanos y un espacio de dimensionalidad muy alta que separa los grupos. Con el hiperplano se trata de encontrar el hiperplano óptimo de separación entre las clases; es decir, el plano para el cual la separación entre clases es máxima. El hiperplano óptimo se define por una línea recta.

En la Figura 6 se muestran los hiperplanos de separación en un espacio bidimensional de un conjunto de datos separables en dos clases. En la izquierda se

ofrecen los posibles planos, en el centro, el hiperplano de separación óptimo, y a la derecha, un ejemplo de clasificación automática de imágenes.



Figura 6. Hiperplanos de separación en un espacio bidimensional

Para la modelización de las SVM se requiere una fase inicial de entrenamiento y aprendizaje. Una vez conformada la modelización el sistema está listo para recibir nuevas entradas y buscar la correspondencia entre el modelo y las entradas para clasificarlas.

Las tres técnicas que acabamos de describir brevemente emplean alguno de estos tres tipos de aprendizaje para la resolución de problemas: a) aprendizaje supervisado: Se suministra un conjunto de datos de entrenamiento con ejemplos ya resueltos en forma de pares de datos que representa el comportamiento, el grupo o el valor adecuados, es decir, el problema y la solución. Tras el proceso de aprendizaje se suele emplear un segundo conjunto de pares de datos para validar la tasa de error. El entrenamiento es un proceso iterativo de adaptación de tal manera que el resultado se acerque a las salidas deseadas; b) aprendizaje no supervisado: No se proporcionan las salidas esperadas. La red maneja los datos de entrada y un conjunto de reglas para encontrar patrones de los que debe aprender; y c) aprendizaje con refuerzo: Se aprende a través de un proceso de ensayo y error y un sistema de recompensa/penalización de acuerdo a los resultados que se van logrando.

Para finalizar este apartado se esquematiza la secuenciación de los procesos ejecutados en la clasificación automática de imágenes.

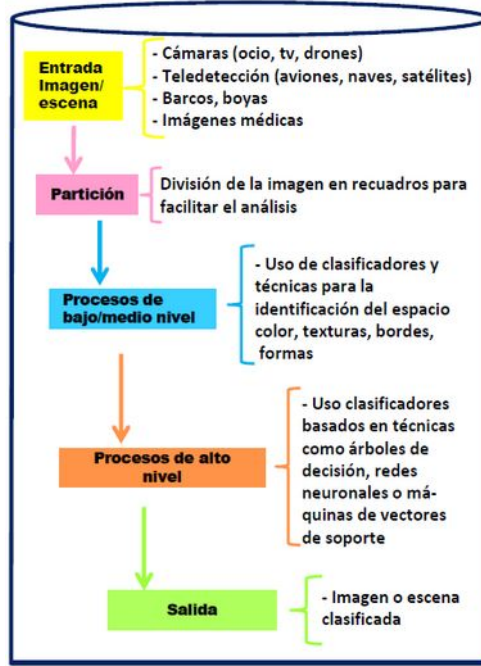


Figura 7. Procesos en la clasificación automática de imágenes.

4.3 Objeto de estudio y aplicaciones de la clasificación automática de imágenes.

El análisis de los treinta textos que conforma la muestra para la realización de esta revisión también ha permitido identificar los tipos de imágenes con las que se está trabajando, así como el objeto de estudio (Tabla 5).

La mayor parte de las imágenes manejadas para las investigaciones provienen de la teledetección (satélites) aunque también se usan imágenes procedentes de microscopios o de cámaras móviles para captar escenas urbanas. La aplicación de la clasificación automática es muy variada y se busca la automatización de procesos repetitivos (como, por ejemplo, la clasificación de hojas del tabaco por calidades); inspección y vigilancia complejas (estado y evolución del hielo marino, cobertura del suelo, biomasa arbórea en bosques, identificación de especies forestales, etc.); inspección, control y desarrollo urbanístico; reconocimiento y valoración de zonas después de catástrofes naturales (daños en edificios o pérdidas de cosechas, etc.) entre otros asuntos.

Tabla 4. Usos y aplicaciones de la clasificación automática.

<i>Autor/es</i>	<i>Aplicaciones y usos</i>
Al-Batah et al. 2009	Características (forma, tamaño y textura de la superficie) de los agregados del hormigón
Amutha y Kavitha, 2011	Imágenes (genérico)
Augereau, et al. 2014	Imágenes en documentos
Cintra et al. 2010	Cobertura del suelo para el monitoreo de la selva tropical
Correia et al. 2018	Imágenes de sensores remotos y satélites
Dimitrios et al., 2010	Parásitosgenes microscópicas 2D
Ghaffarian, 2014	Detección de edificios
HaCohen-Kerner et al., 2015	Identificación de categorías (salud, estilo de vida, ocio, etc.)
Hao, Ge y Wang, 2018	Anotación de imágenes mediante conceptos
Hermosilla et al. 2011	Detección y localización de edificios
Hughes, 2009	Clasificación de tipos de hielo marino
Jabari y Zhang, 2013	Imágenes urbanas satelitales (sombra, vegetación, camino, etc)
Jain, 2013	Imágenes (genérico)
Kupidura et al. 2016	Imágenes satelitales (agua, bosque, baja vegetación, suelo, etc.)
Li et al., 2017	Fracturas y vulgos en rocas
Lin et al. 2018	Tierras desnudas, humedales, agua, construcción, etc.
Ma et al. 2016	Dosel bosque (biomasa de madera, hojas o zona de árboles)
Minetto et al. 2014	Texto legible en escenas urbanas (carteles publicitarios, etc.)
Morioka et al., 2016	Riesgo de aneurismas aórticos abdominales
Hemsley, Mukundan, 2009	Imágenes microscópicas tejidos pulmonar, hepático y arterial
Murphy, Maggioni, 2018	Segmentación de imágenes hiperespectrales.
Raj y SivaSathya, 2016	Formas espectrales, índices de vegetación, datos multisensor
Rezaeian, 2012	Edificios dañados por terremotos
Senthilnath et al., 2016	Tipos cultivo: áreas problemáticas, cambios estacionales, etc.
Shen, 2009	Autoría de pinturas de los siglos XVI al XVIII
Yang et al. 2017	Cantidad de arroz en arrozales para tasaciones tras catástrofes
You et al. 2018	Detección y delimitación de edificios en zonas urbanas
Zhang & Zhang, 2011	Hojas de tabaco para agilizar los procesos de selección
Zhang e al. 2008	Clasificación de especies forestales

5 Conclusiones

En primer lugar, hay que señalar que la producción de información no textual es en la actualidad ingente y que este fenómeno no va a dejar de crecer de modo exponencial en adelante. Las cámaras fijas y móviles en todas sus manifestaciones, así como la teledetección con la toma imágenes de la superficie terrestre mediante aviones o dispositivos espaciales (pensemos que hoy, más de cinco mil satélites sirven datos sobre meteorología, oceanografía, geología, geografía, geolocalización, seguridad, etc.), son instrumentos que suministran petabytes de datos cada día, imposible de procesar manualmente. Debido a esta enorme cantidad de imágenes se recurre a técnicas para la extracción automática de información útil y procesable para la toma de decisiones.

La mayor parte de las imágenes manejadas para las investigaciones provienen de la teledetección (satélites), pero la variedad de otros dispositivos es patente, entiéndase, imágenes procedentes de microscopios, para el campo de la Ciencias naturales, así como el de las cámaras fijas o móviles para captar escenas urbanas, tanto humanas como de territorio. Hay que destacar, que las tres técnicas más usadas en la clasificación automática de imágenes son árboles de decisiones, redes neuronales y máquinas de vectores de soporte. Sin olvidar que otras son utilizadas, apareciendo novedades más sofisticadas con el devenir del tiempo, incluyendo algunas que por su estrategia para los servicios de seguridad no son públicas. Asimismo, el uso de la clasificación y aprendizaje automáticos es diversa, buscando, como es evidente, automatizar procesos repetitivos, de inspección y vigilancia compleja, entre otros asuntos. Finalmente, dejar patente que estos procesos son cada vez más utilizados y su desarrollo en el futuro, además de estratégico, va a significar un avance para cuestiones tales como el Internet de las cosas, como paradigma de la cotidianidad con las que vamos a convivir con ellas.

Referencias

- Al-Batah, M. S., Isa, N. A. M., Zamli, K. Z., Sani, Z. M., y Azizli, K. A. (2009). A novel aggregate classification technique using moment invariants and cascaded multilayered perceptron network. *International Journal of Mineral Processing*, 92(1-2), 92–102.
- Amutha, A. L., y Kavitha, S. (2011). Features based classification of images using weighted feature support vector machines. *Int J Comput Appl*, 26(10), 23–9.
- Augereau, O., Journet, N., Vialard, A., y Domenger, J.-P. (2014). Improving classification of an industrial document image database by combining visual and textual features. En *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on* (pp. 314–318). IEEE.

- Cintra, D. P., Novack, T., Rego, L. F. G., Costa, G., y Feitosa, R. Q. (2010). PIMAR Project-Monitoring the atlantic rainforest remnants and the urban growth of the Rio de Janeiro city (Brazil) through remote sensing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-4/C7.
- Correia, R., Duarte, L., Teodoro, A., y Monteiro, A. (2018). Processing Image to Geographical Information Systems (PI2GIS)—A Learning Tool for QGIS. *Education Sciences*, 8(2), 83.
- Das, A.J. Saikia, N. y Sarma, K.K. (2016). Object classification and tracking in real time: an overview. P. 250-295. En Santhi, V., Acharjya, D.P. y Ezhilarasan, M. (eds.). *Emerging technologies in intelligent applications for image and video processing*. Hershey, PA: IGI Global.
- Dimitrios, A., Rousopoulos, P., Papaodysseus, C., Panagopoulos, M., Loumou, P., y Theodoropoulos, G. (2010). A general methodology for the determination of 2D bodies elastic deformation invariants: Application to the automatic identification of parasites. *IEEE transactions on pattern analysis and machine intelligence*, 32(5), 799–814.
- Ghaffarian, S. (2014). Automatic building detection based on supervised classification using high resolution Google Earth images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 101.
- HaCohen-Kerner, Y., Sabag, A., Liparas, D., Moumtzidou, A., Vrochidis, S., y Kompatsiaris, I. (2015). Classification Using Various Machine Learning Methods and Combinations of Key-Phrases and Visual Features. En *Semantic Keyword-based Search on Structured Data Sources* (pp. 64–75). Springer.
- Hao Z., Ge H. y Wang, L. (2018) Visual attention mechanism and support vector machine based automatic image annotation. *PLoS ONE*, 13(11).
- Hemsley, A., y Mukundan, R. (2009). Multifractal Measures for Tissue Image Classification and Retrieval. 11th IEEE International Symposium on Multimedia(ISM), San Diego, California, pp. 618-623.
- Hermosilla, T., Ruiz, L. A., Recio, J. A., y Estornell, J. (2011). Evaluation of automatic building detection approaches combining high resolution images and LiDAR data. *Remote Sensing*, 3(6), 1188–1210.
- Hughes, N. (2009). Sea ice type classification from multichannel passive microwave datasets. En *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009* (Vol. 3, pp. III–125). IEEE.
- Jabari, S. y Zhang, Y. (2013). Very high resolution satellite image classification using fuzzy rule-based Systems. *Algorithms*, 6, 762-781.
- Jain, S. (2013). A machine learning approach: Svm for image classification in cbir. *Internationa Journal of Aplication or Annovation in Engineering & Management (IJAIEEM)*, 2(4).

- Kupidura, P., Osipińska-Skotak, K., y Pluto-Kossakowska, J. (2016). Automatic Approach to Vhr Satellite Image Classification. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 277–282.
- Li, B., Tan, X., Wang, F., Lian, P., Gao, W., y Li, Y. (2017). Fracture and vug characterization and carbonate rock type automatic classification using X-ray CT images. *Journal of Petroleum Science and Engineering*, 153, 88–96.
- Lin, Y., Li, W. J., Yu, J., y Wu, C. Z. (2018). Ecological Sensitivity Evaluation of Tourist Region Based on Remote Sensing Image—Taking Chaohu Lake Area as a Case Study. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(3).
- Ma, L., Zeheng, G., Eitel, J. y Moskal, L.M. (2016). Improved Salient Feature-Based Approach for Automatically Separating Photosynthetic and Nonphotosynthetic Components Within Terrestrial Lidar Point Cloud Data of Forest Canopies. *IEEE Transactions on geoscience and remote sensing*, 54(2), 679–696.
- Minetto, R. Thome, N., Cord, M., Leite, N.J. y Stolfi, J. (2014). SnooperText: A text detection system for automatic indexing of urban scenes. *Computer Vision and Image Understanding*, 122, 92–104.
- Morioka, C., Meng, F., Taira, R., Sayre, J., Zimmerman, P., Ishimitsu, D. El-Saden, S. (2016). Automatic classification of ultrasound screening examinations of the abdominal aorta. *Journal of digital imaging*, 29(6), 742–748.
- Murphy, J. M., y Maggioni, M. (2018). Unsupervised Clustering and Active Learning of Hyperspectral Images With Nonlinear Diffusion. *IEEE Transactions on Geoscience and Remote Sensing*.
- Ping Tian, D. (2013). A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering*, 8(4), 385–396.
- Raj, K. J., y SivaSathya, S. (2016). A Survey of Various Algorithms Used on Multispectral Satellite Image Classification of Alwar Image Dataset. *Indian Journal of Science and Technology*, 9(45).
- Rezaeian, M. (2012). Automatic classification of collapsed buildings using stereo aerial images. *International Journal of Computer Applications*, 46(21), 35–42.
- Senthilnath, J., Kulkarni, S., Benediktsson, J. A., y Yang, X.-S. (2016). A novel approach for multispectral satellite image classification based on the bat algorithm. *IEEE Geoscience and Remote Sensing Letters*, 13(4), 599–603.
- Shen, J. (2009). Stochastic modeling western paintings for effective classification. *Pattern Recognition*. 42(2), 293–301.
- Tian, Dong Ping (2013). A Review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering*, 8(4), 385–396.

- Velazco Paredes, Y.E. (2014). Recuperación de imágenes por contenido basado en regiones con retroalimentación por relevancia. Universidad Nacional de San Agustín de Arequipa. Tesis doctoral. Disponible en: <http://repositorio.unsa.edu.pe/handle/UNSA/6129> [Consultado: 05-11-2018].
- Yang, M. D., Huang, K. S., Kuo, Y. H., Tsai, H. P., & Lin, L. M. (2017). Spatial and spectral hybrid image classification for rice lodging assessment through UAV imagery. *Remote Sensing*, 9(6), 583.
- You, Y., Wang, S., Ma, Y., Chen, G., Wang, B., Shen, M., y Liu, W. (2018). Building detection from VHR remote sensing imagery based on the morphological building index. *Remote Sensing*, 10(8), 1287. doi: 10.3390/rs10081287
- Zhang, F. y Zhang, X. (2011). Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation. *Sensors*, 11, 2369-2384.
- Zhang, W., Hu, B., Jing, L., Woods, M. E., y Courville, P. (2008). Automatic forest species classification using combined LIDAR data and optical imagery. En *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International* (Vol. 3, pp. III-134). IEEE.

Uso do modelo POLE para recuperação da informação investigativa policial por meio de grafos

Manoel Camilo de Sousa Netto¹, Adilson Luiz Pinto² y Audilio Gonzales Aguilar³

¹ ORCID [0000-0002-7762-7958](https://orcid.org/0000-0002-7762-7958). Agente da Polícia Federal, Discente de Mestrado PGCIN/UFSC, Brasil.
camilo.mcsn@dpf.gov.br

² ORCID [0000-0002-4142-2061](https://orcid.org/0000-0002-4142-2061). Departamento de Ciência da Informação, PGCIN/UFSC, Brasil.
adilson.pinto@ufsc.br

³ ORCID [0000-0001-8693-2076](https://orcid.org/0000-0001-8693-2076). Departamento de Documentación, Université Montpellier 3, França.
audilio.gonzales@gamil.com

Resumen. O estudo apresentou uma metodologia de investigação, baseada em modelos visuais, que usa tecnologias que representem grafos formados por pessoas, objetos, locais, eventos (POLE), e por suas interconexões. Como metodologia utilizou uma rede de mundo pequeno e desenvolveu um modelo conceitual para as interconexões entre as entidades POLE e a sua recuperação por meio da linguagem de padrão aberto OpenCypher. Posteriormente, foi criado um banco de dados baseado em grafos contendo 407.542 vértices e 316.810 conexões, todos relativos a empresas e seus sócios, parentescos e vínculos empregatícios. Os dados se restringiram a uma unidade federativa do Brasil. A pesquisa então recuperou as entidades e suas interconexões até dois graus de separação do líder de uma organização criminosa real. Os resultados de comparação revelou que dezesseis dos vinte e três indivíduos presos e cinco das oito empresas envolvidas no esquema criminoso faziam parte da rede de até dois graus de separação do líder. Grande parte da rede criminosa estava registrada em dados sob posse dos organismos policiais e demonstrou que os dados obtidos estavam dentro dos princípios dos três graus de influência, pois boa parte das relações sociais que se estabeleceram em até três graus no mundo real.

Palabras clave: Rede de Mundo Pequeno. Classificação por Grafos. POLE.

Abstract. The article presented a research methodology, based on visual models, using technologies that represent graphs formed by people, objects, places, events (POLE), and their interconnections. As a methodology it used a

small world network and developed a conceptual model for the interconnections between the POLE entities and their recovery through the OpenCypher open standard language. Subsequently, a database was created based on graphs containing 407,542 vertices and 316,810 connections, all related to companies and their partners, kinship and employment links. Data were submitted to a federative unit in Brazil. The survey then recovered the entities and their interconnections up to two degrees of separation from the leader of a real criminal organization. The comparison results showed that sixteen of the twenty-three arrested individuals and five of the eight companies involved in the criminal scheme were part of the network of up to two degrees of separation from the leader. Much of the criminal network was recorded in data held by police agencies and showed that the data obtained were within the principles of the three degrees of influence because much of the social relations that have been established up to three degrees in the real world.

Keywords: Small World Network. Classification by Graphs. POLE.

1 Introdução

O homem é essencialmente um ser social, o que implica que grande parte de suas interações com o mundo estejam permeadas por laços sociais decorrentes dessa importante característica. Entretanto, as conexões que surgem e se desfazem entre o homem e tudo aquilo que o cerca não se exaurem nos laços sociais, posto que, além das pessoas, outros tipos de entes se relacionam com o ser humano e entre si. Assim, quase todos os cenários do cotidiano humano podem ser considerados como uma grande rede formada por interligações entre diversos atores, em especial pessoas, objetos, locais e eventos (quarteto nominado como “POLE”). Essa rede pode ser representada como um grafo, ferramenta visual que confere grande poder cognitivo sobre vínculos e entidades, inclusive aqueles que representam fatos criminais de amplo interesse policial e que podem ajudar a elucidar delitos.

Isto posto propõe-se, pelo presente trabalho, explanar o uso de um método de investigação policial que se utilize de uma plataforma tecnológica que crie um modelo visual composto por entidades do quarteto POLE e suas interconexões, apresentando um caso de uso concreto. Investigações baseadas em POLE podem otimizar a recuperação da informação e a cognição policial acerca de cenários que possam se configurar como delitos criminais, pois conferem ao policial o poder de travessia em profundidade típico dos grafos.

2 Contexto do trabalho: O modelo POLE

POLE é o acrônimo para o quarteto Person-Object-Location-Event. A tabela abaixo demonstra algumas instâncias dos tipos para cada um desses elementos.

Tabela 1: Exemplos de instâncias do quarteto POLE

Tipo	Pessoa	Objeto	Localização	Evento
Instâncias do tipo	<ul style="list-style-type: none">• Suspeitos• Testemunha• Empregado• Sócio	<ul style="list-style-type: none">• Veículo• Documento• Telefone• E-mail	<ul style="list-style-type: none">• Local de crime• Endereço residencial• Endereço Laboral	<ul style="list-style-type: none">• Crime• Prisão• Encontro• Telefonema

Fonte: Extraído e adaptado de (Depeau, 2019)

O método POLE pode ser sido aplicado em policiamento, contraterrorismo e antiterrorismo, controle de fronteiras e imigração, rede de proteção infantil, serviços sociais, reabilitação de presos, dentre outros usos. Segundo (Eifrem, 2018) agências de aplicação da lei estão avaliando o uso do desse modelo de dados para trabalhar com dados criminais. Essa avaliação decorre da capacidade dos grafos e seus algoritmos conseguirem unir os pontos e encontrar conexões em grandes quantidades de dados, o que os torna uma escolha natural para tal uso.

3 Seis graus de separação e três graus de influência

Stanley Milgram realizou, em 1967, um experimento que chamou de “Small Word Problem”. Segundo Watts (2009), Milgram estava interessado em uma hipótese - não resolvida pelos sociólogos até então - segundo a qual o mundo visto como uma enorme rede de relações sociais era, em certo sentido, pequeno. Ora, esse é um fenômeno comum pois, vez por outra, o homem se surpreende pela revelação da existência de relações sociais antes desconhecidas, fato que popularizou a frase “que mundo pequeno!”. Christakis e Fowler (2009) descrevem parte do experimento de Milgram. Relatam como aquele psicólogo entregou, a algumas centenas de pessoas que moravam em Nebraska-EUA, uma carta endereçada a um homem de negócios em Boston-EUA. O objetivo era que todas os emissários tentassem fazer, através do serviço de correios e usando destinatários intermediários, com que a correspondência chegasse ao seu destino final. Assim, as cartas que puderam ser rastreadas percorreram, em média, seis graus de separação da origem ao destino, sendo que cada grau foi contabilizado como uma rota percorrida por uma carta entre uma origem e um destino.

Cada grau representava, em verdade, uma unidade de relação social. Watts (2009) apresentou a formalização matemática que resultou na solução do “Small World Problem”: a ocorrência conjunta de alta clusterização local com baixa separação global. Christakis (2010) enumera também que, apesar dos seis graus de separação proposto por Milgram, as influências ocorrem principalmente até três graus da origem da informação pois, a partir dessa profundidade, a propagação da mensagem perde força de dissipação. Assim, a reverberação informacional se enfraquece por três fatores principais. O primeiro desses fatores é o decaimento intrínseco, consequência

da deterioração da fidelidade. Seres humanos quase sempre precisam confiar nas suas relações para repassar e receber informações que consideram relevantes e influentes. Em particular, nas organizações criminosas, o sigilo é parte importante da conduta delitiva, pois a revelação dos crimes representa uma ameaça aos transgressores. As entidades mais distantes e periféricas podem até atuar no cerne criminal, mas dificilmente deterão informações relevantes sobre os administradores centrais. Concretamente, observa-se tal comportamento nos casos em que os portadores que carregam drogas, conhecidos no Brasil como “mulas”, quase nunca conhecem os chefes do tráfico que lhes oportunizam o funesto ofício. O segundo fator é a instabilidade da rede, segundo a qual quanto mais dois entes se afastam, mais ligações intermediárias existem entre a origem e o destino da informação. Caminhos maiores aumentam sobremaneira a probabilidade de que uma das ligações intermediárias seja rompida e toda a rota seja desfeita pela quebra. Nos esquemas criminais, manter o distanciamento até três graus mantém a rede estável pela redução da quantidade de ligações intermediárias suscetíveis ao rompimento. O terceiro e último fator é biológico: o ser humano parece ter evoluído em pequenos grupos nos quais todos estariam conectados aos demais por até três graus. Ao que parece, o passado homínido, ausentes as tecnologias atuais que a todos aproximam, impediu uma evolução para manutenção de relações estáveis a mais de três graus.

4 Desenvolvimento de um modelo conceitual de rede para dados investigativos usando POLE

Conhecidos os fatos de que há, em redes de mundo pequeno, poucos graus de separação social e três graus de influência relevante, é possível realizar reflexões acerca da interação do ser humano com o seu entorno, mesmo quando as relações não possuam natureza social. A seguir apresenta-se um modelo conceitual de relações entre POLEs.

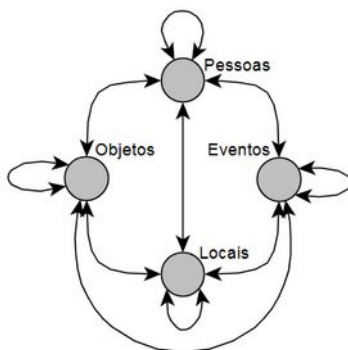


Figura 1. Modelo conceitual de relações para o modelo *POLE*

As conexões entre POLEs, quando importantes nas investigações criminais, são quase sempre consequência da sociabilidade humana. Elas formam um grande conjunto de dados do mundo real em rede, ora denominado Ar. Há dados de Ar que não são parte persistente em sistemas informacionais. Por exemplo, uma relação de amizade pode não estar representada em banco de dados. Entretanto, parte das relações entre POLEs contidas em Ar possuem relevância registral e, por isso, exigem catalogação sistemas informacionais, como no caso de um imóvel (objeto) conectado ao seu proprietário (pessoa). Há, portanto, diversas relações entre POLEs que persistem em bancos de dados, alguns de acesso aberto e outros de acesso restrito. Um subconjunto importante desses dados, ora denominado Ap, está sob domínio dos órgãos policiais. Quando se menciona dados de acesso restrito, vem à tona a discussão acerca da questão ética da violação da privacidade por parte das polícias, entretanto é relevante considerar que o objetivo final é o combate ao crime e as ações dos agentes públicos pautam-se pela legalidade. Os limites legais que comprometam o direito à privacidade não devem ser violados.

Eventualmente, Ap conterá informações que delineiam, preveem – e até mesmo tipificam – condutas criminais. Isto posto, sob certo sentido, Ap pode ser representada e explorada como uma rede de dados sob posse das polícias. Tecnicamente é crível considerar essa rede como um grafo composto por vértices de qualquer um dos tipos manifestos pelo quarteto POLE (pessoas, objetos, locais e eventos) e suas interconexões. Esses dados representam são decorrentes dos elementos do mundo real pertencentes ao conjunto Ar. O conjunto Ap, subconjunto de Ar se configura, portanto, como um conjunto de dados persistentes em sistemas acessíveis à polícia. Ap pode ser representado por um subgrafo decorrente de alguns elementos do mundo real e também pertencentes à Ar, tal qual representado pela figura 1, a seguir.

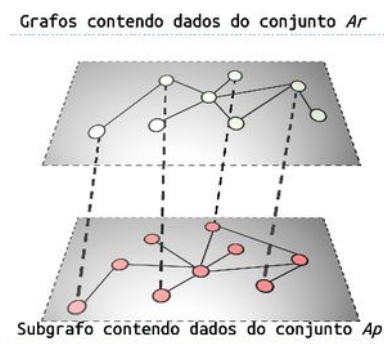



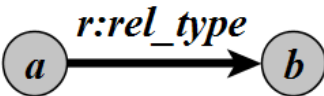
Figura 2. Alguns dados do conjunto Ar (mundo real) geram elementos de Ap, dados sob domínio dos órgãos policiais em sistemas informacionais

5 Representação do conjunto Ap por meio de uma infraestrutura baseada em grafos

A quantidade de dados gerados, transmitidos e armazenados cresceu exponencialmente na vigente era da informação. Essa amplidão é uma das características do conjunto Ap. Para materializá-lo como rede apta a ser analisada em investigações criminais, faz-se necessário uma ferramenta que suporte as dimensões informacionais massivas de um ambiente típico de Big Data. O repositório ideal para a rede de POLEs Ap é um banco de dados baseado em grafos. Usando uma dessas ferramentas, o Neo4j (2018) versão 3.2, o presente trabalho criou uma infraestrutura investigativa baseada em POLEs, para a qual foram carregados dados de sociedade em empresas, vínculos empregatícios e societários e parentescos, criando um banco de dados contendo 407.542 vértices e 316.810 conexões.

Segundo o site Neo4j (2019), o projeto OpenCypher, linguagem suportada pelo banco de dados, visa fornecer uma especificação completa e aberta de uma linguagem de consulta. Considerada candidata a um padrão de mercado, foi adotada por diversos bancos de dados, dentre os quais o Neo4j. OpenCypher tem foco no domínio do conhecimento ao invés da mecânica do acesso aos dados e é facilmente legível por humanos. Adicionalmente usa inglês e iconografia, fatos que minimizam a curva de aprendizagem. Vértices e ligações de um grafo podem ser expressos por Cypher. A Cypher Query Language Reference (Neo4j, 2019) especifica que, um vértice de um grafo segue uma representação simplória mediante o uso de parênteses. A representação de uma ligação é realizada utilizando os caracteres traço, colchetes e dois-pontos. A tabela a seguir resume essas representações:

Tabela 2. Representações iconográficas básicas (vértice e ligação) na linguagem OpenCypher

Descrição da representação	Representação em grafo	Representação iconográfica utilizando OpenCypher
Vértice “a” de um grafo		(a)
Dois vértices “a” e “b”, relacionados por uma ligação “r” do tipo “rel_type” direcionada de a para b		<pre>match (a) - [r:rel_type] -> (b)</pre>

As representações básicas podem ser ampliadas para representar grafos complexos. Isso permite recuperar informações mediante uma funcionalidade própria dos grafos: a travessia pela profundidade dos graus de separação. Uma das principais

consultas que poderia ser realizada seria aquela fundamentada nas observações sobre três graus de influência de Christakis (2010).

6 **Três graus de influência em uma organização criminosa real**

No Brasil, os órgãos públicos contratam serviços de empresas privadas mediante um procedimento denominado licitação. A lei brasileira 8.666, sua instituidora (Presidência da República, 1993), a estabeleceu com fins de garantir a observância do princípio constitucional da isonomia. Trata-se de uma concorrência mediante a qual diversas empresas disputam entre si para, mediante um processo transparente, venderem bens ou serviços aos órgãos públicos brasileiros. A empresa que oferecer o melhor custo-benefício sagra-se vencedora.

Em um caso concreto, em agosto do ano de 2018 foi deflagrada uma operação policial, pela Polícia Federal do Brasil, para combater uma organização criminosa que, sob uma mesma administração central e oculta, criou diversas empresas para agirem em conluio em licitações, com prévio acerto, criando uma falsa aparência de concorrência nos certames. Essa atuação tipifica-se como crime perante a legislação penal brasileira. Qualquer empresa que se sagra-se vencedora do certame estaria atuando em prol do mesmo grupo. A organização criminosa escolheu criteriosamente os participantes da empreitada. As empresas criadas para falsamente competirem entre si possuíam como sócios parentes e ex-funcionários do chefe da organização criminosa, pessoas essas que gozavam da sua plena confiança. Grande parte dessas relações faziam parte de bancos de dados governamentais. A equipe de policiais trabalhou por meses na elucidação. A maioria dos relacionamentos eram vínculos empregatícios e societários (vigentes ou prescritos). Entretanto, após a deflagração da operação, uma consulta em OpenCypher foi executada no banco de dados (Neo4j) ora implementado.

O resultado obtido foram todas as POLEs até dois graus de separação do chefe da organização criminosa (identificado pelo Cadastro de Pessoa Física - CPF). Foram escolhidos dois graus, menos do que os três graus de influência sugeridos por Christakis (2010), porque o número de entidades não relacionadas com o crime aumentou consideravelmente no 3º grau.

Tabela 3. Representação, na linguagem OpenCypher, objetivando obter POLEs até dois graus

Descrição	Representação iconográfica utilizando OpenCypher
Retorne todas as entidades que estão até dois graus de separação de uma pessoa que possui um determinado valor do CPF (o número real foi omitido)	<pre>MATCH (chefe:PessoaFisica)- [*1..2]-(entidades) WHERE chefe.cpf=‘9999999999’</pre>

Muitas horas de trabalho poderiam ter sido poupadas caso a estratégia da consulta tivesse sido adotada de modo prévio. Evidente que o resultado obtido não teria sido o único subsídio para as prisões. Diversas outras evidências são coletadas mediante interceptação telefônica, quebra de sigilo bancário, vigilância, etc. Entretanto, percebe-se que as pessoas que se coadunaram no mundo real para criminosamente agir, podem eventualmente deixar rastros em graus próximos de si em sistemas informacionais. Esses rastros podem ser detectados de forma mais eficaz, conforme a explanação dos resultados da consulta, apresentados a seguir, demonstrarão.

7 Resultados

O grafo retornado pela consulta de até dois graus de separação a partir do chefe do esquema criminoso retornou 16 (dezesseis) dos 23 (vinte e três) indivíduos presos na operação policial. No mesmo grafo surgiram 5 (cinco) das 8 (oito) empresas envolvidas. Esses dados foram decorrentes das formalizações de empresas e dos seus quadros societários e empregatícios vigentes e prescritos e, ainda, dos parentescos entre as pessoas. Até o 2º grau de separação vários indícios de atuação conjunta se revelaram, pois, se tais empresas fossem concorrentes reais em licitações, a probabilidade de que existam densas redes entre elas e as pessoas que as constituem seria mínima. A figura 2 demonstra uma anômala rede de relacionamentos entre entidades cuja regra deveria ser a desconexão. Os resultados, portanto, estiveram dentro dos princípios propostos por Christakis, pois boa parte das relações sociais que se estabeleceram até três graus no mundo real geraram dados a poucos graus de separação entre si em sistemas informacionais.

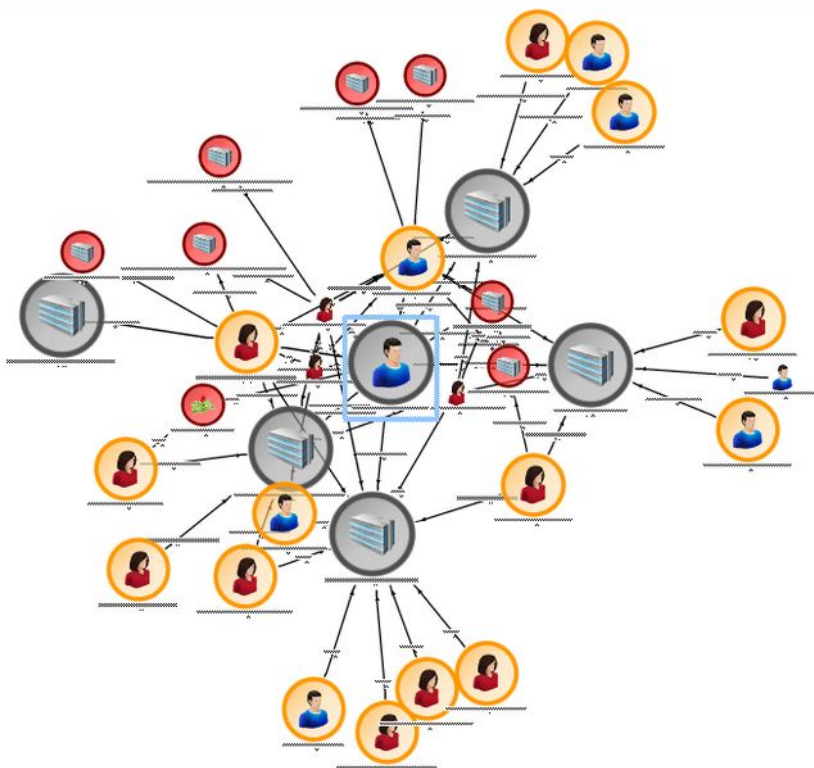


Figura 3. Grafo contendo todas as entidades POLE até dois graus de separação do chefe da organização criminosa

Referencias

- Depeau, J. (2019). *Graph Technology Is in the POLE Position to Help Law Enforcement*. Neo4j Graph Database Platform. Available at: <https://neo4j.com/blog/graph-technology-pole-position-law-enforcement/>
- Milgram, S. (1967). The Small-World Problem. *Psychology Today*, (1), pp.61-67.
- Eifrem, E. (2018). *Pole and Graph: A Duo Set On Interpreting Criminal Networks*. Publicnet.co.uk. Available at: <http://www.publicnet.co.uk/default/2018/07/25/pole-and-graph-a-duo-set-on-interpreting-criminal-networks/>
- J Watts, D. (2009). *Seis Graus de Separação: A Evolução da Ciência de Redes em uma Era Conectada*. 1st ed. São Paulo: Leopardo Editora.

Presidência da República (1993). *Lei 8.666/1993-Institui normas para licitações e contratos da Administração Pública e dá outras providências*. Brasília: Diário Oficial da União.

Christakis, N. and Fowler, J. (2009). *O Poder das Conexões*. 1st ed. Rio de Janeiro.

Neo4j. (2018). San Francisco: Neo4j Corporation.

Neo4j, OpenCypher (2019). *Cypher Query Language Reference*. Opencypher.org. Disponível em: <http://www.opencypher.org/resources>

Generación automática de palabras clave para monitorizar dominios en redes sociales

Miguel Ángel Rozalén Soriano¹ y Álvaro Aranda Jiménez²

¹ ORCID [0000-0002-5897-9018](https://orcid.org/0000-0002-5897-9018). Departamento Knowledge Reuse, Universidad Carlos III de Madrid, España.

miguel.rozalen@kr.inf.uc3m.es

² ORCID [0000-0003-3859-1574](https://orcid.org/0000-0003-3859-1574). Departamento Knowledge Reuse, Universidad Carlos III de Madrid, España.

alvaro.aranda@kr.inf.uc3m.es

Resumen. En la última década las redes sociales se han convertido en uno de los pilares básicos de los canales de comunicación en el cual sus usuarios participan generando contenido de todo tipo e interactuando entre ellos. En el mundo de los medios de comunicación, esta información proveniente de usuarios de redes sociales puede ser una fuente directa de generación de noticias, incluyendo, además de textos, imágenes o videos.

Para poder identificar este tipo de sucesos en tiempo real desde las redes sociales, serán necesarias herramientas sobre las que apoyarse para poder monitorizar las mismas. Dificilmente va a ser posible escuchar, o más bien entender, todo lo que sucede a la vez, pero sí surge la posibilidad de monitorizar sucesos para dominios concretos previamente definidos.

El objetivo tratado en esta publicación será ofrecer un método de generación de la información necesaria, y de manera automática, para poder ‘escuchar’ las redes sociales como monitorización de un dominio concreto. Maximizando, a su vez, la cantidad de palabras de calidad con las que obtener la mayor cantidad de recursos útiles para un dominio de interés. Teniendo como concepto básico la generación y expansión de palabras clave, y la posible aplicación a cualquier otro dominio.

Palabras clave: Redes sociales, monitorización temática, expansión de palabras clave, autogeneración, social media.

Abstract. In the last decade, social networks have become one of the basic pillars of communication channels in which their users participate by generating content of all kinds and interacting with each other. In the world of media, this information from users of social networks can be a direct source of news generation, including, in addition to texts, images or videos.

In order to be able to identify this type of events in real time from social networks, tools will be needed to be supported in order to monitor them. It will

hardly be possible to listen to, or rather understand, everything that happens at the same time, but the possibility of monitoring events for previously defined concrete domains does arise.

The objective addressed in this publication will be to offer a method of generating the necessary information, and automatically, to be able to 'listen' to social networks as a monitoring of a specific domain. Maximizing, in turn, the quantity of quality words with which to obtain the most useful resources for a domain of interest. Having as a basic concept the generation and expansion of keywords, and the possible application to any other domain.

Keywords: Social networks, thematic monitoring, keyword expansion, self-generation, social media.

1 Problema

En la última década las redes sociales se han convertido en uno de los pilares básicos de los canales de comunicación en el cual sus usuarios participan generando contenido de todo tipo e interactuando entre ellos. Esta situación proporciona de forma continua una ingente cantidad de información que se puede analizar para estudiar su impacto y su posible aplicación a multitud de ámbitos. Resulta un campo de estudio multidisciplinar, pudiendo afrontar su análisis desde distintos puntos de vista y perfiles, tratando de extraer conocimiento de la información generada.

En el mundo de los medios de comunicación, se ha comprobado la necesidad de disponer de una política de gestión de medios de difusión digitales. Esta información proveniente de usuarios de redes sociales puede generar noticias desde varios tipos de publicaciones, incluyendo textos enriquecidos mediante contenidos multimedia como imágenes o videos que directamente reflejan sucesos del mundo real hacia el mundo virtual.

Sin duda, se puede tratar de una nueva forma de periodismo en la cual cualquier persona con un dispositivo móvil puede cubrir los aspectos principales de una noticia, incluso, sin tener intención de ello. Alterando así el flujo de la noticia, provocando su nacimiento en las mismas redes sociales. Es entonces cuando los medios se enfrentan a una oportunidad de reaprovechar esa valiosa información que es generada en tiempo real.

Para poder identificar este tipo de sucesos en tiempo real desde las redes sociales, será necesario algún medio de monitorización las mismas. Difícilmente va a ser posible escuchar, o más bien entender, todo lo que sucede a la vez, pero sí surge la posibilidad de monitorizar sucesos para dominios concretos que previamente se hayan definido.

En este punto es donde nace el objetivo tratado en esta publicación: ser capaz de generar la información necesaria, y de manera automática, para poder “escuchar” las redes sociales monitorizando un dominio concreto. Teniendo presente, además, que, aunque todos los recursos puedan salir bajo un mismo contexto, se puede hablar del

mismo concepto de formas muy diversas, y más en un lenguaje tan abierto e interpretable como es el castellano.

En los casos de estudio de otros autores encontrados y analizados previamente a la realización de esta publicación, la mayoría hacen referencia al uso de aplicaciones de terceros para la escucha de la red social sin hacer mención de ninguna lógica previa sobre qué escuchar o qué no de la misma, solamente un caso comenta el uso de información previamente construida por agrupadores semánticos en el mundo de la medicina.

En esta primera publicación (Rodríguez, 2016), se expone un estudio de apps utilizadas para monitorizar una red social por nombres de usuario o palabras clave, sin hacer mención de un proceso para formar esas palabras clave. En el siguiente artículo (de Gracia, 2010), un poco más antiguo, se expone un arduo trabajo para construir una estructura de monitorización de Twitter, pero tampoco hace mención de elaboración específica orientada en la monitorización temática. En este otro ejemplo (Montañés, 2018), tenemos un caso de monitorización, pero, de un único usuario en distintas redes sociales.

Por último, la única publicación (Sarker, 2016) con una aproximación cercana a nuestro experimento está orientada al mundo de la medicina donde se quiere monitorizar el abuso de la prescripción médica. En ella presentan el uso de agrupadores de palabras previamente contruidos, y el posterior uso de clasificadores de técnicas de aprendizaje automático para ampliar y terminar determinando las palabras clave a seguir.

2 Marco conceptual

El caso está basado para el uso de la red social Twitter. En el proceso de ‘escucha’ para la obtención de los recursos que se generan en dicha red, se le debe indicar hasta un máximo de 400 palabras que deben ser seguidas. La obtención de esas palabras que se le indican a la red social son clave para obtener resultados de calidad, y más aún al querer monitorizar un dominio concreto. No todos utilizamos las mismas expresiones para referirnos a los mismos conceptos, y más aún en textos escritos en redes sociales, en las cuales se llegan a utilizar, y estandarizar como uso común, términos que ni si quiera aún están recogidos en el diccionario.

Así pues, en esta comunicación se describirá el procedimiento llevado a cabo para maximizar la cantidad de palabras de calidad con las que se pueda poner a ‘escuchar’ la red social con el objetivo de obtener la mayor cantidad de recursos útiles para un dominio de interés. Teniendo como concepto básico la generación y expansión de palabras clave, y la posible aplicación a cualquier otro dominio.

Dicho proceso deberá tener en cuenta que no se escribe de igual forma en medios formales de comunicación que en redes sociales, donde cada usuario puede escribir libremente teniendo, incluso, faltas de ortografía.

2.1 Implementación tecnológica

En cuanto a la necesidad tecnológica para llevar a cabo el experimento no tiene un coste excesivamente alto. La principal necesidad será el uso de la API que ofrezca la red social con su versión abierta para desarrolladores. En el caso particular de las diferentes librerías del API utilizaremos la que ofrecen en C#, lenguaje de programación de Microsoft. Para interactuar con dicho API, debemos tener una cuenta creada en la red social. A partir de ella, accediendo a la versión para desarrolladores, Twitter generará los tokens necesarios con los que posteriormente permite el acceso al uso del API.

En cuanto a almacenamiento, simplemente con ficheros de texto estándar como salida para almacenar lo producido será suficiente.

Con la lógica en dicho lenguaje de programación para manejar el API, y también la necesaria para la obtención de información de posibles sitios web escogidos tendremos toda la tecnología necesaria.

3 Caso de estudio: accidentes de tráfico en Twitter

Después de hacer una evaluación de redes sociales, se concluyó que la que mejor se ajusta al proyecto para servir como fuente de información en bruto, tanto por sus características técnicas como sociológicas, su manera de estructurar la información, etc, es Twitter. Por lo que el entorno de prueba estará orientado a esta social. Como mencionamos anteriormente presenta una restricción de número de palabras que se pueden seguir, estando además sin normalizar, es decir, que las palabras: ‘niño’ y ‘niña’ serían tan diferentes como ‘hola’ y ‘avión’.

El dominio escogido para la realización del experimento ha sido ‘accidentes de tráfico’. El principal motivo para su elección fueron que son eventos que implican directamente a personas y que, desafortunadamente, suceden con una frecuencia bastante alta.

Haciendo uso de este API el sistema es capaz de recoger la información en bruto, los tweets publicados en tiempo real que están en castellano y contienen palabras relacionadas con el suceso que pretendemos notificar.

El proceso se compone de dos grandes partes. Por un lado, la obtención de los términos formales para la monitorización de los recursos en la red social y, por otro, un proceso con la lógica parecida para obtener aquellos términos claves que puedan ser generados en las redes sociales.

3.1 Paso extracción términos formales

Una vez establecida la temática a monitorizar, es necesario crear un entorno contextual de palabras relevantes para el determinado dominio. Al inicio del proceso que se

va a llevar a cabo, solamente es necesario el nombre de dicha categoría a monitorizar. Sin necesidad de preparar previamente nada en referencia a palabras clave.

En este primer paso se extraerán los términos denominados formales, registrados en el diccionario y escritos correctamente, del contexto que se quiere monitorizar. Mediante la creación de un proceso automático, se recogerá el histórico de noticias registradas en uno de los principales periódicos del país. En nuestro caso utilizaremos el periódico digital ‘El país’, el cual proporciona un buscador a través del cual muestra todas las noticias que tengan registrado en su medio digital para lo que se introduzca como búsqueda. Se puede acceder mediante el siguiente vínculo web: <https://elpais.com/buscador/>

Una vez obtenidas todas las noticias del periódico que contengan la palabra del dominio seleccionado, serán almacenadas en formato texto quedando recogidas para que posteriormente se pueda procesar su contenido. Después, se recorrerán los textos almacenados leyendo palabra por palabra. Durante esta lectura automatizada se llevará a cabo un conteo de palabras para la construcción de un índice de frecuencias de aparición de cada palabra contenida en todas las noticias que fueron extraídas.

Con dicho índice de frecuencias construido, podremos tener una idea de cuáles son las palabras formales que más se han utilizado en la redacción de las noticias que se produjeron referentes a la temática seleccionada. Como último paso, se evaluará cada palabra mediante otro procedimiento automático el cual será capaz de indicarnos a qué categoría gramatical pertenece cada una de las palabras que han sido registradas.

Tras ello, quitaremos de nuestro listado de palabras todas aquellas que su categoría gramatical sea: determinante, pronombre, adjetivo, adverbio, preposición, conjunción o interjección; quedándonos solamente con aquellas que sean verbos o sustantivos. Esta lista final será la considerada como términos formales clave del entorno contextual para el dominio seleccionado.

3.2 Paso extracción términos red social

Llegados a este punto, tendremos clasificados por orden de relevancia los términos formales obtenidos; en un segundo paso se persigue obtener la lista de términos más utilizados en redes sociales para extraer aquellos términos que nos produzcan mejores resultados en la obtención de recursos útiles publicados. A la hora de utilizar términos y expresiones referentes a un mismo dominio, no tiene nada que ver los términos que se puedan utilizar en un periódico, el cual contiene noticias formalmente redactadas, con los términos que se utilizan en las redes sociales, pudiendo ser éstos totalmente anárquicos estando incluso escritos de forma gramaticalmente incorrecta.

En esta fase, vamos a transformar el listado de términos formales anteriormente construido en un listado de palabras de uso frecuente en redes sociales. En este caso no disponemos de un medio para buscar todos los recursos publicados en una red social para una temática concreta, como ocurría con el periódico digital. Para llevar a

cabo esta construcción, será necesario preparar otro proceso automático para ‘escuchar’ la red social. Como comentamos anteriormente, necesitamos indicarle a la red social de qué términos queremos que nos lleguen recursos publicados en ella. Es aquí donde se muestra clave el resultado del procedimiento llevado a cabo en el paso anterior de extracción de términos formales. De dicha lista de términos formales, pondremos a escuchar la red social con los 400 términos con mejor índice de frecuencia, sea cuales sean sin hacer ningún análisis intermedio. De esta manera, queda enlazado este proceso automático para analizar los términos de redes sociales con el procedimiento automático anterior que se iniciaba solamente con el título de una categoría seleccionada.

Una vez preparado el proceso para realizar la ‘escucha’ de la red social, será necesario dejarlo corriendo y almacenando los recursos que nos proporcione hasta tener un número suficientemente representativo de ellos. En nuestro caso, consideramos 100.000 recursos una muestra considerable de la cual poder obtener términos y expresiones utilizados en las redes sociales. En los experimentos utilizados llegar a ese número de recursos nos ha supuesto del orden de dos o tres días de escucha de la red social, el cual consideramos un orden de magnitud de tiempo asumible.

Tras la obtención de los 100.000 recursos extraídos de las redes sociales en base a los 400 términos formales con mayor índice de frecuencia de aparición, repetiremos el análisis realizado en la fase anterior para analizar los nuevos textos obtenidos. Lo que significa, leer cada una de las palabras de todos los textos almacenados, e ir construyendo durante esa lectura el índice de frecuencias de aparición de las palabras para posteriormente quedarnos solo con aquellas palabras que tengan como categoría gramatical sustantivo o verbo.

Finalizado este proceso, obtendremos una nueva clasificación de palabras donde podremos saber cuáles son las que más han sido utilizadas, pero esta vez basado en análisis de textos exclusivamente publicados en redes sociales. Si uno de los términos formales obtenidos en el paso anterior fuese importante en las redes sociales, tendría un reflejo como tal en las mismas y también aparecería como un término destacado en los textos obtenidos de ellas.

De este modo, seleccionando las palabras con mayor índice de aparición, podremos construir, con la mayor calidad posible, nuestra lista final de 400 términos clave de ‘escucha’ de la red social. Siendo cada una de las palabras que componen la lista, contrastadamente importante a la hora de obtener publicaciones respecto al dominio seleccionado.

4 Validación

Tras la implementación anterior, ya estamos preparados para monitorizar la temática escogida y en colaboración con RTVE se puso en marcha un sistema de escucha con el que se pudo evaluar el funcionamiento del procedimiento descrito.

El periodo de validación tuvo lugar durante un tiempo de 6 meses, incluyendo desde enero hasta junio del año 2018. Durante ese tiempo, RTVE recogía todas las notificaciones que le llegaban de la agencia de noticias española ‘EFE’, una de las principales del país, y posteriormente se contrastaba la información con los recursos recogidos de la red social.

En la validación se tenía en cuenta si el sistema de monitorización de la red social había registrado el accidente o no, y en caso afirmativo con cuánta diferencia horaria respecto a la noticia lo había recogido.

Los resultados finales fueron los siguientes:

Tabla 1. Resultados comparativa alertas agencia EFE contra alertas recogidas en redes sociales.

Alertas EFE	Alertas RRSS	Detectadas antes en RRSS	Antelación mayor una hora
67	47	23	15
(100%)	(70.15%)	(34.33%)	(22.39%)

Se consideran unos resultados generales bastantes satisfactorios. Se capturaron con éxito un 70% de los eventos publicados en la agencia de noticias. Teniendo el 51% de ellos con anterioridad en nuestros sistemas, y el 63% de los registrados antes su antelación fue superior a una hora.

Referencias

Abeed Sarker, Karen O'Connor, Rachel Ginn, Matthew Scotch, Karen Smith, Dan Malone, Graciela Gonzálz. (2016). *Social Media Mining for Toxicovigilance: Automatic Monitoring of Prescription Medication Abuse from twitter*. Disponible en: <https://link.springer.com/content/pdf/10.1007%2Fs40264-015-0379-4.pdf>.

Lina Rodríguez de la Cruz, F., Morcelo Ureña, A., de la Morena Taboada, M. (2016). *Apps de monitorización como eje esencial en estrategias de social media en twitter*. Disponible en: https://www.researchgate.net/publication/309726426_Apps_de_monitorizacion_como_eje_esencial_en_estrategias_de_social_media_en_twitter.

Gracia-Soriano. P. de (2010). *La monitorización de Twitter como técnica para el estudio de la sociología del tiempo*. Disponible en: <https://www.fes-sociologia.com/files/congress/12/papers/5746.pdf>.

Montañés, R., Aznar, R., Nogueras, S., Segura, P., Langarita, R., Meléndez, E., Peña, P., Hoyo, R. del. (2018). *Monitorización de Social Media*. Disponible en: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2018-61-26/3399>.

Editora UFPB: uma história fundamentada em dados

Geisa Cavalcante¹ y Guilherme Dias²

¹ ORCID [0000-0002-0905-9012](https://orcid.org/0000-0002-0905-9012). Editora UFPB, Universidade Federal da Paraíba, Brasil.
gfffcavalcante@gmail.com

² ORCID [0000-0001-6576-0017](https://orcid.org/0000-0001-6576-0017). Departamento de Ciência da Informação, Universidade Federal da Paraíba, Brasil.
guilhermeataide@gmail.com

Resumo. Objetiva investigar a história de uma editora universitária, a partir dos dados bibliográficos dos seus livros. Caracteriza-se como pesquisa exploratória, de cunho qualitativo: pesquisa documental para coleta de dados e uso dos princípios referentes à teoria fundamentada e ao *big data* para análise dos dados. Os aspectos analisados foram organizados nas dimensões: volume, diversidade e qualidade, o cruzamento desses eixos possibilitou caracterizar a linha do tempo da Editora UFPB, onde, observam-se os seguintes períodos históricos: 1962 a 1975 - Imprensa Universitária da Paraíba; 1975 a 1981 - De imprensa à editora; 1981 a 1993 - Grande recessão; 1993 a 2001 - O retorno; 2001 a 2009 - Síndrome do livro pronto; 2009 a 2014 - Fênix. É sabido que o volume de dados analisados não foi suficiente para traçar um panorama definitivo, a avaliação de novos conjuntos de dados poderá levar a novos caminhos, possibilitando a descoberta de novos conhecimentos a partir da teoria fundamentada.

Palavras chave: memória institucional; teoria fundamentada; *big data*; editoras universitárias.

Abstract. It aims to investigate the history of a university publisher, based on the bibliographic data of its books. It is characterized as an exploratory research, of a qualitative nature: documentary research for data collection and use of the principles concerning the grounded theory and the big data for data analysis. The analyzed aspects were organized in the dimensions: volume, diversity and quality, the intersection of these axes made it possible to characterize the Editora UFPB time line, where the following historical periods are observed: 1962 to 1975 - University Press of Paraíba; 1975 to 1981 - Press to publisher; 1981 to 1993 - Great recession; 1993 to 2001 - The return; 2001 to 2009 - Book Syndrome Ready; 2009 to 2014 - Phoenix. It is known that the volume of data analyzed was not enough to draw a definitive picture, the evaluation of new datasets could lead to new paths, enabling the discovery of new knowledge from the grounded theory.

Keywords: institutional memory; grounded theory; big data; university publishers.

1 Contextualização

O desenvolvimento das tecnologias de informação e comunicação e o aumento da capacidade de processamento de dados levou os pesquisadores a um cenário em que já se torna possível a análise do universo de dados, caracterizando a era do *big data* (Mayer-Schonberger & Cukier, 2013).

Para o trato do universo de dados sugere-se a adoção de métodos quantitativos de pesquisa. No entanto, o desenvolvimento de outros métodos de pesquisa, como a teoria fundamentada, torna possível que o grande volume de dados também seja analisado pelo viés qualitativo.

A teoria fundamentada propõe o desenvolvimento de uma teoria durante a pesquisa, esta teoria deverá provir da análise dos dados coletados, ao invés de advir da fundamentação teórica.

O *big data* e a teoria fundamentada abrem, pois, um novo horizonte de possibilidades de pesquisa. Diante desse cenário, pergunta-se: o que os dados bibliográficos dizem a respeito de uma editora universitária?

Considera-se, aqui, as universidades como centros de pesquisa e desenvolvimento, as editoras universitárias como responsáveis pela publicidade do conhecimento produzido através das pesquisas realizadas em sua Instituição de Ensino Superior (IES) (Bufrem & Garcia, 2014) e as bibliotecas como instituições tradicionais de memória (Pessach, 2008).

Acredita-se, pois, na possibilidade de recuperação da memória institucional de uma editora universitária a partir da exploração dos seus dados que constem no acervo do sistema de informação de bibliotecas de sua universidade de origem.

A guarda, a preservação e a recuperação da memória institucional é fonte de registro da aprendizagem organizacional por tornar possível o aprendizado com o passado, garantindo a melhoria contínua da organização. Dessa forma, o resgate material deixa de ter o papel de celebração de marcos históricos, passando a ser elemento constitutivo do futuro ao proporcionar o aprendizado com o passado e com a atualidade (Barbosa, 2013).

Costa e Saraiva (2011) ratificam esta perspectiva ao afirmarem que a memória formalizada possibilita reelaborar realidades, transformando-as e sustentando-as. Descrevem, também, a existência de uma hierarquização das memórias institucionais, sendo as memórias oficiais apenas as que refletem a ideologia defendida pela instituição, opondo-se a estas as demais memórias. Daí a importância do resgate da memória em seu contexto completo e complexo, a fim de se obter um retrato mais fidedigno da instituição estudada.

2 Objetivo

Visando apreender o que os dados revelam em sua totalidade para construção da teoria, deve-se ir a campo apenas com o problema de pesquisa e o objetivo geral, sem aprofundamento teórico, para que se evite uma concepção prévia do que pode ser encontrado (Glaser, 1978 citado por Charmaz, 2009). Dessa forma, é objetivo geral deste estudo: investigar a história de uma editora universitária, a partir dos dados bibliográficos dos seus livros publicados.

3 Metodologia

Antes de caracterizar esta pesquisa, faz-se necessário descrever o percurso metodológico utilizado. De início, intencionava-se que esta pesquisa envolvesse uma pesquisa bibliográfica sobre editoras universitárias, pesquisa documental sobre a Editora UFPB e a análise dos dados extraídos do Sistema de Automação de Bibliotecas (SAB) sobre os títulos da Editora UFPB constantes no sistema de bibliotecas da Universidade Federal da Paraíba (UFPB).

Antevia-se a existência de certas dificuldades para análise desses dados em razão dos diversos nomes utilizados pela Editora UFPB ao longo de sua história e pelas diversas formas de cadastro dos mesmos no SAB. No momento de recebimento dos dados, as dificuldades foram maiores do que as previstas, uma vez que não era possível distinguir quais títulos, de fato, pertenciam à editora em questão. Decidiu-se, então, consultar os metadados dos livros no site da Agência Brasileira de ISBN, a fim de corrigir as discrepâncias encontradas. No entanto, descobriu-se que vários dados dos livros, em conformidade com o que se havia extraído do SAB, não correspondiam aos metadados cadastrados na Agência Brasileira de ISBN, configurando-se como dados inconsistentes, conforme a classificação de Castro e Ferrari (2016).

Um dos métodos para a resolução dos casos de dados inconsistentes é a análise manual auxiliada por rotinas específicas (Castro & Ferrari, 2016). Diante disso, priorizou-se a consulta dos livros físicos, a fim de confirmar os dados em sua fonte original. A Editora UFPB possui 3.039 títulos e 13.228 exemplares no sistema de bibliotecas da UFPB, o que impossibilitaria a consulta ao universo de dados. Limitou-se, então, a consulta dos títulos depositados na Coleção Autores Paraibanos da Biblioteca Central da UFPB.

A Coleção Autores Paraibanos destina-se à guarda memorial de obras publicadas por paraibanos. O fato de essa ser a naturalidade da maioria dos autores da Editora UFPB, faz com que grande parte do acervo de títulos publicados por esta editora também estejam depositados nesta coleção, o que justifica a sua escolha.

No processo de consulta à fonte original, foram capturadas, aproximadamente, 7.000 fotos dos títulos, referentes à capa, falsa folha de rosto, folha de rosto e colofão. Tais fotos compreendiam todos os títulos da Editora UFPB depositados na Coleção Autores Paraibanos, a pesquisa segue em andamento, constando, aqui, a análise de

300 desses títulos. No momento de seleção dos 300 títulos, foram priorizados os títulos de coleções ou selos editoriais, além de publicações realizadas sobre a UFPB e títulos que tratassem de temas de rara publicação ou em língua estrangeira.

Dessa forma, esta pesquisa caracteriza-se como pesquisa exploratória, executada por métodos qualitativos que consistiu no uso da pesquisa documental para coleta de dados e uso dos princípios referentes à teoria fundamentada e ao *big data* para análise dos dados.

Para construção de uma teoria fundamentada completa, Glaser (1978, 1992 citado por Charmaz, 2009, p. 19) e Glaser e Strauss (1967 citado por Charmaz, 2009, p. 19) é necessário o atendimento aos critérios: ajuste adequado aos dados, utilidade, densidade conceitual, durabilidade ao longo do tempo, suscetível a alterações e detentor de poder explicativo.

Para garantir o ajuste adequado aos dados foram realizadas mudanças de rumos e métodos, a partir das descobertas encontradas na análise dos dados. A utilidade da pesquisa constitui-se por proporcionar o conhecimento do passado desta instituição, de modo que seja possível entender o seu presente e redefinir o seu futuro. A densidade conceitual, por sua vez, é proporcionada pelas diversas dimensões que puderam ser observadas através dos dados: volume editorial, diversidade editorial e qualidade editorial.

A durabilidade ao longo do tempo reflete-se no cuidado de analisar documentos de diversas gestões organizacionais diferentes, a fim de evitar que as relações de poder prejudicassem a análise (Le Goff, 1990), além da análise das facetas citadas acima. A suscetibilidade de alterações garante-se na construção contínua da memória organizacional e na possibilidade de outras descobertas ao serem analisadas a totalidade dos dados. Já o poder explicativo é garantido pelo entrelaçamento dos fatos que marcaram a história da Editora UFPB e que ficaram registrados nos títulos que compõem a sua produção editorial.

A ótica do *big data*, por sua vez, foi utilizada ao se investigar o maior número de possibilidades advindas dos dados coletados, compreendendo o que tais informações representavam para a Editora UFPB e sua história, desde a avaliação de seu volume editorial à identificação dos ciclos temporais que marcaram a sua história, convertendo o volume de dados em inovação, conhecimento e valor, através da estruturação e análise dos dados (Davenport, 2014), obtendo finalidade através desses dados (Mayer-Schonberger & Cukier, 2013) e utilizando não apenas a finalidade inicial, mas também, as finalidades secundárias descobertas (Mayer-Schonberger & Cukier, 2013).

4 Apresentação de resultados

A partir das fotografias coletadas dos 300 títulos selecionados para análise nessa dissertação, foram catalogados 7.128 dados. Além dos dados advindos das análises

das fotos coletadas, os mesmos foram cruzados com os metadados disponíveis na Agência Brasileira de ISBN e com os dados extraídos do SAB.

Utilizou-se a codificação linha a linha para catalogação dos dados e a sua categorização foi realizada em conformidade com as de teoria fundamentada defendidas por Strauss e Corbin (1990 citados por Charmaz, 2009), utilizando a abordagem objetivista da teoria fundamentada. Foi possível analisar os mais diversos aspectos sobre a história da Editora UFPB, organizados nas vertentes: volume editorial, diversidade editorial e qualidade editorial.

O demonstrativo de volume editorial retrata a quantidade de publicações realizadas pela Editora UFPB. Os anos encontrados com maior número de publicações foram 1977 (17), 1979 (20), 1980 (20), 2012 (21) e 2013 (24). Os períodos de maior volume editorial ocorreram entre 1975 e 1983, 1995 e 2003 e entre 2005 e 2014, sendo o maior volume alcançado entre 2012 e 2014.

A vertente diversidade ilustra as diversas áreas em que a Editora UFPB tem publicado, comprovando que a sua produção nunca esteve voltada para uma área em específico, mas que sempre buscou atender às mais diversas pesquisas da UFPB. Os períodos de maior diversidade são localizados temporalmente próximos aos períodos de maior volume. Descobriu-se, também, que a Editora, por meio dos seus títulos, atua na promoção ao desenvolvimento: científico, social e da UFPB, enquanto instituição de ensino superior, compreendendo a sua influência na comunidade em que se encontra inserida.

A dimensão qualidade, por sua vez, avalia a qualidade editorial dos títulos analisados através dos fatores: presença de Conselho Editorial e número de erros cometidos no processo de edição. Quanto ao fator Conselho, observou-se sua presença com maior frequência nos períodos entre 1975-1982, 1995-2001, 2009-2014. Quanto ao fator erros editoriais, constatou-se que, os períodos de crescimento do volume editorial também são os períodos com maior frequência de erros, havendo um crescimento mais significativo a partir do início de uso do número de ISBN.

Através do cruzamento dos dados de volume editorial, diversidade editorial e qualidade editorial, foi possível traçar uma linha do tempo da Editora UFPB, conforme apresentado na Figura 1.

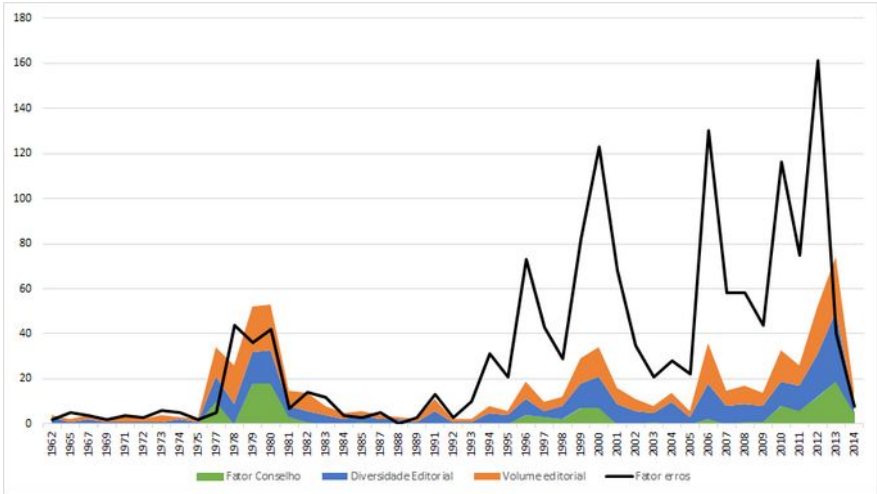


Figura 1. Gráfico da linha do tempo da Editora UFPB

A análise do gráfico torna possível a identificação de períodos históricos distintos na trajetória da Editora UFPB, caracterizados em conformidade com as descrições que seguem.

1962 a 1975 - Imprensa Universitária da Paraíba: representada visualmente na Figura 2.

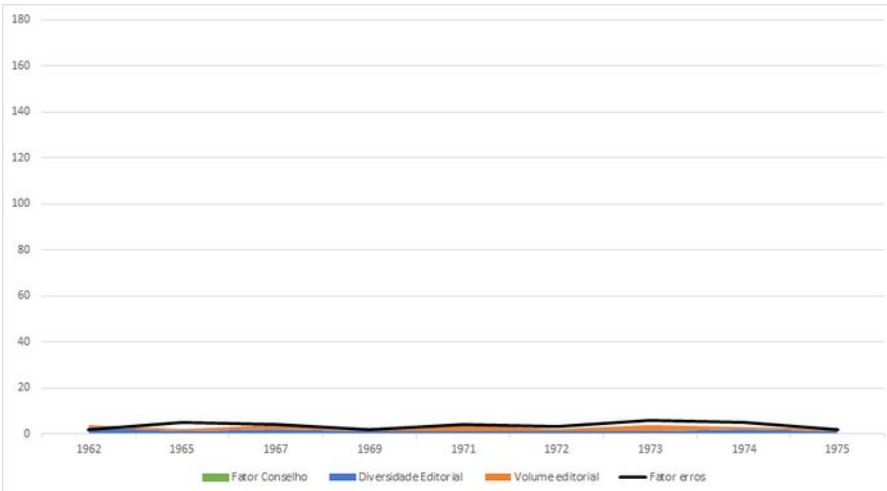


Figura 2. Imprensa Universitária da Paraíba

A Editora UFPB surge enquanto imprensa universitária, mas ainda apresenta baixos valores de volume e diversidade editorial. Ainda não apresenta publicações com Conselho Editorial.

1975 a 1981 - De imprensa à editora: cuja representação encontra-se na Fig 3.

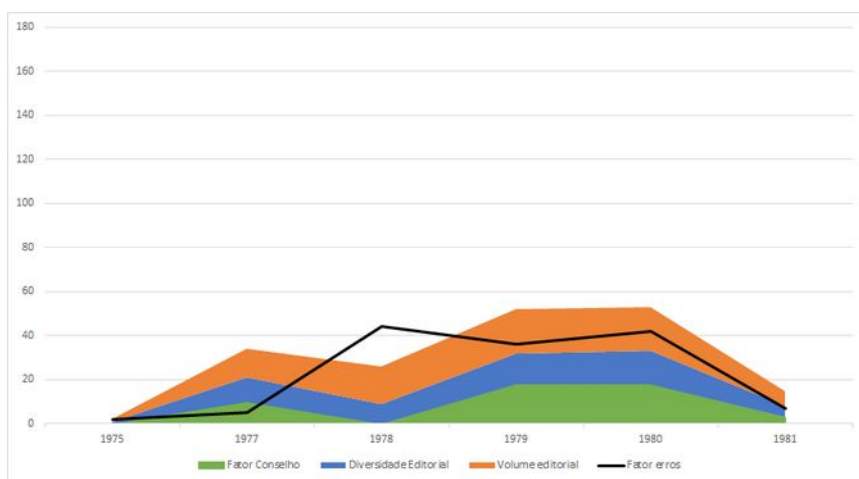


Figura 3. De imprensa à editora

A Editora UFPB apresenta sua primeira onda de crescimento, com a publicação das coleções “Estudos Universitários”, “Textos Didáticos”, além de várias outras publicações a respeito da própria UFPB. Apresenta suas primeiras publicações com Conselho Editorial, mas o crescimento de volume, também provoca o crescimento no número de erros editoriais.

1981 a 1993 - Grande recessão: retratada na Figura 4.

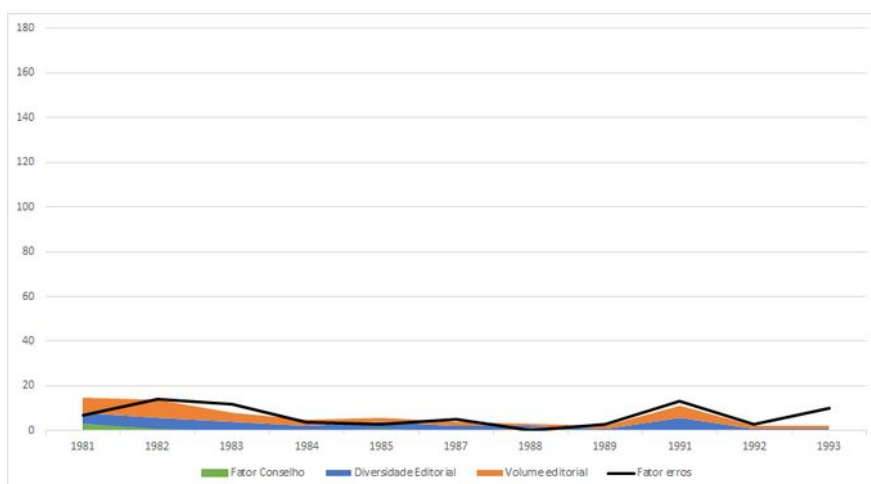


Figura 4. Grande recessão

Com exceção dos períodos entre 1981-1983 e 1989-1992, a Editora apresenta uma redução significativa no volume e diversidade editorial, também volta a apresentar publicações sem Conselho Editorial. A queda em volume, proporciona a queda no número de erros, com exceção do período entre 1992-1993, onde foi adotado o seu primeiro prefixo editorial.

1993 a 2001 – O retorno: caracterizado na Figura 5.

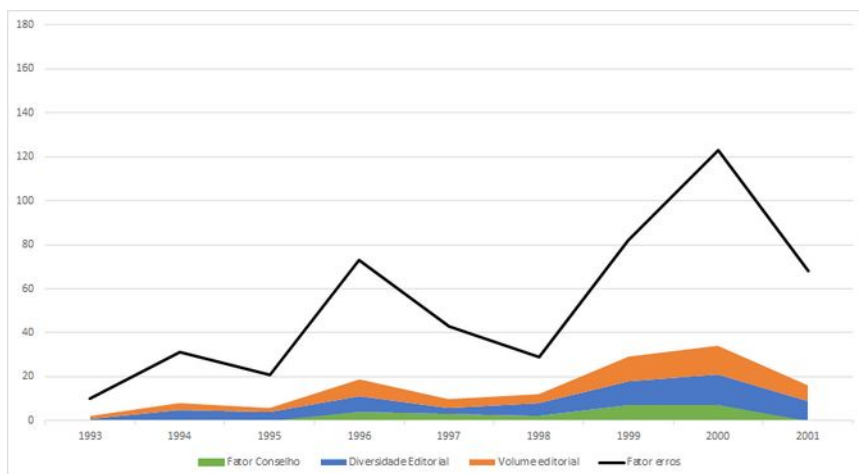


Figura 5. O retorno

Apesar de não alcançar os índices obtidos no período entre 1975-1981, a Editora UFPB volta a apresentar crescimento de volume e diversidade editorial, volta, também, a publicar títulos com Conselho Editorial. O aumento do volume provoca o crescimento, de modo drástico no número de erros editoriais. Nesse período são publicadas a “Série Sala de Aula”, a “Série Extensão” e a “Coleção Novos Autores Paraibanos”.

2001 a 2009 - Síndrome do livro pronto: simbolizada na Figura 6.

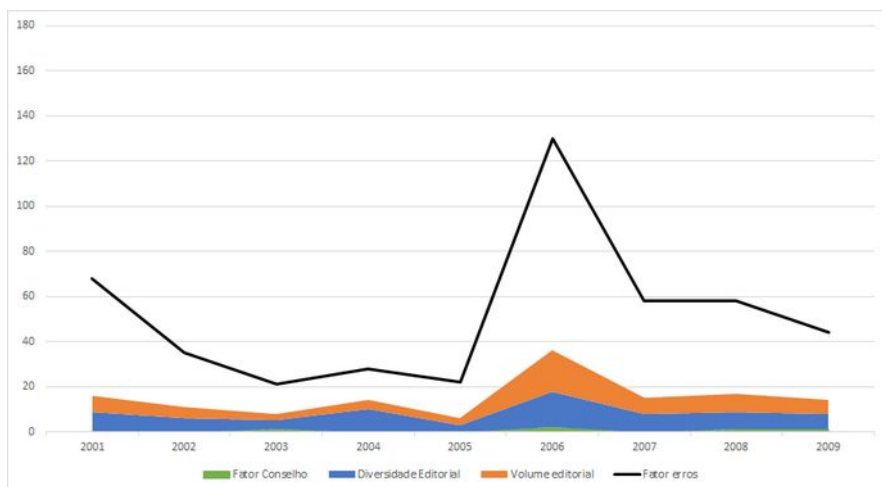


Figura 6. Síndrome do livro pronto

A Editora UFPB apresenta grave oscilação de volume e diversidade editorial, os níveis de erros seguem em proporções elevadas e é registrado mais um período de ausência de Conselho Editorial. É, também, o período de surgimento do “Selo Autor Associado”.

2009 a 2014 – Fênix: representada visualmente na Figura 7.

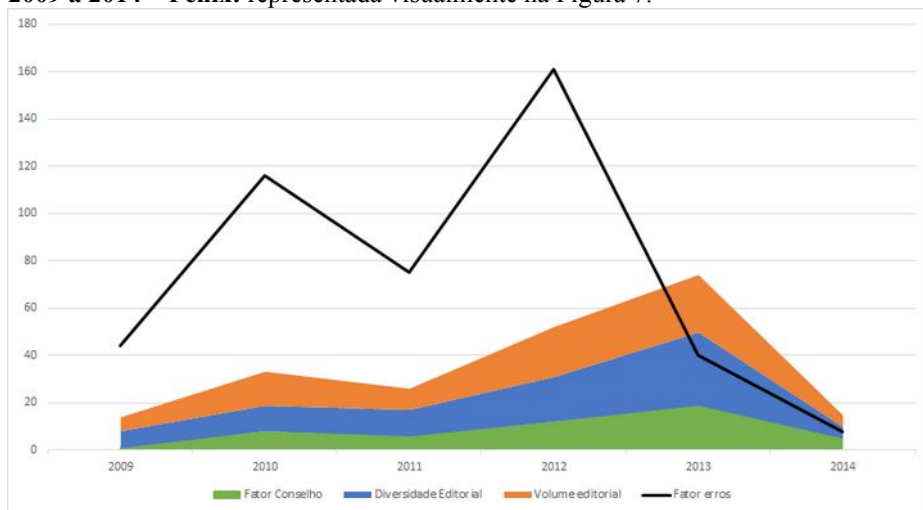


Figura 7. Fênix

Assim como na mitologia grega, a Editora UFPB parece ressurgir das cinzas no período, volta a publicar com Conselho Editorial, cresce em volume e diversidade e,

no período entre 2012-2014, alcança os maiores níveis de produção editorial de sua história, além de apresentar uma queda significativa no número de erros editoriais. É o período em que é publicada a “Coleção Humanidades”.

Observa-se que a publicação de Coleções e Séries têm influência positiva sob o volume e a diversidade editorial; os principais momentos de crescimento do volume e da diversidade ocorreram com a presença do Conselho Editorial. Por fim, nota-se como nítido ponto fraco da Editora UFPB, a numerosa quantidade de erros editoriais encontrados, sendo necessária a tomada de providências cabíveis a respeito, como a profissionalização da revisão final de edição, catalogação e cadastro dos metadados na Agência Brasileira de ISBN, que demonstrou ser o maior gargalo encontrado.

Dessa forma, a análise dos dados emergiu na caracterização histórica da Editora UFPB, tendo sido possível entender melhor suas fases, desafios e tipos de publicações realizadas. Sabe-se que, tal caracterização pode (e deve) ser aprofundada ao se considerar o universo de publicações da Editora UFPB, mas este já é um grande passo para entendimento de seu histórico e do que constitui o seu diferencial competitivo.

É sabido que o volume de dados analisados não é suficiente para traçar um panorama definitivo sobre a história da Editora UFPB, a avaliação de novos dados poderá levar há novos caminhos, mostrando que esta teoria é passível de alterações, conforme Glaser e Strauss (1967 citado por CHARMAZ, 2009).

Quanto à possibilidade de desenvolvimento de novas pesquisas, além da conclusão desta com maior número de títulos ou com o universo do acervo, sugere-se que seja realizada uma análise semiótica da evolução dos projetos editoriais desta casa editorial, assim como, da evolução de sua marca ao longo dos anos, a fim de que outros aspectos sobre o histórico da Editora UFPB possam ser analisados e registrados. Também pode ser sugeridas novas pesquisas que utilizem a teoria fundamentada enquanto método de análise dos conteúdos disponibilizados pelo *big data*.

Referencias

- Barbosa, A. A. Memória institucional: possibilidade de construção de significados no ambiente organizacional. (2013) *9o Encontro Nacional de História da Mídia*. Ouro Preto. Disponível em: <http://www.ufrgs.br/alcar/encontros-nacionais-1/9o-encontro-2013/artigos/gt-historiografia-da-midia/memoria-institucional-possibilidade-de-construcao-de-significados-no-ambiente-organizacional>.
- Bufrem, L. S. & Garcia, T. M. B. A editora universitária e o compromisso da universidade com as práticas de divulgação do conhecimento produzido. (2014) *Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*. Porto Alegre, vol. 20, nº1, Jan./Jun. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/40816/32989>
- Charmaz, K. (2009) *A construção da teoria fundamentada: um guia prático para análise qualitativa*. Porto Alegre: Artmed.

- Costa, A. de S. M. da. & Saraiva, L. A. S. Memória e formalização social do passado nas organizações. (2011) *Rev. Adm. Pública [online]*. vol. 45, nº6, Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122011000600007&lng=en&tlng=en
- Davenport, T. H. (2014) *Big data no trabalho: derrubando mitos e descobrindo oportunidades*. 1. ed. Rio de Janeiro: Elsevier.
- Le Goff, J. (1990) *História e memória*. Campinas: Editora da UNICAMP.
- Mayer-Schonberger, V. & Cukier, K. (2013) *Big Data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. Rio de Janeiro: Elsevier.
- Pessach, G. (2008). Memory Institutions: Social Remembering, Privatization and its Discontents. *Cardozo Arts & Entertainment Law Journal, Forthcoming*. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1085267

An approach to the Contextual Design methodology in the context of Information Science

Luís Miguel Oliveira Machado¹, André Pacheco²

¹ORCID [0000-0003-3403-5618](https://orcid.org/0000-0003-3403-5618). Faculdade de Letras da Universidade de Coimbra, Portugal.
luismmachado70@gmail.com

²ORCID [0000-0002-1810-4866](https://orcid.org/0000-0002-1810-4866). Faculdade de Letras da Universidade de Coimbra, Portugal.
andrez.pacheco@gmail.com

Abstract. Based on an operative definition of Information Science (IS) as a study of the mediation aspects of the phenomena: data, information, knowledge and message, and their implementation in the cultural domain, the potential of Contextual Design (CD) as a methodology for the development of optimized interfaces between information resources and users is analyzed. The goals are (i) to identify papers that address or make use of the CD methodology in databases related to IS; (ii) to describe the approaches to the concept of CD in those papers; (iii) to point out the potentials and limitations of this methodology. We draw an exploratory / descriptive study based on a systematic review and categorical analysis. Eight papers were retrieved, which may act as a potential indicator of the reduced use of this methodology in the IS area. Furthermore, its application is essentially practical. Although the limitation of non-involvement of users throughout all the design process is pointed out, the CD methodology shows the ability to collect implicit information through user activities and the possibility of shared systematization of complex data. These characteristics make CD a potential value for the development of appropriate interfaces between information resources and their users.

Keywords: contextual design; information science; user-centered methodology; informational resources.

1 Introduction

Despite a certain difficulty in the delimitation of the field of knowledge designated by Information Science (IS), or perhaps precisely due to that difficulty, several authors have engaged in its epistemological study (Araújo, 2013; Barreto, 2008; Buckland, 2012; Capurro, 2003; Le Coadic, 1996; Machado, Simões, & Souza, 2017; L. V. R. Pinheiro, 2006; Saracevic, 2009; A. M. da Silva & Ribeiro, 2008; J. L. C. Silva & Freire, 2012; Souza & Almeida, 2009; Zins, 2006). In the 1970s, Shera and Cleveland (1977, p. 260) accounted for the many approaches regarding the origin of IS. The further increase of studies of this nature in the following years have led some authors

to consider IS as one of the most introspective scientific fields (Souza, Almeida, & Baracho, 2013).

Amongst the several works that attempt to provide a better understanding of the field, we highlight a study by Chain Zins that included the contributes of an international panel formed by 57 researchers working in IS, which presents a mainstream vision of the field at the beginning of the 21st century (Zins, 2007a, 2007d, 2007c, 2007b).

In an effort to relate this vast number of contributions towards an agreement on the epistemology of the field, a definition of IS is suggested as a scientific field that studies the mediation aspects related to the phenomena of data, information, knowledge and message (D-I-K-M), and their implementation in the cultural domain, which is seen as the several ways in which society faces reality. Therefore, the area of activity of IS lies at the intersection between the D-I-K-M phenomenona and the cultural domain or, more specifically, in the components that mediate users' information resources (cf. Figure 1).

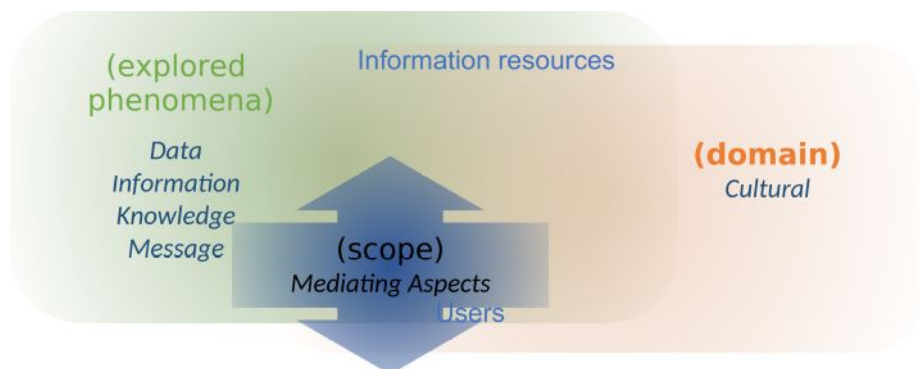


Figure 1. Schematic representation of the area of activity of IS, according to the definition suggested in this study (elaborated by the authors).

The option to include the D-I-K-M phenomena in the pole identified ad-hoc as *information resources* reflects a broad conception of IS, as expressed by Buckland in (Zins, 2007b, p. 336) that includes the four phenomena, or in (Buckland, 2012) where he stresses the cultural aspect of the field. It is important to point out that this option should not be taken as an undifferentiated use of these concepts, a topic highly pertinent to IS (Marques, 2017, p. 64). It is emphasized, instead, the intrinsic relation shared by the distinct concepts, as stated by Silva and Ribeiro: “*Information is distinct, although not separated, from both knowledge and communication*” (Silva & Ribeiro, 2008, p. 48, italic in the original). Similarly, Marques (2015, p. 58) relates *information* with *message* as part of a *communicative* process from which results *knowledge*.

Considering this perspective, we adopt a position similar to Zins (2007c, p. 487), situating the D-I-K-M phenomena in the internal and external context of an individual

and, as such, able (as artifacts) to undergo a mediation between them and the user. It is stressed the provisory nature of this position, since the complexity and controversy of this topic¹ far exceeds the framework of the current study. As a result, considering the previous discussion and the goals of this study, we adopt a similar interpretation to the one described by Zins:

[The] analysis of the panel's definitions of D-I-K-M made it clear that the wording can be deceptive. Panel members often misused the terminology. Therefore, I adopt Begthol's definition as an ad hoc position that IS explores D-I-K-M phenomena, without differentiating, however defined and in whatever relation to each other. (Zins, 2007b, p. 340).

Moving on towards the identification of the aspects that participate in the mediation between information resources and their users, Zins (2007d, p.528) indicates that they should answer the following questions: who? — the mediators (human and/or mechanical); what? — the matter (desired contents/subjects); why? — the motive (query reason); how? — the means (intervening methods); where and when? — the milieu (historical and social context).

In this framework, user-centered methodologies feature as potential processes not only for the creation of optimized interfaces between information resources and their users, but also to shape resources (perceived as artifacts). *Contextual Design* (CD) is one of these methodologies,² introduced by its authors as “a user-centered design process that uses in-depth field research to drive innovative design, [...] a step-by-step process for collecting field data and using it to design any sort of technical product”, (Holtzblatt & Beyer, 2014, p. 1).

Since the introduction of the methodology is focused on the specific characteristic of the concept, i.e., in the term *contextual*, it is important to specify what should the broader term *design* encompasses. In this study, the definition suggested by Ralph and Wand is adopted, according to which design is “a *specification* of an object, manifested by some *agent*, intended to accomplish *goals*, in a particular *environment*, using a set of *primitive components*, satisfying a set of *requirements*, subject to some *constraints*” (Ralph & Wand, 2009, p. 108, italics in the original).

The combination of these two definitions allows us to posit CD as a specification process of an artifact or service, based on systematically-collected data from users while these execute tasks related to the products to be

1 Vide the debate on the definitions of the concepts of data, information, knowledge and message, in Chaim Zins (2007c), or the critical review performed by Jennifer Rowley (2007) about the data-information-knowledge-wisdom hierarchy suggested by Russell Ackoff (1999).

2 In addition to Contextual Design, other user-centered methodologies exist, such as Goal Directed Interaction Design, Scenario-Based Design and Human-Centered Systems Development Life Cycle. Iivari and Iivari (2011) identify advantages and limitations in the use of each of them.

developed. This monitoring constitutes the essential aspect of the methodology, as the observation of how users engage with a situation will serve as a starting point for the creative process (Iivari & Iivari, 2011, p. 140).

The methodology comprises three stages: i) immersive experience in the “world” of the end-users (current, future or potential); ii) idealization of new concepts based on the global panorama obtained from the first stage; iii) development and test of the product using end-users (Holtzblatt & Beyer, 2014, p. 1). Amongst all these stages, the first is pivotal as it provides the big picture that will be the foundation for all subsequent developments. The creation of the visual models (diagrams) is most crucial step of this stage, as they not only organize and systematize collected data, but also play a pivotal role in addressing what Holtzblatt and Beyer (2014, p. 18) regard as the greatest challenge in development teams — to achieve a shared understanding of the “world” of the user.

Based on the aforementioned assumptions, we aim to identify and analyze studies in IS that address or use the CD methodology as part of the development of mediating solutions between information resources and users. Specifically, our goal is to: (i) identify peer-reviewed papers that discuss or use the CD methodology (in databases related with IS, as mentioned ahead); (ii) describe the approach to the CD concept used in the retrieved papers; (iii) highlight potential contributes and limitations of this methodology, as pointed out by those papers.

In order to address these objectives, we endeavored on an exploratory/descriptive study based on a systematic review. A categorical analysis was performed on the corpus of papers selected using the methodology described in the next section.

2 Methodology

The following services were used as sources for the constitution of the corpus of the study: a) *EBSCO Discovery Service* (EDS); b) *ProQuest* (PQ); c) *Web of Science* (WoS); d) *Directory of Open Access Journals* (DOAJ); e) *E-prints in Library & Information Science* (E-LIS). With the exception of the latter, which is field-specific, queries in the remaining platforms were filtered to the following collections: in DOAJ, to *Bibliography - Library Science - Information Resources*; in WoS, to *Information Science & Library Science*; in PQ, to *Library and Information Science Abstracts*; and in EDS, to *Library & Information Science Source*; to *Library, Information Science & Technology Abstracts* and to *Information Science & Technology Abstracts*.

Since we intended to retrieve only papers in which CD plays a significant part, it was opted to restrict the query to papers in which the term ‘contextual design’ occurs in at least one of the following indexing fields: “title”, “subjects” or “abstract”. In addition to the restriction to peer-reviewed articles, no other formal or temporal filters were applied. We decided not to apply a temporal filter since we observed, in the exploratory queries, week occurrence of relevant results.

The corpus was submitted to a hermeneutic approach that seeks to form a progressive understanding of the texts under analysis (Kuckartz, 2014, p. 19). The technique of categorical and contingency analysis was used, namely categorization by “corpus” (Bardin, 2011, pp. 149-259). For the categories that emerged during the gradual classification process, meaning units formed by textual segments extracted from the corpus under analysis were attributed (Bardin, 2011, p. 134; Kuckartz, 2014, p. 44).

3 Results and discussion

Collection occurred in April 4th, 2018, in the mentioned platforms, according to the aforementioned methodology. The results were as follows: five papers in PQ; four in EDS and WoS; one in DOAJ and none in E-LIS. In the end, after the removal of duplicates, the corpus was comprised of eight studies (see Table 1).

Table 1. Corpus of the study.

tRef.	Source	Publisher	Author	Title
#01	WoS.	(2001) <i>Information & Management</i>	Smart & Whiting	<i>Designing systems that support learning and use: a customer-centered approach</i>
#02	EDS; PQ.	(2001) <i>Journal of Library Administration</i>	Normore	<i>Reference in Context Project</i>
#03	EDS; PQ; WoS.	(2004) <i>Library & Information Science Research</i>	Nesset & Large	<i>Children in the information technology design process: A review of theories and their applications</i>
#04	EDS.	(2006) <i>Universal Access in the Information Society</i>	Marsico et al.	<i>A proposal toward the development of accessible e-learning content by human involvement</i>
#05	EDS; PQ; WoS.	(2008) <i>Electronic Library and Information Systems</i>	Keshavarz	<i>Human information behavior and design, development and evaluation of information retrieval systems</i>
#06	DOAJ; PQ.	(2008) <i>Journal of Library and Information Science</i>	Kruse et al.	<i>A User Field Study: Communication in Academic Communities and Government Agencies</i>
#07	PQ.	(2009) <i>Journal of Access Services</i>	Kelly et al.	<i>Accessibility 2.0: Next Steps for Web Accessibility</i>
#08	WoS.	(2016) <i>Research and Advanced Technology for Digital Libraries</i>	Heuwing et al.	<i>Contextual Design Methods for Information Interaction in the Workplace</i>

As far as the discussion of the CD concept in these papers is concerned, it was observable a high amount of descriptive and/or comparative studies of

experimental projects using this methodology (six), in contrast with more theoretical studies (two), as shown in Table 2.

Table 2. Typology of the studies and corresponding approaches identified in the corpus.

Study typology	tRef.	Approach
a) Descriptive studies of specific projects in which CD was used as a methodology, in its original form or adapted.	#01	This study uses contextual design methods to gather and analyze information on users' wants, needs, and work habits when using computers.
	#02	The method we used is called Contextual Design [...] we were interested in how people look for information.
	#04	The present proposal stems from blending issues related to the contextual design approach with characteristics of learner-centered design.
	#06	The study was conducted within the academic community at British and Danish universities and government agencies in The Netherlands, using the 'Contextual Design' approach and 'Cultural Probes'.
b) Literature reviews on user-centered methodologies for the development of information systems.	#03	This article reviews the literature on the role that children can play in the design of information technology applications intended for young users themselves.
	#05	The paper takes the form of a literature review with particular concentration on the efforts made by information science researchers.
c) Descriptive study of a project for a model of contextual approach.	#07	The paper describes a tangram model which provides a pluralistic approach to Web accessibility, and provides case studies which illustrate use of this approach.
d) Comparative study of projects that use different user-centered methodologies.	#08	This contribution presents and compares methods which can be used to elicit information about users in the workplace, and to analyze and to create requirements based on these results, especially from scenario-based design and contextual design.

In descriptive/comparative studies, this methodology is used autonomously (#01 and #02) and in combination with other methods (#04, #06 and #08). Moreover, study #07 is worthy of note for the fact that it does not use CD as created and defined by Holtzblatt and Beyer, who are not cited. In this paper, the term 'contextual design' appears only in the key-words. The use of CD as a common-name to identify the specification of the model developed can only be inferred by an analysis of the study. Due to this, this study (#07) was not included when analysis the contributes towards the potentials and limits of the CD methodology.

Regarding the potentials made explicit in retrieved papers, five out of seven stress how effectively CD serves as a method to collect information amongst users (category a) of Table 3). Its relevance is also reflected in the several specifications listed in Table 3, amongst which is stressed its effectiveness in the collection of data about user's information behavior.

Table 3. Potential advantages and corresponding specifications of the CD methodology, as made explicit in the corpus.

Potential advantages	Specification (if applied)	tRef.
a) CD provides an effective method to gather information from users...	... in a systematic way. ... in a way that is implicit in their actions. ... in their work environment. ... regarding their information behavior.	#01 #01; #05 #01 #02; #05; #06; #08
b) CD adds value to information systems and to the organizations responsible for them.		#01
c) CD includes techniques to organize sets of complex and diversified data.		#02
d) CD provides a structure that allows individuals without prior experience to perform interviews.		#02
e) CD incorporates an interdisciplinary approach.		#03; #04

As far as limitations are concerned, only four were identified. The only limitation mentioned in more than one paper was associated with the lack of involvement of users in all the stages of a system's development (category c) of Table 4). This limitation, in addition to the one expressed in category b), are explicitly mentioned in papers classified as literature reviews in the methodology.

Table 4. Limitations of the CD methodology, as made explicit in the corpus.

Limitations	tRef.
a) CD implies a new form of action that may hinder its use.	#01
b) CD "demands" cooperation with untrained individuals.	#03
c) CD does not involve the users in all stages of the process.	#03; #05
d) CD and other similar methods are not sufficient when dealing with complex information systems.	#08

Regarding the limitations found in comparative/descriptive studies, it was observed that one paper (#08) does not restrict it specifically to CD, but extends it to all user-centered methodologies. Moreover, paper #01 does not express that limitation as a methodological problem. In contrast with the others, this limitation suggests that the difficulty in implementing a CD methodology lies in its novelty. However, considering the use history of CD (Holtzblatt & Beyer, n.d.) in the Human-Computer Interaction (HCI) field, this methodology can hardly be considered a novelty in 2001, when this study was published.

Figure 2 shows a graphical representation of the diachronic comparison between the publication years of the studies that constitute the corpus, and representative works of the evolutionary process of CD as a methodology,

according to its entry in the *Encyclopedia of Human-Computer Interaction*, available online.

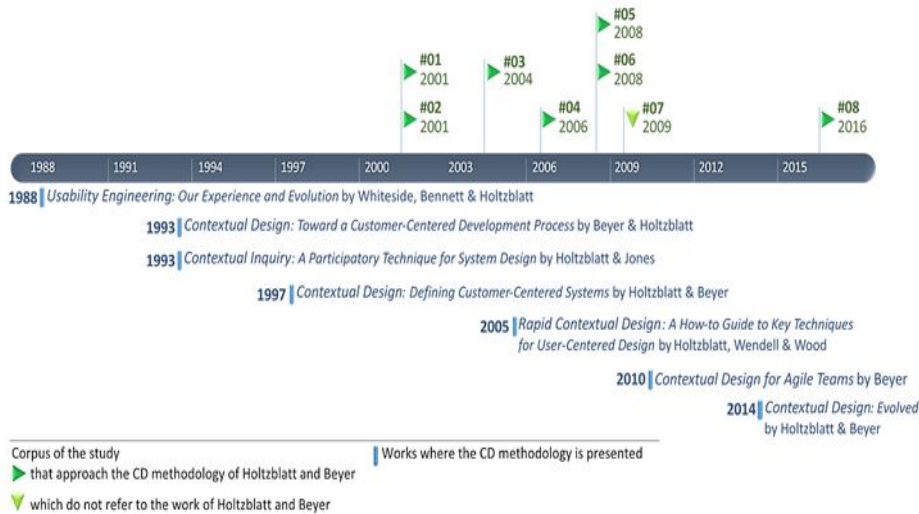


Figure 2. Schematic representation of the diachronic comparison between the publication years of studies that constitute the corpus and representative works of the evolutionary process of CD as a methodology (elaborated by the author).

It is possible to verify in Figure 2 that retrieved papers in the IS field appear only after almost a decade, in relation to the first works carried out in 1993.

4 Conclusions

This study enables us to draw some indicators regarding the use of CD as a methodology, applied to the development of mediation solutions between information resources and users. However, conclusions are limited by the exploratory nature of research.

The first indicator is related to the low number of retrieved papers (eight), considering the diversity of databases queried and the nearly three decades of existence of this methodology. This number indicates a low use-rate of this methodology in IS.

In regard to the approaches to the concept of CD, it was concluded that it is essentially empirical, as the methodology is used to obtain the desired results. Furthermore, it was observed a trend to combine methodologies, which is reflected in the mixed approaches revealed by each descriptive/comparative study since 2004.

Two potential uses of CD as a methodology were also noted, each related to a different moment of the process: data collection and organization. In the first, the

possibility of performing contextual interviews that collect implicit data in user's actions, in addition to information they made explicit, allows the collection of data that reflects the real needs of the target audience of the system, which individuals are sometimes not aware of. In the latter, it was highlighted the ability to systematize in categories usually large and complex amounts of data, through visual models of the "world" of the user. As a result, it can be concluded that CD provides added value for the development of suitable interfaces between information resources and its users.

Finally, there might be a potential link between how limitations are made explicit and the methodological orientation of studies, since the two limitations most directly concerned with intrinsic characteristics of CD were revealed by the literature review. These two limitations refer to user engagement, and can be seen as opposite: some authors indicate as a limitation the non-engagement of users in the whole design process, whereas others call out to the difficulty of working with individuals without specific training.

5 Final remarks

This study stresses the interdisciplinary nature of CD as a methodology, both at a domain-level and in its use in multidisciplinary research teams. This plural nature might be the echo of the equally interdisciplinary nature of IS, although this topic is not consensual.

The overview conveyed by this research concerns the explicit use of this methodology, without considering partial uses of methods or techniques that do not refer to CD as a whole. Some examples can be found in studies that apply techniques such as 'contextual inquiry' or 'contextual interview', which can be considered a part of CD. However, by not making this bond explicit, they were not included in this study as it is questionable whether or not the authors know about the CD methodology.

These issues may constitute interesting starting points for future research. It might be interesting to perform a more extensive survey on the use of CD in IS, or to verify if researchers that use adjacent techniques — e.g. contextual inquiry/interview — relate them to other components of the CD methodology. Such endeavors would complement and complete the initial overview that this study provides.

References

- Ackoff, R. L. (1999). *From data to wisdom*. In Ackoff's Best: His classic writings on management (pp. 170–172). New York: John Wiley & Sons.

- Araújo, C. A. Á. (2013). O que é Ciência da Informação? *Informação & Informação*, 19(1), 1–30. <https://doi.org/10.5433/1981-8920.2014v19n1p01>
- Bardin, L. (2011). *Análise de conteúdo*. (L. A. R. A. Pinheiro, Trans.). São Paulo: Almedina.
- Barreto, A. de A. (2008). Uma quase história da ciência da informação. *DataGramaZero*, 9(2), 1–18. Retrieved from <http://eprints.rclis.org/17637/>
- Buckland, M. K. (2012). What kind of science can information science be? *Journal of the American Society for Information Science and Technology*, 63(1), 1–7. <https://doi.org/10.1002/asi.21656>
- Capurro, R. (2003, November 10). Epistemologia e Ciência da Informação. (A. M. R. Cabral, E. W. Dias, I. Paim, L. M. M. Dumont, M. P. Aun, & M. E. N. Borges, Trans.), *V Encontro Nacional de Pesquisa Em Ciência Da Informação*. Belo Horizonte. Retrieved from http://www.capurro.de/enancib_p.htm
- De Marsico, M., Kimani, S., Mirabella, V., Norman, K. L., & Catarci, T. (2006). A proposal toward the development of accessible e-learning content by human involvement. *Universal Access in the Information Society*, 5(2), 150–169. <https://doi.org/10.1007/s10209-006-0035-y>
- Heuwing, B., Mandl, T., & Womser-Hacker, C. (2016). Contextual Design Methods for Information Interaction in the Workplace. *Journal of Library and Information Science*, 42(1), 72–78. <https://doi.org/10.6245/JLIS.2016.421/>
- Holtzblatt, K., & Beyer, H. (n.d.). *Contextual Design*. Retrieved June 26, 2018, from <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/contextual-design>
- Holtzblatt, K., & Beyer, H. (2014). *Contextual Design: Evolved*. (J. M. Carroll, Ed.). Morgan & Claypool. <https://doi.org/10.2200/S00597ED1V01Y201409HCI024>
- Iivari, J., & Iivari, N. (2011). Varieties of user-centredness: an analysis of four systems development methods. *Information Systems Journal*, 21(2), 125–153. <https://doi.org/10.1111/j.1365-2575.2010.00351.x>
- Kelly, B., Sloan, D., Brown, S., Seale, J., Lauke, P., Ball, S., & Smith, S. (2009). Accessibility 2.0: Next steps for web accessibility. *Journal of Access Services*, 6(1–2), 265–294. <https://doi.org/10.1080/15367960802301028>
- Keshavarz, H. (2008). Human information behaviour and design, development and evaluation of information retrieval systems. *Program: Electronic Library and Information Systems*, 42(4), 391–401. <https://doi.org/10.1108/00330330810912070>
- Kruse, F., Sørensen, A. B., Ballaux, B., Christensen-Dalsgaard, B., Hofman, H., Nielsen, M. P., ... Thøgersen, J. (2008). A User Field Study: Communication in Academic Communities and Government Agencies. In B. Christensen-Dalsgaard, D. Castelli, J. B. Ammitzbøll, & J.

- Lippincott (Eds.), *Research and Advanced Technology for Digital Libraries* (pp. 447–449). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-87599-4_60
- Kuckartz, U. (2014). *Qualitative Text Analysis: A guide to methods, practice and using software*. (K. Metzler, Ed.). London: SAGE Publications Ltd.
- Le Coadic, Y.-F. (1996). *A ciência da informação*. (M. Y. F. Gomes, Trans.). Brasília: Briquet: de Lemos/Livros.
- Machado, L. M. O., Simões, M. da G. de M., & Souza, R. R. (2017). Relações disciplinares entre a Ciência da Informação e a “triade” Biblioteconomia, Arquivística e Documentação (1960-2000): Subsídios para uma reflexão sobre a área. *Ciência Da Informação*, 46(2), 33–50. Retrieved from <http://revista.ibict.br/ciinf/article/view/3077/3698>
- Marques, M. B. (2015). Informação, comunicação e conhecimento: os desafios da sociedade do século XXI. In C. Camponez, B. Araújo, F. Pinheiro, I. Godinho, & J. Moraes (Eds.), *IX Congresso Sopcom. Comunicação e Transformações Sociais (vol.1): Ciência da Informação Comunicação e Educação* (pp. 48–62). Coimbra. Retrieved from <http://www.bocc.ubi.pt/pag/sopcom/1-ix-congresso.pdf>
- Marques, M. B. (2017). *Gestão da informação em sistemas de informação complexos*. Pesquisa Brasileira Em Ciência Da Informação e Biblioteconomia, 12(2), 60–76.
- Neset, V., & Large, A. (2004). Children in the information technology design process: A review of theories and their applications. *Library & Information Science Research*, 26(2), 140–161. <https://doi.org/10.1016/j.lisr.2003.12.002>
- Normore, L. F. (2001). Reference in Context Project. *Journal of Library Administration*, 34(3–4), 345–353. https://doi.org/10.1300/J111v34n03_15
- Pinheiro, L. V. R. (2006). Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In M. N. González de Gómez & E. G. Dill Orico (Eds.), *Políticas de memória e informação* (pp. 111–142). Natal: EDUFRRN. Retrieved from <http://ridi.ibict.br/handle/123456789/18>
- Ralph, P., & Wand, Y. (2009). A proposal for a formal definition of the design concept. In K. Lyytinen, P. Loucopoulos, J. Mylopoulos, & B. Robinson (Eds.), *Design Requirements Engineering: A Ten-Year Perspective* (pp. 103–136). Cleveland: Springer. Retrieved from https://doi.org/10.1007/978-3-540-92966-6_6
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>

- Saracevic, T. (2009). Information Science. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences* (pp. 2570–2585). New York: Taylor & Francis. <https://doi.org/10.1081/E-ELIS3-120043704>
- Shera, J. H., & Cleveland, D. B. (1977). History and foundations of information-science. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 12, pp. 249–275). New York: Knowledge Industry Publications Inc.
- Silva, A. M. da, & Ribeiro, F. (2008). *Das “ciências” documentais à ciência da informação: Ensaio epistemológico para um novo modelo curricular* (2nd ed.). Porto: Edições Afrontamento.
- Silva, J. L. C., & Freire, G. H. de A. (2012). Um olhar sobre a origem da ciência da informação: indícios embrionários para sua caracterização identitária. *Encontros Bibli*, 17(33), 1–29. <https://doi.org/10.5007/1518-2924.2012v17n33p1>
- Smart, K. L., & Whiting, M. E. (2001). Designing systems that support learning and use: a customer-centered approach. *Information & Management*, 39(3), 177–190. [https://doi.org/10.1016/S0378-7206\(01\)00089-1](https://doi.org/10.1016/S0378-7206(01)00089-1)
- Souza, R. R., & Almeida, M. B. (2009). Representação do conhecimento: identidade ou esvaziamento da Ciência da Informação? In A Ciência da Informação criadora de conhecimento, vol.1, *IV Encontro Ibérico EDIBCIC 2009* (pp. 157–165). Coimbra: Imprensa da Universidade de Coimbra.
- Souza, R. R., Almeida, M. B., & Baracho, R. M. A. (2013). Ciência da Informação em transformação: Big Data, Nuvens, Redes Sociais e Web Semântica. *Ciência Da Informação*, 40(2), 159–173. Retrieved from <http://revista.ibict.br/index.php/ciinf/article/view/2283/1906>
- Zins, C. (2006). Redefining information science: From “information science” to “knowledge science.” *Journal of Documentation*, 62(4), 447–461. <https://doi.org/10.1108/00220410610673846>
- Zins, C. (2007a). Classification Schemes of Information Science: Twenty-Eight Scholars Map the Field. *Journal of the American Society for Information Science and Technology*, 58(4), 645–672. <https://doi.org/10.1002/asi.20506>
- Zins, C. (2007b). Conceptions of Information Science. *Journal of the American Society for Information Science and Technology*, 58(4), 335–350. <https://doi.org/10.1002/asi.20507>
- Zins, C. (2007c). *Conceptual Approaches for Defining Data, Information, and Knowledge*. *Journal of the American Society for Information Science and Technology*, 58(4), 479–493. <https://doi.org/10.1002/asi.20508>
- Zins, C. (2007d). Knowledge Map of Information Science. *Journal of the American Society for Information Science and Technology*, 58(4), 526–535. <https://doi.org/10.1002/asi.20505>

Repositório de dados como forma de organização do conhecimento e desenvolvimento científico

Elizabete Cristina de Souza de Aguiar Monteiro¹, Priscila Machado Borges Sena²,
Ricardo César Gonçalves Sant'Ana³ y Ursula Blattmann⁴

¹ ORCID [0000-0002-3797-8139](https://orcid.org/0000-0002-3797-8139). Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, Brasil.
ecsamonteiro@gmail.com

² ORCID [0000-0002-5612-4315](https://orcid.org/0000-0002-5612-4315). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Brasil.
priscilasena.ufsc@gmail.com

³ ORCID [0000-0003-1387-4519](https://orcid.org/0000-0003-1387-4519). Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, Marília, São Paulo.
ricardo.santana@unesp.br

⁴ ORCID [0000-0002-8834-0987](https://orcid.org/0000-0002-8834-0987). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Brasil.
ursula.blattmann@ufsc.br

Resumo: Os dados armazenados em repositórios de dados podem contribuir para o resgate da memória de um pesquisador em relação a sua pesquisa e de uma instituição, uma vez que são organizados e representados de forma a revelar os métodos e os instrumentos utilizados pelos pesquisadores em determinados períodos de tempo, bem como as temáticas pesquisadas, os tipos de dados coletados ou gerados e o contexto histórico que fizeram parte, tornando-se passíveis de constituir uma memória coletiva. Buscou-se apresentar como a organização e representação de dados em repositório de dados pode contribuir para a constituição e recuperação da memória institucional. A metodologia foi exploratória e descritiva. O universo foi composto por 36 repositórios recuperados das cem melhores universidades do mundo. Para coleta dos dados sobre o padrão de metadados utilizados pelos repositórios foi utilizado o Registry of Research Data Repositoy. Os resultados demonstram que para a representação dos conjuntos de dados, os repositórios utilizam o esquema de metadados Dublin Core e alguns repositórios criaram seus próprios requisitos a partir do DC para atender suas particularidades na representação. Conclui-se que os repositórios de dados agregam valor aos repositórios institucionais das universidades tendo o compromisso com a formação da memória acadêmica e institucional.

Palavras-chave: Repositório de dados; Dados científicos; Organização do conhecimento; Desenvolvimento científico.

Abstract: Data stored in data repositories can contribute to the rescue of a researcher's memory in relation to your research and institution, since they are organized and represented in order to reveal the methods and instruments used by the researchers in certain periods of time, as well as the themes researched, the types of data collected or generated and the historical context that did part, making it possible to constitute a collective memory. It was tried to present how the organization and representation of data in data repository can contribute to the constitution and retrieval of institutional memory. The methodology was exploratory and descriptive. The universe was composed of 36 repositories retrieval from the 100 best universities in the world. To collect the data on the metadata standard used by the repositories, the Registry of Research Data Repository was used. The results demonstrate that for the representation of data sets, repositories use the Dublin Core Metadata Scheme and some repositories have created their own requirements from DC to meet their particularities in representation. It is concluded that the data repositories add value to the institutional repositories of universities and are committed to the formation of academic and institutional memory.

Keywords: Data repository. Scientific data. Knowledge organization. Scientific development.

1 Introdução

Os Repositórios de dados científicos de universidades são ambientes digitais com infraestrutura para dar suporte aos pesquisadores na gestão e na disponibilização de dados, fatores que potencializam a reutilização por outros pesquisadores (Monteiro, 2017; Monteiro & Sant'Ana, 2018). Esta reutilização pode ocorrer para diversos fins, desde aplicações técnicas à estudos científicos. A realização da representação e organização de dados em repositórios pode colaborar para que estes sejam preservados e reutilizados na construção da memória institucional ou da memória de uma área do conhecimento a fim de serem resgatados no futuro.

Entende-se que os dados armazenados em repositórios possam contribuir para o resgate da memória de uma instituição, uma vez que são organizados e representados de forma a revelar os métodos e os instrumentos utilizados pelos pesquisadores em determinados períodos de tempo, bem como as temáticas pesquisadas, os tipos de dados coletados ou gerados e o contexto histórico que fizeram parte. Deste modo, compreende-se que o resgate dos dados revela o registro da memória individual de um pesquisador em relação a sua pesquisa e, em conjunto com as memórias de outros pesquisadores e da instituição, tornam-se passíveis de constituir uma memória coletiva.

Nesse sentido, torna-se pertinente investigar como ocorre a coleta, organização e representação nos repositórios de dados científicos das universidades, uma vez que os dados quando organizados de forma coerente, podem trazer a luz o contato com o conhecimento dos responsáveis por suas coletas e, desta forma, serem reutilizados com credibilidade. Sendo assim, buscou-se apresentar como a organização e

representação de dados em repositório de dados podem contribuir para a constituição e recuperação da memória institucional.

A metodologia empregada foi exploratória e descritiva. O universo pesquisado foi composto por 36 repositórios decorrente de uma pesquisa sobre repositórios de dados com desdobramentos discutidos e apresentados em outros trabalhos sobre direito autoral dos dados (Monteiro, 2017), privacidade dos dados (Affonso et al., 2017; Monteiro, Affonso, Borba & Sant’Ana, 2018) e Plano de gerenciamento de dados (Monteiro, Sant’Ana, 2017).

Para coleta dos dados sobre o padrão de metadados utilizados pelos repositórios foi utilizado o Registry of Research Data Repositoy, (re3data.org)¹, um registro global de repositórios de dados de pesquisa. Os resultados foram compilados e apresentados em um quadro.

2 Repositório de dados e memória

Os repositórios digitais, presentes nas comunidades científicas, surgiram com propósitos de preservação da memória e visibilidade institucional. Trata-se de ambientes que contribuem no processo da comunicação científica, que armazenam objetos digitais, que gerenciam e preservam esses objetos por longo tempo, que promovem o acesso e o uso dos objetos, impulsionam a visibilidade da produção científica e minimiza os custos de publicações (Arellano, 2008; Camargo & Vidotti, 2009).

Os objetivos dos repositórios digitais são armazenar, preservar e disseminar a memória institucional da universidade a qual pertencem, contribuir com a legislação de acesso aberto à informação e proporcionar maior visibilidade às produções científicas, técnicas, culturais, artísticas, administrativas e tecnológicas da Universidade (Camargo & Vidotti, 2009; Sayão & Marcondes, 2009; Varelal; Barbosa & Guimarães, 2009).

Os repositórios de dados estão alinhados a essas premissas a medida que são uma extensão dos repositórios digitais, mas com o foco em dados científicos pois gerenciam, armazenam, preservam, disseminam os dados de suas comunidades. Esses repositórios contribuem com a memória da produção técnica-científica e cultural tanto dos pesquisadores quanto das instituições que os implementaram, devido suas vinculações a essas produções. Destaca-se que os repositórios de dados estão além desta visão de repositório digitais, uma vez que cada conjunto de dados tem características próprias e por isso pode requerer um tratamento diferenciado (Rodrigues et al., 2010) verificado em sua representação nos repositórios com uso de padrões de metadados próprios.

Posto isso, na próxima seção aborda-se a relação entre organização e representação do conhecimento para a constituição e recuperação da memória.

¹ <https://www.re3data.org/>

3 Organização e representação do conhecimento para a memória

A organização do conhecimento pode ser compreendida como um procedimento de modelagem do conhecimento que objetiva a elaboração de representações do conhecimento. Já a representação do conhecimento pode ser entendida como um modelo de abstração do mundo real, elaborado para determinado fim (Brascher & Café, 2008). Nessa perspectiva, é possível relacionar a organização e representação à constituição de uma dada memória.

Ao abordar memória torna-se relevante ressaltar que esta pode ser individual ou coletiva. De acordo com Halbwachs (2013), a memória coletiva trata-se da análise de um fenômeno de recordação e localização das lembranças a partir dos contextos sociais que vivem como alicerce para o processo de reconstrução da memória. Assim, as memórias de um indivíduo nunca são apenas suas na medida que nenhuma lembrança pode coexistir isolada de um grupo social.

Constata-se nos conceitos apresentados por Halbwachs (2013), que a memória individual é um enfoque acerca da memória coletiva. Neste sentido, um sujeito integra dois tipos de memória, individual e coletiva.

Desta forma, a organização e representação do conhecimento na gestão de dados nos repositórios universitários apresentam-se como processos que podem agregar à construção da memória institucional e científica (coletiva), e ainda à construção da memória individual baseada em coletivos de memória.

Ante o que foi apresentado, na próxima seção são descritos os resultados e realizada a discussão desses.

4 Resultados e discussão

No Quadro 1 estão relacionados os repositórios e os nomes dos esquemas de metadados coletados no re3data.org, quadro este que permitiu a estruturação da sustentação da identificação das classificações propostas em relação aos repositórios pesquisados. Dos 36 repositórios, em 18 não foram localizadas as informações.

UNIVERSIDADE	REPOSITÓRIO	METADATA STANDARD SCHEME	METADATA STANDARD NAME
Harvard University	Harvard Dataverse	DCC	DDI - Data Documentation Initiative
Massachusetts Institute of Technology	DSpace @ MIT	DCC	Não disponível
Stanford University	Stanford Data Repository	DCC - Dublin Core	Não disponível
University of California Berkeley	Data Repository	Não disponível	Não disponível
University of Michigan	Deep Blue Data	DCC - Dublin Core	Repository-Developed Metadata Schemas
University of Washington	ResearchWorks Archive	Não disponível	Não disponível
University of Pennsylvania	ScholarlyCommons	Não disponível	Não disponível

University of Oxford	Oxford University Research Archive	DCC	RDF Data Cube Vocabulary
University of California Los Angeles UCLA	Merritt	DCC	Dublin Core
Yale University	ISPS Data Archive	Não disponível	Não disponível
University of Cambridge	Data Repository	Não disponível	Não disponível
University of Wisconsin Madison	Research Data Service	Não disponível	Não disponível
Michigan State University	MSU Code Repository	Não disponível	Não disponível
University of Texas Austin	Texas ERC	Não disponível	Não disponível
University of California San Diego	iDASH	Não disponível	Não disponível
Pennsylvania State University	ScholarSphere	Não disponível	Não disponível
University of Illinois Urbana Champaign	Illinois Data Bank IDEALS	DCC	Repository-Developed Metadata Schemas DataCite Metadata Schema
University of North Carolina Chapel Hill	Carolina Digital Repository	Não disponível	Não disponível
Princeton University	DataSpace	DCC	Dublin Core
University College London	PARNASSUS	Não disponível	Não disponível
University of British Columbia	cIRcle		DDI - Data Documentation Initiative
University of Maryland Baltimore	DRUM	Não disponível	Não disponível
Purdue University	Purdue University Research Repository	DCC	Dublin Core
California Institute of Technology Caltech	Caltech	DCC	DataCite Metadata Schema
University of Virginia	Libra Data	DCC	DDI - Data Documentation Initiative / Dublin Core / ISA-Tab / ABCD - Access to Biological Collection Data / DataCite Metadata Schema
University of California Irvine	UC Irvine Machine Learning Repository	DCC	Repository-Developed Metadata Schemas
University of Arizona	UA Campus Repository	DCC	AVM - Astronomy Visualization Metadata / Darwin Core / DDI - Data Documentation Initiative / Dublin Core / EML - Ecological Metadata Language
University of Edinburgh	Edinburgh DataShare	DCC	Dublin Core
Washington University Saint Louis	CNDA	Não disponível	Não disponível
Simon Fraser University	Radar	DCC	DDI - Data Documentation Initiative
Utrecht University /	Utrecht University	Não disponível	Não disponível

Universiteit Utrecht			
Virginia Polytechnic Institute and State University	VTechData	Não disponível	Não disponível
Tufts University	GIS	Não disponível	Não disponível
Ruprecht Karls Universität Heidelberg	heiDATA	DCC	ISA-Tab DataCite Metadata Schema Dublin Core DDI - Data Documentation Initiative
University of Copenhagen/ Københavns Universitet	GBIF	Não disponível	Não disponível
University of Amsterdam/ Universiteit van Amsterdam	UvA / AUAS figshare	DCC	Dublin Core / DataCite Metadata Schema

Quadro 1. Metadados para a organização dos conjuntos de dados em repositórios de dados

Diversos esquemas de metadados são utilizados para a representação de conteúdo digital com diferentes graus de especificidade sendo, o Dublin Core, mais amplamente utilizado (Digital Curation Centre, c2014-2019). O esquema é composto por 15 elementos sendo possível adaptações e customizações para as particularidades de cada conteúdo. Além dos 15 elementos, podem ser adicionados qualificadores formando um “perfil de aplicação” que ampliam a semântica deles.

Para a representação dos conjuntos de dados, os repositórios de dados analisados utilizam esquema de metadados Dublin Core com adição de elementos tendo como atributos em comum título, autor, palavras-chave, assunto, versões e descrição dos conjuntos de dados. Destaca-se que o padrão de metadados utilizado nos repositórios é o Dublin Core (DC) e que alguns repositórios criaram seus próprios requisitos a partir do DC para atender suas particularidades na representação.

Observa-se nos repositórios dimensões descritas por Ramírez Céspedes (2006) sobre critérios de avaliação de bibliotecas digitais que configuram elementos os quais compõem e descrevem a memória institucional a medida que representam elementos do contexto histórico-temporal, considerando os repositórios com algumas características das bibliotecas digitais no que se refere à:

1) biblioteconômica – os componentes que contextualizam os repositórios, suas políticas e seus conjuntos de dados, seus usuários, seus serviços e produtos, elementos de organização e representação de conteúdos com esquemas de metadados, vocabulários controlados ou ontologias, recursos de recuperação e busca;

2) tecnológica – dimensão que parte da concepção do repositório como um site na Web e da análise de hardware, software, banco de dados e infraestrutura de rede disponíveis para a sua implementação;

3) interação usuário-sistema – a ênfase está no design das interfaces, na arquitetura da informação, nas funcionalidades do sistema nas quais os usuários interagem com os conteúdos digitais gerenciados pelo repositório.

Os resultados demonstram que os repositórios organizam seus conjuntos de dados em coleções que denominam como:

- comunidades e coleções ou disciplinas: representam as comunidades, os departamentos ou institutos que compõem a universidade, sendo elementos que

representam a memória do que cada área coletou ou gerou de dados e as pesquisas desenvolvidas;

- cobertura temporal: cobre o período histórico ao qual os dados estão relacionados e representam as memórias anuais da comunidade;
- cobertura geográfica: incluem os dados de determinadas cidades, países ou regiões, e representam a memória dos locais que fizeram parte das pesquisas;
- financiador: as agências financiadoras das pesquisas e representam a memória ligada às agências que financiaram as pesquisas em determinados período de tempo.

Com a organização dos datasets nas coleções, as universidades disponibilizam ambientes digitais de memória onde os dados são gerenciados, tratados, armazenados e preservados por longo prazo e que podem ser recuperados, inclusive, para pesquisas históricas e sobre memória, análises bibliométricas e cientométricas, resgatando o que foi pesquisado por cada comunidade em determinado período, o que foi pesquisado em áreas geográficas em determinado tempo, quais períodos foram os que mais publicaram no repositório, inferindo os mais produtivos.

Ressalta-se que de acordo com a Special Libraries Association (SLA, 2016), uma das competências dos profissionais da informação é a Organização de Dados, Informação e Recursos de Conhecimento, por meio do gerenciamento de recursos de dados, informações e conhecimento utilizáveis e acessíveis ao longo da vida definida (Sena, Blattmann & Teixeira, 2017).

Os repositórios de dados têm-se tornado preponderantes para as universidades reunirem os conjuntos de dados de sua comunidade, facilitando a organização, a gestão e a recuperação dos mesmos. Eles reúnem os dados e a documentação relacionada aos dados disponibilizando informações profícuas aos pesquisadores locais, regionais e internacionais reconstituindo a memória dos pesquisadores e da instituição.

5 Considerações Finais

Os repositórios de dados são serviços vinculados organicamente aos ambientes institucionais; agregam valor aos repositórios institucionais que já fazem parte do contexto das universidades a mais tempo e, portanto, têm compromisso com a formação da memória acadêmica e institucional, com a preservação a longo prazo de ativos de valor contínuo com aderência à área da Ciência da Informação. A reunião e organização da memória do pesquisador e da instituição favorece a rastreabilidade e resgate dos elementos que compõem os repositórios de dados.

Para os profissionais da Ciência da Informação o contexto dos repositórios e de sua memória científica e institucional é um caminho com muitos desafios e oportunidades. Desafios, contudo, na gestão dos dados enfocando sua preservação a longo prazo e, consequentemente, a preservação da memória. Oportunidades em expandir sua área de atuação e participar das discussões sobre o contexto apresentado, visto que na literatura a temática possui abordagem recente.

Ademais, acredita-se que com o objetivo proposto para este trabalho, apresentar como a organização e representação nos repositórios de dados científicos das universidades pode contribuir para a constituição e recuperação da memória institucional foi possível inferir que a organização e disponibilização de dados podem contribuir para o resgate na memória organizacional.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) do Brasil por financiar a bolsa de estudos do doutorado em andamento.

Referências

- Affonso, E. P., Monteiro, E. C. S. A., BORBA, V. U., & SANT'ANA, R. C. G. (2017). A privacidade e os planos de gerenciamento de dados de repositórios de dados científicos. *Informação & Tecnologia*, 4, 35-53. Recuperado em 02 fevereiro de 2019, de <http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/37586>
- Arellano, M. A. M. (2008). *Crîtérios para a preservação digital da informação científica*. 2008. 354 f (Doctoral dissertation, Tese (Doutorado em Ciência da Informação)– Universidade de Brasília, Departamento de Ciência da Informação, Brasília). Recuperado em 02 de febrero de 2019, de
- Brascher, M., & Café, L. (2008). Organização da informação ou organização do conhecimento? In: *Anais do Encontro Nacional de Pesquisa em Ciência da Informação*, 9.
- Camargo, L. S. A., & Vidotti, S. A. B. G. (2009). Arquitetura da informação para repositórios científicos digitais. In: Sayão, L. F. et al. (Org.). *Implantação e gestão de repositórios institucionais*: políticas, memória, livre acesso e preservação. Salvador: EDUFBA.
- Digital Curation Centre (c2014-2019). *Dublin Core*. Recuperado de <http://www.dcc.ac.uk/resources/metadata-standards/dublin-core>
- Halbwachs, M. (2013). *A memória coletiva* (2a ed.). Tradução de Beatriz Sidou. São Paulo: Centauro.
- Monteiro, E. C. S. A. (2017). *Direitos autorais nos repositórios de dados científicos*: análise sobre os planos de gerenciamento dos dados. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília. Recuperado em 02 de febrero de 2019, de https://repositorio.unesp.br/bitstream/handle/11449/149748/monteiro_eca_me_mar.pdf?sequence=3&isAllowed=y
- Monteiro, E. C. S. A., BORBA, V. U., & SANT'ANA, R. C. G. (2017). Repositório de Dados Científicos: aspectos sobre privacidade de dados. *Anais do I Workshop de Informação, Dados e Tecnologia (WIDaT 2017)*, Florianópolis, p. 100-106. Recuperado em 02 agosto de 2018, de <https://repositorio.ufsc.br/handle/123456789/180294>
- Monteiro, E. C. S. A., & Sant'Ana, R. C. G. (2018) Plano de gerenciamento de dados em repositórios de dados de universidades. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, 23(53), 160-173. Recuperado em <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p160/37296>.

- Monteiro, E. C. S. A., Affonso, E. P., BORBA, V. U., & SANT'ANA, R. C. G. (2018). A privacidade e os planos de gerenciamento de dados de repositórios de dados científicos. *Informação & Tecnologia*, 4(1), 35-53. Recuperado em <http://www.brapci.inf.br/index.php/res/download/100671>
- Ramírez Céspedes, Z. (2006). Criterios e indicadores para evaluar las bibliotecas digitales. *ACIMED*, 14(6). Recuperado em 02 de febrero de 2019, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352006000600004&lng=es&tlng=pt.
- Rodrigues, E. et al.(2010). *Os repositórios de dados científicos: estado da arte*. Recuperado em 02 de febrero de 2019, de http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=271&lang=pt
- Sayão, L. F., & Marcondes, C. H. Software livres para repositórios institucionais: alguns subsídios para a seleção In: Sayão, L. et al. (2009). *Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*. Salvador: EDUFBA. Recuperado em 02 de febrero de 2019, de https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.
- Sena, P., Blattmann, U., & Teixeira, C. (2017). Ecossistema de Startups em Florianópolis: possibilidades para profissionais da Biblioteconomia. *RBBB. Revista Brasileira de Biblioteconomia e Documentação*, 13, 2571-2588. Recuperado em 5 janeiro, 2019, de <https://rbbd.febab.org.br/rbbd/article/view/864/957>
- Special Libraries Association (SLA). *Competencies for Information Professionals*. 2016. Recuperado em 10 julho, 2018, de <https://www.sla.org/about-sla/competencies/>
- Varelal, A. V.; Barbosa, M. L. A. & Guimaraes, I. B. (2009). Dos processos analógicos às tecnologias digitais contemporâneas de recuperação da informação: caminhos cognitivos na mediação para o acesso ao conhecimento. In: *Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*. Salvador: EDUFBA. Recuperado em 02 de febrero de 2019, de https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.

Estudio de la opinión del profesorado de la Universitat de Barcelona de la nueva herramienta de descubrimiento del CRAI, *Cercabib*

Andreu Sulé¹, Juanjo Boté² y Marina Salse³

¹ ORCID [0000-0002-2467-3678](https://orcid.org/0000-0002-2467-3678). Departament de Biblioteconomia, Documentació i Comunicació Audiovisual & Centre de Recerca en Informació, Comunicació i Cultura. Universitat de Barcelona, España.
sule@ub.edu.

² ORCID [0000-0001-9815-6190](https://orcid.org/0000-0001-9815-6190). Departament de Biblioteconomia, Documentació i Comunicació Audiovisual & Centre de Recerca en Informació, Comunicació i Cultura. Universitat de Barcelona, España.
juanjo.botev@ub.edu

³ ORCID [0000-0003-2003-7225](https://orcid.org/0000-0003-2003-7225). Departament de Biblioteconomia, Documentació i Comunicació Audiovisual & Centre de Recerca en Informació, Comunicació i Cultura. Universitat de Barcelona, España.
salse@ub.edu

Resumen. El objetivo del trabajo es conocer la opinión del profesorado de la Universitat de Barcelona (UB) de la nueva herramienta de descubrimiento del Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI), *Cercabib*. En 2018, *Cercabib* sustituyó los anteriores sistemas de recuperación de información de la UB, fusionando en un único entorno de búsqueda el *Catàleg de les biblioteques* y el *ReCercador+*, y convirtiéndose en la ventanilla única de búsqueda simultánea de todos los fondos del CRAI independientemente del soporte, tipología o ubicación del recurso.

Como en cualquier proceso de cambio, la implementación de una nueva herramienta de descubrimiento puede generar diversidad de opiniones entre sus usuarios. La introducción de nuevas interfaces gráficas, de nuevas funcionalidades y de una nueva filosofía de búsqueda basada en la consulta simultánea de todos los fondos del CRAI puede satisfacer a una parte de ellos, pero también puede generar reacciones adversas, ya sea por el propio hecho del cambio (la incomodidad de tener que aprender algo nuevo), por el desconocimiento de la herramienta o por posibles carencias o errores en el diseño del sistema. Conocer las opiniones de los usuarios de *Cercabib* ha de servir para mejorar el servicio ofrecido por el CRAI ya sea por medio de campañas de difusión y formación orientadas a necesidades no detectadas inicialmente o a través de la modificación de aquellos elementos de información o de acción del buscador identificados por los usuarios como problemáticos. Además, tanto la metodología del estudio como sus conclusiones pueden ser de

utilidad para otras bibliotecas universitarias que recientemente también hayan sustituido su catálogo por una nueva herramienta de descubrimiento.

Palabras clave: Herramientas de descubrimiento; Cercabib; Estudio de usuarios; Profesorado; Universitat de Barcelona.

Abstract. The objective of the work is to know the opinion of the faculty of the University of Barcelona of the new discovery tool of the Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI), *Cercabib*. In 2018, *Cercabib* replaced the previous information retrieval systems of the UB, merging into a single search environment the *Catàleg de les Blliblioteques* and the *ReCercador+*, and becoming the single window of simultaneous search of all CRAI funds regardless of the support, type or location of the resource.

As in any change process, the implementation of a new discovery tool can generate a diversity of opinions among its users. The introduction of new graphical interfaces, new functionalities and a new search philosophy based on the simultaneous consultation of all CRAI funds, can satisfy a part of them, but it can also generate adverse reactions, either due to the change itself (the discomfort of having to learn something new), by the lack of awareness of the tool or by possible scarcities or errors in the design of the system. Knowing the opinions of users of *Cercabib* must serve to improve the service offered by the CRAI, either through dissemination and training campaigns aimed at needs not detected initially or through the modification of those elements of information or action of the search system identified by users as problematic. In addition, both the study methodology and its conclusions may be useful for other university libraries that have recently replaced their catalog with a new discovery tool.

Keywords: Discovery tools; Cercabib; User studies; Faculty; University of Barcelona.

1 Introducción

En los últimos años, las herramientas de descubrimiento están sustituyendo los antiguos OPAC en buena parte de las bibliotecas del mundo, incluyendo las de nuestro país. Las dos principales diferencias de estas herramientas en comparación con los antiguos catálogos (OPAC) son la asimilación del modelo de búsqueda de Google (priorización de la búsqueda simple en una única casilla sin limitadores de campo ni operadores booleanos explícitos) y la recuperación de resultados en todos los fondos de la biblioteca, independientemente de su naturaleza (analógica o digital) y de su procedencia (colección propia, bases de datos, revistas electrónicas, repositorios, etc.). El objetivo final de las herramientas de descubrimiento es reproducir, en la medida de lo posible, la exitosa experiencia de usuario que ofrece Google, tratando de simplificar al máximo el proceso de búsqueda, por medio de una

casilla única, de todos los recursos bibliográficos adquiridos y contratados por la biblioteca (ÁVILA-GARCÍA; ORTIZ-REPISO; RODRÍGUEZ-MATEOS, 2015).

En el caso del Centre de Recursos per a l'Aprenentatge i la Investigació (CRAI) de la Universitat de Barcelona (<http://crai.ub.edu/>) el cambio de sistema se produjo el 2018 cuando *Cercabib* (<http://Cercabib.ub.edu/iii/encore/>) sustituyó los anteriores sistemas de recuperación de información de la UB, fusionando en un único entorno de búsqueda el *Catàleg de les biblioteques* (<https://cataleg.ub.edu/>) y el *ReCercador+* (http://mlplus.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do), y convirtiéndose en la ventanilla única de búsqueda simultánea de todos los fondos del CRAI, independientemente del soporte, tipología o ubicación del recurso. Tal y como se expone en la página de presentación de *Cercabib*, con esta herramienta se pueden “realizar búsquedas de manera simultánea en todos los fondos del CRAI al margen del soporte, tipología o ubicación del recurso. Es decir: libros, revistas, artículos de revista, tesis, audiovisuales, entre otros; independientemente de si son recursos en papel o formato electrónico y de si están ubicados físicamente en alguno de nuestros CRAI Bibliotecas o alojados en un servidor propio o ajeno. En conjunción con el SIRE (Servicio Intermediario de acceso a los Recursos Electrónicos), el *Cercabib* os dará acceso a los recursos electrónicos suscritos por el CRAI.” (<http://crai.ub.edu/es/recursos-de-informacion/Cercabib>).

Ahora bien, más allá de las declaraciones formales, el éxito de una herramienta de búsqueda se mide por la mayor o menor satisfacción de sus usuarios. Como en cualquier proceso de cambio, la implementación de una nueva herramienta de descubrimiento puede generar diversidad de opiniones entre sus usuarios. La introducción de nuevas interfaces gráficas, de nuevas funcionalidades y de una nueva filosofía de búsqueda basada en la consulta simultánea de todos los fondos del CRAI puede satisfacer a una parte de ellos, pero también puede generar reacciones adversas, ya sea por el propio hecho del cambio (la incomodidad de tener que aprender algo nuevo), por el desconocimiento de la herramienta o por posibles carencias o errores en el diseño del sistema.

Es por ello que el objetivo del presente trabajo es conocer la opinión, la experiencia de los usuarios de *Cercabib*, y más concretamente del profesorado de la Universitat de Barcelona (UB). Conocer sus opiniones ha de servir para mejorar el servicio ofrecido por el CRAI, ya sea por medio de campañas de difusión y formación orientadas a necesidades no detectadas inicialmente o a través de la modificación de aquellos elementos de información o de acción del buscador identificados por los usuarios como problemáticos. Además, puesto que el perfil y las necesidades de los profesores de la Universitat de Barcelona son similares a las de otros centros académicos del mundo, tanto la metodología del estudio como sus conclusiones pueden ser de utilidad para otras bibliotecas universitarias que recientemente también hayan sustituido su catálogo por una nueva herramienta de descubrimiento.

Este estudio forma parte de un proyecto más amplio de evaluación de *Cercabib* que se irá completando con el estudio de la experiencia de otros colectivos universitarios (estudiantes, bibliotecarios, etc.), así como con la evaluación de

funcionalidades específicas de la herramienta (búsqueda simple, búsqueda avanzada, limitadores, facetas, etc.), de sus interfaces gráficas y del uso del fondo por medio de *logs* y estadísticas de préstamo.

2 Metodología

Para recoger la opinión y la experiencia de los profesores de la UB sobre *Cercabib* hemos optado por la entrevista personal basada en un cuestionario previo (Anexo). Ello nos ha permitido definir a los entrevistados con mayor precisión el objetivo del estudio, y también aclarar posibles dudas sobre el significado de las preguntas. Además, la entrevista permite la modulación de las preguntas adaptándolas al perfil del entrevistado e incluso derivar la conversación a otros aspectos relacionados con el objetivo del trabajo que ayudan a una mejor interpretación de los resultados.

Debido a este carácter cualitativo de nuestro estudio, el muestreo ha sido no probabilístico intencional (GONZÁLEZ-TERUEL; BARRIOS CERREJÓN, 2012), es decir, hemos seleccionado el profesorado no de manera aleatoria sino en base a una serie de criterios que consideramos adecuados para nuestro objetivo:

- Profesorado con una antigüedad mínima de cinco años en la Universitat de Barcelona. Con ello queremos asegurar una cierta experiencia en el uso de las herramientas de búsqueda del CRAI i especialmente del anterior *Catàleg de les biblioteques*.
- Paridad de profesorado dedicado principalmente a actividad docente y profesorado dedicado principalmente a actividad investigadora. Con ello queremos detectar posibles diferencias de opinión de *Cercabib* relacionadas con diferentes necesidades de uso de la información.
- Profesorado perteneciente a diferentes áreas y disciplinas. Para detectar posibles diferencias de opinión de *Cercabib* relacionadas con conocimientos, habilidades y necesidades de información de grupos de usuarios específicos.
- Paridad de profesorado por edad. Con ello queremos detectar posibles diferencias de opinión de *Cercabib* relacionadas con diferentes expectativas y experiencias de búsqueda de información.

Las entrevistas se han realizado de febrero a marzo de 2019. El número total de profesores entrevistados ha sido de 22, cifra suficientemente representativa puesto que a partir de la veintena de encuentros ya pudimos constatar que habíamos llegado a la saturación de información que podían aportar las entrevistas, es decir, que los nuevos profesores dejaban de aportar datos adicionales o información relevante para el estudio de *Cercabib*.

El perfil de estos 22 profesores ordenado por disciplinas es el siguiente:

Tabla 1. Perfil de los profesores entrevistados ordenado por disciplinas.

Disciplina	Antigüedad	Edad	Dedicación	
			% docente	% investigadora
Arqueología	1990	56	50	50
Biblioteconomía y Documentación	1987	59	80	20
Biblioteconomía y Documentación	1994	57	62	38
Biblioteconomía y Documentación	1995	51	50	50
Biblioteconomía y Documentación	2002	43	86	15
Biblioteconomía y Documentación	2002	46	50	50
Biblioteconomía y Documentación	2007	49	40	60
Comunicación Audiovisual y Publicidad	2010	47	50	50
Enfermería	2007	60	80	20
Filosofía	2003	43	50	50
Historia	1977	68	40	60
Historia	1981	64	50	15
Historia	1986	63	50	50
Historia	1993	59	50	50
Historia	1995	61	40	60
Historia	1996	58	0	0
Historia	1997	49	60	40
Historia	2000	50	51	49
Historia	2003	56	50	50
Interacción Persona-Computadora	1998	50	75	25
Microbiología	2012	37	33	66
Psicología	1994	52	50	50

3 Antecedentes

La revisión bibliográfica muestra el interés en este tipo de estudios, si bien son escasos los trabajos centrados en el colectivo del profesorado. Tal y como indican NICHOLS et al. (2017), los estudiantes son el grupo de usuarios que más se ha tenido en consideración para la evaluación de las herramientas de descubrimiento y, de hecho, son los principales destinatarios de la formación de las bibliotecas en este tipo de sistemas. Y, aunque en menor medida, también son abundantes las investigaciones sobre la opinión y percepción de los bibliotecarios sobre las herramientas de descubrimiento.

Los estudios con estudiantes muestran que sus expectativas están fuertemente condicionadas por su experiencia con *Google*, *Google Scholar* y *Amazon*, tanto por lo

que hace a la simplicidad de la búsqueda como a la gratificación instantánea (NICHOLS et al., 2017). DALAL, KIMURA y HOFMANN (2015) llevaron a cabo un estudio con estudiantes de la Rider University (Estados Unidos de América) donde observaron lo que puede considerarse una de las principales deficiencias de este tipo de servicio, a saber, el desconocimiento del alcance de la herramienta en tanto que no sabían que ésta une el catálogo, varias bases de datos y publicaciones académicas. De hecho, la mayoría de ellos desconocían las diferencias entre Google y la herramienta de descubrimiento.

Las opiniones de los bibliotecarios recogidas por diferentes autores muestran críticas sobre cuestiones técnicas de las herramientas. FAWLEY y KRYSAK (2014) realizaron una encuesta a través de la lista de distribución ili-listserv, de la American Libraries Association, a la que respondieron 152 subscriptores. Se trataba de bibliotecarios universitarios que utilizaban una herramienta de descubrimiento (EBSCO Discovery Service, Summon de Serials Solutions y Primo de Ex Libris, principalmente) para la formación en alfabetización informacional. Los resultados mostraron una alta valoración de la herramienta (76%) al considerarla un buen punto de partida para la investigación (especialmente en los estudios de nivel inferior) y por ofrecer resultados en formatos diferentes. Por su parte, los bibliotecarios que no valoraron bien la herramienta (24%) criticaron la poca fiabilidad de la ordenación por relevancia, el abrumador número de resultados de búsqueda y la creación de falsas expectativas de acceso inmediato a todos los materiales. Preguntados por las mejoras que introducirían en el sistema, una de las más seleccionadas fue mejor comprensión de cómo la herramienta busca y ordena los resultados. En este punto vale la pena destacar el trabajo de AHARONY y PREBOR (2015), que contempla, entre otras valoraciones más técnicas, un estudio de personalidad orientado a descubrir por qué algunos bibliotecarios son más propensos a aceptar o a rechazar las herramientas de descubrimiento. Los investigadores realizaron una encuesta entre 125 bibliotecarios y profesionales de la información de Israel con al que descubrieron que las personas con una actitud más negativa ante la adopción de nuevas tecnologías tenían una visión más pesimista y una menor satisfacción con las herramientas de descubrimiento; por contra, aquellas personas que tenían un nivel de satisfacción más elevado eran aquellos que tenían una actitud abierta y positiva a la incorporación nuevas tecnologías.

Los estudios sobre la percepción de los profesores de las herramientas de descubrimiento son, como se ha comentado anteriormente, escasos. De hecho, algunos de ellos no son el producto de la opinión directa de los docentes sino de valoraciones realizadas por los bibliotecarios del centro. Por ejemplo, NICHOLS et al. (2017), en un estudio que recoge la opinión de 56 bibliotecarios referencistas y de formación de usuarios en universidades de los estados Unidos de América, constatan el poco predicamento que las herramientas de descubrimiento tienen entre los profesores. Únicamente en el 13% de las bibliotecas encuestadas la mayoría o todo el personal docente hacen uso las herramientas de descubrimiento; por el contrario, el

mismo estudio identificó altas tasas de uso entre estudiantes y bibliotecarios (en el 55% de las bibliotecas).

WELLS y RICHARDSON (2014) llevaron a cabo un estudio sobre las expectativas de los usuarios de la Curtin University (Australia) de una instalación de la herramienta Primo implementada cinco años atrás. A diferencia de investigaciones sobre herramientas recién instaladas, el trabajo de WELLS y RICHARDSON tiene la particularidad de encuestar usuarios con experiencia en el manejo del sistema. Quizás por ello los resultados obtenidos mostraron una alta valoración de las características de Primo, en concreto su capacidad de integración de múltiples sistemas en el proceso de búsqueda, el fácil acceso a los recursos en línea, las opciones de filtrado y la cobertura del sistema. En definitiva, los usuarios valoraron Primo como una herramienta fácil de usar.

4 Resultados del estudio

El cuestionario que ha servido de guía para la realización de las entrevistas personales a los profesores de la UB consta de dos partes:

- Una primera para recoger la información demográfica del docente (fecha de nacimiento, año de inicio de la vinculación con la UB, dedicación, etc.);
- Una segunda compuesta por 15 preguntas de respuesta abierta de carácter cualitativo sobre la actitud a los cambios tecnológicos, los hábitos de búsqueda de información y la parte más importante dedicada a la percepción y valoración de *Cercabib*.

A continuación se exponen de forma sumariada los resultados de las entrevistas agrupados en cinco preguntas clave:

- Conocimiento de la existencia de *Cercabib*
- Conocimiento del alcance de *Cercabib*
- Experiencia (positiva o negativa) en el uso de *Cercabib*
- Aspectos más positivos y más negativos de *Cercabib*
- Preferencia entre el catálogo anterior y *Cercabib*

4.1 Conocimiento de la existencia de *Cercabib*

El conocimiento de la existencia de *Cercabib* es muy amplio entre el profesorado entrevistado (86%, 19 de 22). Algunos docentes han sido más concretos que otros en la denominación del sistema, pero casi todos ellos saben que el servicio de bibliotecas de la UB dispone de una nueva herramienta de búsqueda.

Segmentados los resultados por disciplina, no se observan diferencias significativas que permitan detectar un conocimiento específico de *Cercabib* derivado de la especialización temática de los entrevistados.

Por contra, en el caso de la antigüedad en la UB sí que se aprecia un menor conocimiento de *Cercabib* en las personas que llevan menos años vinculadas a la universidad (franja de incorporación 2000-2010 y especialmente 2011-2012). Una posible explicación es que estos docentes acostumbran a ser personas más jóvenes que en las entrevistas manifestaron su preferencia por el uso de herramientas alternativas a *Cercabib* para buscar información (*Google*, *Google Scholar*, *Web of Science*, *Scopus*, *ResearchID*, *Academia*, etc.).

Segmentados por edad los resultados ratifican la valoración hecha anteriormente: a medida que disminuye su edad los profesores tienen un menor conocimiento de la existencia de *Cercabib*.

Por último, los datos segmentados por dedicación muestran una cierta tendencia al desconocimiento de *Cercabib* a medida que la docencia no es la actividad principal del profesorado. Este hecho es posible que esté relacionado con la preferencia por el uso de herramientas de búsqueda más específicas en los entornos de investigación. Algunos profesores manifestaron preferir buscar información en bases de datos bibliográficas directamente relacionadas con sus líneas de investigación.

4.2 Conocimiento del alcance de *Cercabib*

Únicamente el 32 % (7 de 22) de los profesores entrevistados manifestaron tener un conocimiento claro del alcance de *Cercabib*, es decir, qué tipos de documentos se pueden encontrar, de qué fuentes provienen o cuál es la diferencia en relación al anterior sistema de búsqueda del CRAI de la UB.

4.3 Experiencia (positiva o negativa) en el uso de *Cercabib*

La experiencia de uso de *Cercabib* es, en general, poco satisfactoria. Únicamente 6 de los 22 profesores entrevistados (27%) manifestaron una valoración positiva de la herramienta. Del resto, 11 la valoraron negativamente y 5 no manifestaron ninguna opinión puesto que no empleaban *Cercabib* para buscar información.

Si nos atenemos a la disciplina, los profesores de Biblioteconomía y Documentación son de los más críticos con *Cercabib*. Creemos que este hecho se debe a los conocimientos avanzados que estos docentes tienen de las herramientas de búsqueda, lo que les permite identificar como limitaciones o carencias del sistema cuestiones que otros docentes atribuyen a errores propios en el proceso de búsqueda (no saber buscar bien).

Segmentados por antigüedad en la UB, los resultados reflejan una valoración más negativa entre los profesores más nuevos (franja 1991-2012), constatación que se ve

confirmada por los resultados por edad, donde se aprecia que la insatisfacción aumenta a medida que se reduce la edad de los entrevistados.

Los datos por dedicación muestran un ligero aumento de la satisfacción en la experiencia de uso de *Cercabib* entre el profesorado con una mayor actividad docente que investigadora. Aun así, las diferencias con el resto de profesores no son especialmente significativas en este caso.

4.4 Aspectos más positivos y más negativos de *Cercabib*

Los aspectos más positivos y negativos de *Cercabib* expresados por los profesores entrevistados se muestran en las dos tablas siguientes:

Aspectos positivos	Num. Resp.
Mucha Información	5
NS / NC	5
Nada	3
Facilidad de uso	5
Disponibilidad de texto completo	1
Integración con servicios de la UB	1
Relación entre recursos	1
Atractivo visual	1

Aspectos negativos	Num. Resp
Ruido	9
NS/ NS	6
Nada	3
Poco amable	1
Silencio	1
Falta información	1
Visualización deficiente	1

Como se puede apreciar, en el caso de los aspectos positivos las respuestas son muy variadas, si bien destacan la gran cantidad de información que recupera *Cercabib* (tanto en número como en tipología documental) y la facilidad de uso. En este último aspecto vale la pena puntualizar que lo que la mayoría del profesorado destacaba como facilidad de uso hacía referencia a la búsqueda simple tipo *Google* (casilla única) y no tanto otras funcionalidades más especializadas como la búsqueda avanzada o los filtros en la pantalla de resultados.

Como aspecto negativo destaca por encima de todos la gran cantidad de ruido en los resultados de las búsquedas. De una forma u otra, muchos profesores manifestaron su insatisfacción con la ordenación de los resultados de *Cercabib* y la contrapusieron con los criterios de relevancia de otras herramientas (especialmente *Google*) que consideran más ajustados a sus necesidades. De hecho, algunos profesores argumentaron esta carencia de *Cercabib* como uno de los motivos de su poco o nulo uso.

4.5 Preferencia entre el sistema de búsqueda anterior y *Cercabib*

Los resultados de las entrevistas muestran una clara preferencia por el anterior sistema de búsqueda (50%, 11 de 22). Muchos profesores manifestaron sentirse más cómodos con el anterior servicio del CRAI de la UB, y especialmente con el *Catàleg de les biblioteques* (<https://cataleg.ub.edu/>). Para ellos el *Catàleg* era una herramienta donde sabían qué buscaba y dónde lo buscaban. Sabían que los resultados hacían referencia, básicamente, a la colección física de las bibliotecas de la UB; mientras que con *Cercabib* no saben bien a qué corresponden los resultados, de dónde proceden las referencias, por qué algunas parecen enlazables pero de hecho no lo son, etc. En definitiva, la preferencia por el *Catàleg de les biblioteques* obedece a menudo tanto a cuestiones funcionales (ruido en los resultados) como a la incomprensión de qué es *Cercabib*.

5 Conclusiones

Los resultados obtenidos sobre la opinión del profesorado de la Universitat de Barcelona de la nueva herramienta de descubrimiento del CRAI, *Cercabib*, no difieren substancialmente de los estudios sobre herramientas de descubrimiento expuestos en el apartado “Antecedentes”. Como en la mayoría de éstos, la numerosa y diversa información recuperada y la facilidad de uso (búsqueda como *Google*) son los elementos mejor valorados, mientras los peores son el desconocimiento del alcance de la herramienta, la poca fiabilidad de la ordenación por relevancia y el abrumador número de resultados de búsqueda. Es especialmente relevante, y en cierto sentido preocupante, la mayoritaria preferencia por el anterior sistema de búsqueda.

Las posibles soluciones a estos problemas creemos que pasan, en primer lugar, por el uso de nuevas estrategias de difusión de *Cercabib* entre el profesorado, puesto que aun reconociendo el encomiable esfuerzo realizado hasta ahora por el personal del CRAI de la UB (sesiones de formación, vídeos, manuales, difusión a través de las redes sociales, etc.), lo cierto es que buena parte de los docentes entrevistados no saben a ciencia cierta qué es *Cercabib*. En segundo lugar, creemos que hay que mejorar los algoritmos que determinan la ordenación de los resultados de las búsquedas para conseguir una mayor personalización de la recuperación (LIU; LIU; BELKIN, 2019). Sólo de esta manera se podrá superar la sensación de abrumación que impele a muchos profesores a no usar *Cercabib* y a utilizar herramientas de búsqueda con criterios de relevancia más ajustados (por ejemplo, *Google*) o de ámbitos temáticos más específicos que garantizan una recuperación más selectiva.

6 Anexo

Información demográfica

Nombre completo:

Fecha de nacimiento:

Año de inicio de la vinculación con la UB:

Dedicación: docencia (%), investigación (%), aproximado

Disciplina:

Cuestionario sobre *Cercabib*

- a) ¿Te consideras una persona abierta a los cambios tecnológicos?
- b) ¿Qué herramientas de búsqueda usas para encontrar información? *Google*, *Google Scholar*, *Web of Science*, *Scopus*, otros (indica cuáles). ¿Cuál es el primer lugar donde buscas?
- c) ¿Utilizas habitualmente los servicios de las bibliotecas de la UB? ¿Con qué frecuencia?
- d) ¿Sabes qué es *Cercabib*?
- e) ¿Conoces las principales diferencias entre *Cercabib* y el anterior sistema de búsqueda del CRAI de la UB? ¿Cuál te gusta más? ¿Por qué?
- f) ¿Conoces el alcance de *Cercabib* (qué puedes encontrar)?
- g) ¿Cuántas veces a la semana/mes usas *Cercabib*?
- h) ¿Con qué fines usas *Cercabib* (apoyo a la docencia, investigación, motivaciones personales, etc.)?
- i) ¿Qué tipo de documento buscas principalmente en *Cercabib* (libros, artículos de revistas, ponencias de congresos, material audiovisual, etc.)?
- j) ¿Te resulta fácil utilizar *Cercabib*?
- k) ¿Crees que localizas rápidamente la información?
- l) ¿Qué opciones de búsqueda usas? ¿Búsqueda simple, avanzada, refinamiento de los resultados?
- m) ¿Qué esperarías de un sistema de búsqueda de información de la UB?
- n) ¿Qué destacarías como positivo de *Cercabib*?
- o) ¿Qué destacarías como negativo de *Cercabib*?

Referencias

- Aharony, N.; Prebor, G. (2015). Librarians' and Information Professionals' Perspectives Towards Discovery Tools—An Exploratory Study. *The Journal of Academic Librarianship*, v. 41, n. 4, 429–440.

- Ávila-García, L.; Ortiz-Repiso, V.; Rodríguez-Mateos, D. (2015). Herramientas de descubrimiento: ¿una ventanilla única? *Revista Espanola de Documentacion Cientifica*, v. 38, n. 1, 1–17.
- Dalal, H. A.; Kimura, A. K.; Hofmann, M. A. (2015). Searching in the wild: observing information-seeking behavior in a discovery tool. *Proceedings of the ACRL*, 668–675.
- Fawley, N.; Krysak, N. (2014). Learning to love your discovery tool: strategies for integrating a discovery tool in face-to-face, synchronous, and asynchronous instruction. *Public Services Quarterly*, v. 10, n. 4, 283–301.
- González-Teruel, A.; Barrios Cerrejón, M. T. (2012). *Métodos y técnicas para la investigación del comportamiento informacional: fundamentos y nuevos desarrollos*. Gijón: Trea.
- Liu, J.; Liu, C.; Belkin, N. J. (2019). Personalization in text information retrieval: a survey. *Journal of the Association for Information Science and Technology*, Early View.
- Nichols, A. F. et al. (2017). What does it take to make discovery a success?: a survey of discovery tool adoption, instruction, and evaluation among Academic Libraries. *Journal of Web Librarianship*, v. 11, n. 2, 85–104.
- Shapiro, S. (2018). Academic librarians, information overload, and the Tao of discovery. *The Journal of Academic Librarianship*, v. 44, n. 5, 671–673.
- Wells, D.; Richardson, C. (2014). *How do library clients use discovery systems?* Paper presented at LIANZA Conference 2014, Auckland, New Zealand.

A Representação Arquivística na Tradição Canadense: subsídios para elaboração de um modelo ideal de ensino por meio da semântica textual.

Glenda da Rocha Monteiro¹ y Thiago Henrique Bragato Barros²

¹ ORCID [0000-0001-8612-8686](https://orcid.org/0000-0001-8612-8686). Universidade Federal do Pará, , Pará, Brasil.
glendasax@yahoo.com.br

² ORCID [0000-0001-7439-5779](https://orcid.org/0000-0001-7439-5779). Universidade Federal do Rio Grande do Sul Rio Grande do Sul, Brasil.
bragato.barros@ufrgs.br

Resumo: Apresenta um estudo relacionando entre a Linguística e a Arquivologia por meio da semântica e a Representação em Arquivos. O objetivo geral desenvolvido foi contribuir para a construção de um referencial teórico e metodológico a respeito da Representação Arquivística no contexto canadense, visando a criação de um modelo de ensino tendo como base metodológica a semântica textual, e seus objetivos específicos foram analisar instrumentos de pesquisa de duas instituições canadenses; observar os critérios de análise linguística estabelecidos para esta pesquisa, nestes instrumentos; comparar as abordagens em representação nas duas instituições selecionadas. O método empregado para alcançar tais objetivos, caracteriza-se por ser um estudo exploratório, teórico e documental, tendo a semântica e critérios de construção textual de sentido e elementos constituintes da coerência e coesão textual como metodologia de análise para a compreensão dos caminhos da Representação Arquivística nos instrumentos de pesquisa dos Arquivos canadenses- Library and Archives Canada (LAC) e Provincial Archives of Manitoba. Delimitou-se estas duas instituições canadenses de esferas de atuação federal e estadual respectivamente, para a análise das práticas de classificação e descrição de arquivos devido à importância para o cenário do país escolhido, uma vez que, essas instituições apresentam diferentes abordagens em representação.

Palavras-chave: Representação Arquivística. Classificação. Descrição. Linguística. Semântica.

Abstract: It presents a study relating Linguistics and Archival Science through semantics and Archival Representation. The general objective was to contribute to the construction of a theoretical and methodological reference regarding the Archival Representation in the Canadian context, aiming the creation of a teaching model based on textual semantics, and its specific objectives were to analyze research instruments of two Canadian institutions; to observe the linguistic analysis criteria established for this research, in these instruments;

compare the approaches in representation in the two selected institutions. The method used to achieve these objectives is characterized by being an exploratory, theoretical and documentary study, having the semantic and textual construction criteria of meaning and constituent elements of textual coherence and cohesion as a methodology of analysis for understanding the paths of Archival Representation in the research tools of the Canadian Archives - Library and Archives Canada (LAC) and Provincial Archives of Manitoba. We delimited these two Canadian institutions from federal and state action spheres, respectively, to analyze the classification and description of archives practices due to the importance to the scenario of the chosen country, since these institutions present different approaches in representation.

Key-words: Archival Representation. Classification. Description. Linguistics. Semantics

1 Introdução

A representação arquivística faz parte de um desdobramento teórico- conceitual recente que visa integrar as atividades de classificação e descrição enquanto parte do processo de representação, organização e acesso a documentos de arquivo. O ato de classificar e descrever, tem absorvido a multiplicidade de caráter, devido a visão gerada pelas novas concepções sociais e organizacionais surgidas na década de 1980, incluindo mais de uma proveniência ao documento na produção contemporânea e assim expondo o apoio a atividades variadas por um mesmo documento, tornando o campo da descrição contemporânea um espaço dinâmico e contínuo.

A dinâmica contextual explorada para esta pesquisa foi a relação traçada entre a Linguística e a Arquivologia, por meio da semântica textual e a Representação em Arquivos. O Canadá, país enfoque deste trabalho cuja a tradição arquivística é reflexo da Arquivística contemporânea de Schellenberg e do trabalho arrojado e ousado de W. Kyle Lamb, dá início a uma visão particular para arquivística canadense e mundial a partir dos anos de 1950 e 1960, e apresenta mais recentemente Terry Cook e outros autores no debate das temáticas pertinentes ao cenário.

O objetivo geral do estudo desenvolvido foi contribuir para a construção de um referencial teórico e metodológico a respeito da Representação Arquivística no contexto canadense, visando a criação de um modelo de ensino tendo como base metodológica a semântica textual. Os objetivos específicos foram os de analisar instrumentos de pesquisa das duas instituições; observar os critérios de análise linguística, estabelecidos para esta pesquisa, nestes instrumentos; e comparar as abordagens em representação nas duas instituições selecionadas.

A pesquisa apresenta natureza qualitativa, cujo o método aplicado caracterizou-se por ser um estudo exploratório, teórico e documental, tendo a Semântica Textual com uma metodologia de análise para a compreensão dos caminhos da Representação Arquivística por meio dos instrumentos de pesquisa das instituições analisadas. A análise dos dados ocorreu em três fases teórico-metodológicas, baseando-se

principalmente na análise bibliográfica nas áreas de representação e semântica textual e nos produtos da Representação Arquivística nas instituições estabelecidas.

No primeiro momento foi efetuado um levantamento bibliográfico exaustivo a respeito da Representação Arquivística e Semântica, seus princípios teóricos, seus usos na atualidade por parte da Ciência da Informação, aspectos de sua constituição histórica e como efetuar a análise semântica dentro da Ciência da Informação. Para este fim, na primeira etapa foram analisadas bibliografias que estabelecem visões conceituais quanto a Representação Arquivística sob a ótica canadense e autores que contribuem para a construção dos conceitos das funções que constituem o processo representativo da informação.

Delimitou-se duas instituições canadenses de duas esferas de atuação, estadual e federal, para a análise das práticas de classificação e descrição de arquivos, são elas: Library and Archives Canada (LAC) e Provincial Archives of Manitoba. Selecionou-se essas instituições devido à importância no cenário local e federal do país escolhido, uma vez que, essas instituições apresentam diferentes abordagens em representação, por exemplo, em relação a descrição propriamente dita.

2 Organização e representação da Informação Arquivística: aspectos teóricos e conceituais.

A representação arquivística é o ato que busca criar meios que possibilitem o acesso às informações contidas em documentos de arquivos envolvendo processos organizacionais e descritivos de forma fluida. Este processo evolui de acordo com o contexto social ao qual o documento está inserido e a representação procura por meio dos processos de classificação e descrição, estabelecer comunicação entre o órgão produtor da informação e o usuário, se valendo da ordem original e do contexto de produção documental, da criação de ferramentas de acesso e de sistemas resultantes das atividades para desempenhar a representação de maneira eficaz (Yakel, 2003).

Parafraseando (Yakel, 2003) propõe-se a representação como uma atividade ética, que assimila os fatores sociais, culturais, políticos e econômicos nos esquemas de representação, o que ocorre mediante a observação de processos já estabelecidos para a criação de novos processos que melhor compreendam a realidade em questão e as necessidades dos documentos alocados naquele arquivo.

Pensamento este que é ilustrado pelo exemplo da primeira tentativa de padronizar a representação por J. C. Fitzpatrick na Biblioteca do Congresso, suas notas relacionadas ao cuidado, catalogação, calendário e organização dos manuscritos publicadas em 1913, contém questões implícitas quanto a rearranjo e expõe a tensão do tratado de Fitzpatrick quanto a incapacidade de conciliar a representação em um esquema de arranjo mais amplo e a descrição muito detalhada dentro de um mesmo esquema, questão que àquele momento parecia não beneficiar a atividade representacional de maneira eficaz.

O termo representação assim compreende então as funções de classificação e de descrição. Estas duas são as que melhor captam o trabalho realizado pelo profissional de arquivo na organização, reorganização, interpretação, criação de termos substitutos e outros que integram o processo representacional. Yakel (2003) afirma que a classificação arquivística é também um ato representativo, e que mesmo perdido o sentido da ordem original devido a manutenção da massa documental ou a desordem provocada pela acumulação, ainda assim é possível representar a informação.

Na Arquivística Pós-Moderna vários fatores implicam para um constante pensar no fazer da representação. O princípio da proveniência na representação, apresenta multifaces de contextos na produção contemporânea, neste busca-se apoiar atividades variadas e estágios de produção em órgãos diferenciados, onde é necessário estabelecer a conexão entre os contextos de criação e órgãos criadores, de maneira que todas estas sejam inteiramente representadas no momento da classificação, arranjo e descrição do conhecimento (Tognoli, 2012).

A representação não está preocupada somente com o documento em si mesmo, mas com suas funções e seus órgãos produtores (Barros, 2016), assim:

Logo, o estudo da proveniência como um pressuposto para a representação arquivística encontra fulcro não apenas na compreensão do conteúdo imediato do documento, mas também em sua relação com os criadores, as funções e os sistemas de gerenciamento e manutenção. Mais uma vez, todos os contextos devem ser representados no momento da classificação/arranjo e descrição do conhecimento arquivístico. (Tognoli, 2012, p.85).

Fatores como a realidade da instituição arquivística e a realidade social a qual a instituição está inserida influenciam na representação da informação, e o papel do profissional arquivista se faz importante, pois como sujeito que media a relação da informação com o público, é responsável pela construção do saber histórico e gerenciamento do fluxo informacional, sendo ativo no processo de representação (Tognoli, 2012).

As funções de Descrição, Avaliação e Classificação são importantes para a organização da informação e permissão de acesso por parte do usuário, pois são metodologias de difusão em que a representação da informação encontra seu alicerce para a ampliação e renovação de seus modos de fazer e instrumentos (Barros & Martins, 2015).

3 Linguística e semântica: aplicação em Arquivologia e CI

A linguística aplicada à Ciência da Informação, em especial ao que se relaciona a textos, vem atrelar o estudo da linguagem aos processos de análise- classificação e descrição dos documentos. A análise documental possibilita o entendimento do assunto documental para que sejam tomadas decisões cabíveis quanto aquele documento, de modo que na descrição a compreensão do profissional deva levá-lo a extrair as informações mais importantes do documento a fim de elaborar um resumo para uma posterior representação da informação.

Como auxiliar no processo descritivo e de organização para a representação a Linguística, em especial a semântica textual, contribui como método para melhor construção das descrições, por meio do estudo da linguagem e também das figuras de linguagem que possam estar contidas no corpo dos documentos e que esclareçam o contexto de criação e tramitação tornando-a mais completa e clara Quanto a relação da Linguística e Ciência da informação, Mendonça (2000, p. 50) afirma:

A contribuição da linguística, via linguagem, e a função social da ciência da informação, via comunicação, fornecem uma das questões prioritárias no meio documental neste final de século. Foucault diz que “a existência da linguagem é soberana, pois que as palavras receberam a tarefa e o poder de representar o pensamento”. Pensamento representado pela extensão do vocabulário e pela fertilidade de seus elementos na qual se incluem dois modos operantes: o fixo, em que a palavra é a base da construção limitada, restrita, e o flexível, em que as palavras têm múltiplas definições e são dotadas de combinações associativas, segundo Moles.

A linguística por meio da semântica e da semiótica propiciam bases para o melhor entendimento das palavras e contextos que envolvem o assunto a respeito de um campo ou a área a ser retratado. Na Arquivologia, os instrumentos organizacionais têm papel semelhante aos instrumentos de pesquisa com o princípio de promover o acesso a informação de forma eficaz e que também sofrem interferência do estudo da semântica.

A orientação dos conteúdos que formam um fundo, uma série ou subsérie são melhor esclarecidos por meio da análise dos sentidos e significados expressos pelas palavras e pelo contexto que as envolvem e que revelam as intenções do documento e produtores (sua função e o processo a qual se inclui).

Martins e Barros (2017) esclarecem que a classificação e a descrição são funções resultados da realização de atividades analíticas e de pesquisa específicas empregadas nos arquivos/acervos e estas atividades resultam em pesquisas que acabam por ser representações textuais, geralmente em formatos de plano de classificação e instrumentos de pesquisa.

Esta análise influencia em todo processo de gestão documental, no qual cada etapa é interdependente da anterior, ou seja, se há a compreensão da “ação” do documento e suas palavras e sentidos, há a possibilidade maior de serem representados de maneira que o usuário tenha clareza na localização das informações, por isso a Linguística textual, em especial o estudo da semântica aplicada ao texto, tem importância para as atividades representacionais, pois pode estabelecer parâmetros linguístico-textuais condutores e normalizadores que auxiliem na produção e análise dos textos.

3.1 Competências linguísticas: construção textual de sentido e elementos constituintes da coerência e coesão textual

As categorias de análise deste trabalho foram estabelecidas de acordo com os princípios Apresentados por Koch (2015) e Fávero e Koch (2012), de Construção Textual de Sentido e Elementos Constituintes da Coerência e Coesão Textual, tais como: coesão textual, coerência, situacionalidade, informatividade, intertextualidade,

intencionalidade, aceitabilidade, fatores de contextualização; consistência e relevância e focalização.

No desenvolver dos estudos linguísticos que envolvem a semântica e suas nuances se revela um vasto campo de exploração. Observa-se que há a possibilidade de desdobramento em relação ao “sentido das palavras”, indo em direção a questão da interpretação e aos fatores de coerência e construção de sentido são próprias do ser humano. Almeida (2005), trabalha a ideia de que a capacidade de alguém ler e ouvir não o torna um simples receptor, mas parte da construção de sentido do item lido ou ouvido. Para corroborar com essa afirmativa Almeida (2005) cita Orlandi (2004, p. 58): “Não é só quem escreve que significa; quem lê também produz sentidos. E o faz, não como algo que se dá abstratamente, mas em condições determinadas, cuja especificidade está em serem sócio-históricas.”.

A partir desta afirmativa entende-se que no momento de leitura de um texto há a interação da capacidade cognitiva do indivíduo de abstrair o sentido e o que está escrito, fazendo da leitura um ato social indireto entre autor e leitor: o texto como um produto não fechado e as expectativas e cargas de conhecimentos que o leitor possui e infere enquanto faz a leitura. Por meio da leitura então o leitor acaba por torna-se co-autor do texto devido a interpretação que faz deste.

[...]segundo o conceito de Orlandi (2004b:64), de que interpretar é —explicitar o modo como um objeto simbólico produz sentidos. Assim, reitera-se que leitor não é um receptor passivo, interage no processo de leitura, tornando-se co-autor, atribuindo às palavras do texto outros sentidos. (Almeida, 2005, p. 20).

Quanto a elaboração do sentido das palavras ou do conjunto delas perpassa aos aspectos semânticos, e encontra a direção da construção de sentidos durante a leitura. O autor emite sua ideia, que expressa um sentido semântico próprio ocasionado pela escolha das palavras na construção do texto e a interpretação da ideia é avariada pela interferência da passagem do tempo, pois as palavras mudam de forma e significado e também pelos aspectos sociais e culturais que separam o autor do leitor.

Dentro da construção do sentido são observados aspectos que auxiliam o autor na produção e o leitor na interpretação do texto. Estes fatores e critérios formam a construção textual de sentido e elementos constituintes da coerência e coesão textual. Estes, segundo Martins e Barros (2017) são abordados por Koch (2015) e Fávero e Koch (2012); e Almeida (2005) expõe Koch e Travaglia (2003), todos estes autores tem por base os elementos elencados por Beugrande e Dressler (1981). Martins e Barros (2017) apresenta os elementos descritos pelas autoras, aqui sintetizamos em apenas um quadro:

Quadro 1. Definições dos critérios de construção textual de sentido e elementos constituintes da coerência e coesão textual.

CRITÉRIOS E ELEMENTOS CONSTITUINTES	DEFINIÇÕES
-------------------------------------	------------

COESÃO TEXTUAL	A forma como os elementos linguísticos presentes na superfície textual se interligam se interconectam, por meio de recursos também linguísticos, de modo a formar um “tecido” (tessitura), uma unidade de nível superior à da frase, que dela difere qualitativamente. (Koch, 2015, p. 45)
COERÊNCIA	Diz respeito ao modo como os elementos subjacentes à superfície textual entram numa configuração veiculadora de sentidos. (Koch, 2015, p. 49)
SITUACIONALIDADE	No primeiro sentido, a situacionalidade refere-se ao conjunto de fatores que tornam um texto relevante para a situação comunicativa em curso ou passível de ser reconstituída. Trata-se, neste caso, de determinar em que medida a situação comunicativa, tanto o contexto imediato de situação como o entorno sócio-político-cultural em que a interação está inserida, interfere na produção/recepção do texto, determinando escolhas em termos, por exemplo, de grau de formalidade, regras de polidez, variedade linguística a ser empregada, tratamento a ser dado ao tema etc. No segundo sentido, é preciso lembrar que o texto tem reflexos importantes sobre a situação, visto que o mundo textual não é jamais idêntico ao mundo real.. (Koch, 2015, p. 49)
INFORMATIVIDADE	Quanto à distribuição da informação, é preciso que haja um equilíbrio entre a informação dada e a informação nova.[...] Quanto ao grau de previsibilidade ou expectabilidade da informação, um texto será tanto menos informativo quanto mais previsível (redundante) for a informação que trás. Há, portanto, graus de informatividade: um texto cuja informação seja toda apresentada da forma mais previsível terá baixo grau de informatividade; se a informação for introduzida, pelo menos em parte, de forma menos esperada, menos previsível, haverá um grau médio de informatividade; e, se toda informação for apresentada de maneira imprevisível, o texto terá um grau máximo de informatividade e exigirá um grande esforço de processamento, podendo assim, à primeira vista, parecer pouco coerente. (Koch, 2015, p. 50)
INTERTEXTUALIDADE	Compreende as diversas maneiras pelas quais a produção/recepção de um dado texto depende do conhecimento de outros textos por parte dos interlocutores, ou seja, dos diversos tipos de relação que um texto mantém com outros textos. (Koch, 2015, p. 51)
INTENCIONALIDADE	Refere-se aos diversos modos como os sujeitos usam textos para perseguir e realizar suas intenções comunicativas, mobilizando, para tanto, os recursos adequados à concretização dos objetivos visados; em sentido restrito, refere-se à intenção do locutor de produzir uma manifestação linguística coesa e coerente, ainda que essa intenção nem sempre se realize integralmente [...]. (Koch, 2015, p. 51)

ACEITABILIDADE	É a contraparte da intencionalidade. Refere-se à concordância do parceiro em entrar num “jogo de atuação comunicativa” e agir de acordo com suas regras, fazendo o possível para levá-lo a um bom termo, visto que, como postula Grice (1975), a comunicação humana é regida pelo Princípio de Cooperação. (Koch, 2015, p. 51).
FATORES DE CONTEXTUALIZAÇÃO	Para Maruschi (1983) é preciso que se inclua este fator pelo fato de ancorar o texto nas situações de comunicação. Dos quais propõe dois subtipos: os contextualizadores propriamente ditos (data, local, assinatura, timbre, em documentos oficiais, diagramação, localização na página ou em cadernos, em se tratando de textos jornalísticos, recursos gráficos em geral); e prospectivos, que permitem avançar expectativas sobre o texto (título, nome do autor, início do texto). Esses elementos são, muitas vezes, decisivos para a interpretação (Koch, 2015, p. 52)
CONSISTÊNCIA E RELEVÂNCIA	Apresentado por Giora (1985) compreende que a condição de consistência exige que todos os enunciados de um texto possam ser verdadeiros, isto é, não contraditórios dentro de um mesmo mundo ou dentro dos diversos mundos representados num texto. O critério da relevância exige que o conjunto de enunciados que compõem o texto seja relevante para um mesmo tópico discursivo, isto é, que os enunciados sejam interpretáveis como predicando algo sobre um mesmo tema. Assim, a relevância não se dá linearmente entre pares de enunciados, mas entre conjuntos de enunciados e um tópico discursivo (Koch, 2015, p. 53).
FOCALIZAÇÃO	A focalização permite determinar, o significado, no texto, de palavras homônimas e polissêmicas, bem como o uso adequado de certos elementos linguísticos de valor dêitico, como é o caso dos verbos ir e vir, cujo emprego depende da direção do movimento focalizado. A focalização é, ainda, responsável pela escolha de descrições ou expressões nominais, na construção/reconstrução de referentes textuais. (Koch, 2015, p. 53)

Fonte: Martins e Barros, 2017– adaptado pelos autores.

Estes critérios e elementos se participativos no processo da construção de texto levarão o leitor a compreender melhor a informação e a construção de sentido a ser expressa pelo autor em seu texto, mas também no próprio leitor mantendo o aspecto interativo entre os dois envolvidos neste mecanismo. O principal aspecto influenciador no processo da produção de sentido dos textos e que contribui para a funcionalidade dos critérios de construção textual de sentido e elementos constituintes da coerência e coesão textual é a semântica.

4 Resultados

De acordo com as categorias de análise estabelecidas, escolheu-se três instrumentos de pesquisa de instituições canadenses, sendo dois da Library and Archives Canada (LAC) e um do Provincial Archives of Manitoba. A configuração dos instrumentos consultados nesta parte da pesquisa é utilizada para o acesso e recuperação de informações on-line. As disposições e formas no site dos Arquivos, a princípio causam certo estranhamento devido aos formatos bem diferenciados ao que é comum a realidade brasileira.

Esta diferença de disponibilidade de dados aos usuários foi levada em consideração na análise final apresentada neste capítulo, o que pode deixar, a princípio, como um estudo de usabilidade do site, porém a principal intenção da observação do funcionamento das páginas é pela configuração em que se encontram os dados representados. Vale lembrar também que as informações analisadas estão dispostas num sistema on-line, passivo de modificação a qualquer momento.

Disponíveis em formato on-line os instrumentos estão dispostos como parte da navegação do site, este fator levou para a demora da compreensão do que seria o instrumento de pesquisa em si. A configuração de rede interligada entre Arquivos e a integração entre Arquivo e Bibliotecas torna possível a recuperação de informações dos mais variados lugares, esta é a realidade aplicada ao contexto canadense, arquivos em rede.

No caso do LAC, o material analisado estava disposto na forma de conjuntos pertencentes a um assunto/conteúdo provenientes de fundos diversificados e com formatos e suportes diversos. Ao pesquisar no site o usuário deve estar atento a configuração do resultado. Escolheu-se o tema Music e mediante esta escolha abriu-se um catálogo com uma infinidade de documentos, fotos e livros sobre esta temática. Selecionou-se apenas os relacionados a Arquivo, obtendo-se no total 22.344 de arquivos. Dois acervos foram selecionados dentro do catálogo do LAC: o arquivo do Alliance Communications Corporation and A & F Music Limited - Purchase of shares of Partisan Music Productions Incorporated- 1992-1994 e o arquivo Block 55 - Cultural Affairs [textual record].

O arquivo do Alliance Communications Corporation and A & F Music Limited - Purchase of shares of Partisan Music Productions Incorporated- 1992-1994, segundo a descrição contém, entre outros itens correspondências, contratos e registros financeiros da referida empresa. Segundo informações descritas na página, este faz parte do Arquivo General management [textual record, graphic material] e encontra-se vinculado a sub- séries que consistem a 375 descrições de nível inferior incluindo várias mídias em sua constituição.

A página do LAC apresenta um glossário de termos que aponta termos e denominações próprias utilizadas nas descrições, como exemplo o que é arranjo, extensão, e significação dos números relacionados aos códigos de acesso e outros. Este glossário é aqui considerado o fator de intertextualidade que possibilita o entendimento das partes quando se abre o referido arquivo da Alliance, porém

acredita-se que pela pequena extensão, apenas 2cm (em metro linear), deixe sua descrição no site empobrecida, apesar de seus conteúdos estarem classificados como registros abertos – código 90 (Open Records -code 90).

A descrição de Alliance Communications Corporation and A & F Music Limited - Purchase of shares of Partisan Music Productions Incorporated- 1992-1994, apresenta coesão textual, coerência, situacionalidade, intertextualidade (em referência ao glossário de termos e as descrições de outras sub- séries), aceitabilidade, fatores de contextualização; consistência e relevância. Porém os quesitos de informatividade, intencionalidade e focalização, são afetados pela ausência de textos mais detalhados, pois alguns dos itens apresentam apenas códigos, restringindo a análise semântica.

Quadro 2. Análise Textual do Alliance Communications Corporation and A & F Music Limited- Purchase of shares of Partisan Music Productions Incorporated- 1992-1994.

CRITÉRIOS E ELEMENTOS CONSTITUINTES	ANÁLISE
COESÃO TEXTUAL	Os textos que compõe as seções da descrição deixam a relação entre os itens descritivos visíveis, por exemplo: a relação do título do fundo expressa muito bem os dados descritos em datas, parte da estrutura de arranjo, âmbito e conteúdo e linguagem do material.
COERÊNCIA	Estrutura as ideias do texto de modo lógico, compreendendo itens propostos pelas Normas de Descrição.
SITUACIONALIDADE	É indicada pela presença de datas, âmbito e conteúdo e indicação de séries e subséries relacionada
INFORMATIVIDADE	Visualizada apenas nesta descrição como parte da ação informativa: intenção principal da descrição
INTERTEXTUALIDADE	Compreendendo o conceito exposto por Koch (2015) para este elemento observa-se que a intertextualidade ocorre devido a relação que este traça com o glossário de termos da página
INTENCIONALIDADE	Visualizada apenas nesta descrição como parte da ação informativa: intenção principal da descrição.
ACEITABILIDADE	Apesar de o fator de INTENCIONALIDADE estar comprometido pela ausência de itens semânticos que dê a contrapartida para o quesito de ACEITABILIDADE, este se considera este existente devido a existência de informações contidas no todo da descrição
FATORES DE CONTEXTUALIZAÇÃO	É proposto principalmente pelos itens de título, pelo campo de âmbito e conteúdo e data.
CONSISTÊNCIA E RELEVÂNCIA	Apresenta como meio de confirmação de veracidade a influência da Instituição publicadora da informação e não apresentar contradição ao longo da descrição. A RELEVÂNCIA é constatada pela função de informar e indicar conteúdos próprios da descrição arquivística.
FOCALIZAÇÃO	Apresenta ausência de textos com itens de valor semântico que supram este critério

Fonte: Elaborado pelos autores. Dados da pesquisa 2018.

Outro instrumento analisado da LAC, foi o Block 55 - Cultural Affairs [textual record], segundo ao campo de âmbito e conteúdo este consiste em registros do Bloco 55 - Assuntos Culturais, parte do principal sistema de registro de departamentos que foi criado em 1963 e era conhecido como a série 63. Esse sistema foi projetado para simplificar o sistema de arquivamento departamental, atribuindo números de blocos de arquivos principais às principais áreas de atividade. Esses blocos de arquivos são então, divididos em tópicos comuns, como relatórios, tratados e legislação. Os arquivos também podem incluir designadores para países individuais ou organizações internacionais.

A extensão do acervo é grande, pois conecta-se a subséries. Compreende os anos de 1963 a 1994 e sua adesão (grupo de registros ou arquivos da mesma entidade - fonte, com o mesmo local de origem ou histórico- proveniência, aceito todos de uma vez como propriedades em um repositório arquivístico) é de 16 volumes que se desdobram de acordo com os anos de registro e subséries associadas com restrições variadas. A intertextualidade é manifesta neste inventário por meio da conexão que suas partes têm com o glossário de termos do site, já citado anteriormente.

Por esta descrição já verificamos pontos semânticos de coesão textual, coerência, situacionalidade, informatividade e focalização contidas na descrição, expressas por linguagem simples que preenchem campos com textos que trazem estes elementos, porém o não estabelecimento claro quanto a procedência exata dos documentos acarreta uma perda de parte da consistência e relevância e fatores de contextualização que acabam por ficar comprometidos, pois há procedência governamental, mas não especifica de quais departamentos são exatamente na descrição.

Quadro 3. Análise Textual do Block 55 - Cultural Affairs [textual record].

CRITÉRIOS E ELEMENTOS CONSTITUINTES	ANÁLISE
COESÃO TEXTUAL	Há a existência de COESÃO TEXTUAL mediante a relação notável entre as partes da descritas.
COERÊNCIA	Estrutura as ideias do texto de modo lógico, compreendendo itens propostos pelas Normas de Descrição.
SITUACIONALIDADE	É indicada pela presença de datas, âmbito e conteúdo e indicação de séries e subséries relacionadas.
INFORMATIVIDADE	O texto da descrição cumpre o papel de informar apesar de haver lacunas referentes ao FATORES DE CONTEXTUALIZAÇÃO.
INTERTEXTUALIDADE	Compreendendo o conceito exposto por Koch (2015) para este elemento observa-se que a intertextualidade ocorre devido a relação que este traça com o glossário de termos da página.
	Visualizada apenas nesta descrição como parte da ação

INTENCIONALIDADE	informativa: intenção principal da descrição.
ACEITABILIDADE	É percebida pela noção e entendimento proporcionado pela leitura da descrição.
FATORES DE CONTEXTUALIZAÇÃO	Afetado pela ausência de informações relevantes que indiquem a procedência exata dos documentos na descrição.
CONSISTÊNCIA E RELEVÂNCIA	Afetado pela ausência de informações relevantes que indiquem a procedência dos documentos, porém se faz relevante a medida que se compreende o conjunto de descrições deste.
FOCALIZAÇÃO	Este fator se faz presente pela coerência dos textos apontando sentido referente ao objeto descrito e tendo possibilidade de as partes textuais da descrição serem utilizadas para reescrever novas descrições do objeto a partir dela.

Fonte: Elaborado pelos autores. Dados da pesquisa 2018.

Os instrumentos de pesquisa escolhido do Provincial Archives of Manitoba foi o Constitutional Reform Files, trata-se de arquivos referentes a reforma constitucional incluem cópias de atas de conferências / reuniões constitucionais e material de apoio; documentos informativos e pareceres jurídicos elaborados pela sucursal ou outras jurisdições; e correspondência relacionada. Este, em seu formato no site, assemelha-se mais ao formato brasileiro de instrumento de pesquisa.

Mais extenso que o primeiro analisado da LAC, a coleção Constitutional Reform Files apresenta 7 registros datados de 1991 a 2005. A descrição geral da coleção apresenta coesão textual, coerência, situacionalidade, informatividade, intertextualidade, intencionalidade, aceitabilidade, consistência e relevância, focalização e fatores de contextualização, incluindo estes itens a outras coleções relacionadas a Constitutional Law.

Porém no acesso a lista online para visualizar registros relevantes relacionados a esta coleção, as únicas informações disponíveis nestas descrições é o número de acesso, tamanho do arquivo, formato e data, deixando algumas características semânticas nulas devido à falta de texto descritivo nelas.

Quadro 4- Análise Textual do Constitutional Reform Files.

CRITÉRIOS E ELEMENTOS CONSTITUINTES	ANÁLISE
COESÃO TEXTUAL	As partes da descrição estão interligadas por itens linguísticos que conferem a descrição unidade temática e de sentido.
COERÊNCIA	Apresenta sentido único mediante as informações dispostas nos campos descritivos a respeito do objeto descrito.
SITUACIONALIDADE	Preenchida pelos campos de data, escopo e conteúdo e notas.

INFORMATIVIDADE	Permeia toda a descrição do fundo por meio das informações apresentadas em todos os campos descritivos.
INTERTEXTUALIDADE	Expressado pela indicação de relação do fundo com a História, a Constituição e a reforma constitucional.
INTENCIONALIDADE	Visualizada apenas nesta descrição como parte da ação informativa: intenção principal da descrição
ACEITABILIDADE	É percebida pela noção e entendimento proporcionado pela leitura da descrição.
FATORES DE CONTEXTUALIZAÇÃO	Apresenta datas, faz referências as possíveis legislações as quais este arquivo está sujeito, e dados da constituição do fundo.
CONSISTÊNCIA E RELEVÂNCIA	Apresenta como meio de confirmação de veracidade a influência da Instituição publicadora da informação e não apresentar contradição ao longo da descrição. A RELEVÂNCIA é constatada pela função de informar e indicar conteúdos próprios da descrição arquivística.
FOCALIZAÇÃO	As partes textuais da descrição podem ser utilizadas para reescrever novas descrições do objeto a partir dela

Fonte: Elaborado pelos autores. Dados da pesquisa 2018.

5 Considerações finais

A perspectiva da representação e organização em Arquivos, juntamente com o imperativo tecnológico crescente no final do século XX e início do século XXI, traz a facilidade de propagação e acesso via web dos trabalhos arquivísticos dos mais variados lugares do mundo. Percebe-se que a configuração de arquivos em rede, a união entre Biblioteca e outros órgãos gestores de informação social e técnico-científicas no Canadá é um dispositivo comum de divulgação, disponibilização e acesso a serviços, conteúdos e ações variadas a respeito do fluxo informacional gerado por departamentos, público ou a privado em determinada região.

Observa-se também, que os estudos linguísticos vêm desenvolvendo relação com a Arquivologia mediante a necessidade de reflexão nos processos de criação de elementos relacionados a organização e representação dos documentos e da informação, em especial, os instrumentos de pesquisa.

A análise de instrumentos de pesquisa de duas instituições públicas canadenses por meio de aspectos linguístico expressa parte da relação em aprofundamento entre Arquivologia- Representação e a Linguística- campo da Semântica, visto que há certa quantidade de trabalhos que envolvem questões teóricas entre ambas áreas, porém que pouco exploram proposições práticas.

Os instrumentos de pesquisa, quer sejam eles catálogos, índices, guias ou inventários, nas duas instituições estão dispostos on-line e são configurados como parte dos sistemas de navegação dos sites das duas instituições. O tipo de organização correspondente a realidade canadense, apresenta facilidades e também ressalta a integração dos contextos implícitos aos documentos, em especial aos de Arquivos.

As descrições analisadas transparecem as interconexões de dados e os múltiplos contextos aos quais estas informações pertencem. No entanto, assim, torna-se uma rede complexa, em vista que o usuário deve estar bem familiarizado com o sistema de busca para efetuar pesquisa, mesmo com as duas instituições oferecendo suporte on-line para as atividades de pesquisa.

A linguagem dos temas visualizados nos itens das páginas das instituições que possuíam descrição, apresenta funcionalidades semânticas, de acordo com as categorias de análise aplicadas a este estudo: linguagem clara, referência a outros conteúdos de modo explícitos, descrição de âmbito e histórico, apresentados de forma concisa.

Contudo, ao retomarmos aos objetivos da pesquisa, observa-se que a contribuição para a construção de referencial teórico foi alcançada e possível, assim como a instituição de uma metodologia baseada na semântica textual por meio de categorias semânticas para a análise.

Referências

- Almeida, V. C. de. (2005). *A interpretação de textos com base nos fatores de coerência*. Dissertação de Mestrado em Língua Portuguesa, Faculdade de Letras/UFRJ,. Rio de Janeiro, Brasil.
- Barros, T. H. B. (2016). A Indexação e a Arquivística: aproximações iniciais no universo teórico da organização e representação do conhecimento. Encontros *Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 21:46, 33-44.
- Barros, T. H. B., & Martins, W. R. (2015). A informação orgânica enquanto um objeto interdisciplinar: as relações entre a arquivística e a ciência da informação no âmbito da representação em arquivos. *ÁGORA*, Florianópolis, v. 25, n. 51, 132-149.
- Fávero, L. L., & Koch, I. G. V. (2012). *Linguística Textual: introdução*. – 10. Ed. – São Paulo: Cortez.
- Koch, I. V. (2015). *Introdução à Linguística Textual: trajetória e grandes temas*. – 2. ed. – São Paulo: Contexto.
- Martins, W. R., & Barros, T. H. B. (2017, novembro). Estruturas semânticas da representação do conhecimento arquivístico: analisando elementos estruturantes e de conteúdo. *Atas do III Congresso ISKO Espanha-Portugal & XIII Congresso ISKO Espanha- Coimbra*, Beira Litoral, Portugal, 23 e 24.
- Mendonça, E.S. (2000). *A Linguística e a Ciência da Informação: estudos de uma interseção*. Ci. Inf., Brasília, v. 29, n. 3, 50-70.
- Tognoli, N. B. (2012). A representação na arquivística contemporânea. *Ci. Inf*, Brasília, 5: 2, 79-92
- Yakel, E. (2003). Archival Representation. *Archival Science*, n. 3, 1-25.

e-EMGDE, RiC, NEDA y las normas de codificación: más allá de un perfil de aplicación en archivos

Baños-Moreno, M.J.¹, Valentín Ruiz, F.J.² y Blázquez Martín de las Mulas, A.³

¹ ORCID [0000-0001-9137-1330](https://orcid.org/0000-0001-9137-1330). Odilo, España.
mbm41963@um.es

² ORCID [0000-0001-5343-920X](https://orcid.org/0000-0001-5343-920X). Odilo, España.
fjvalentin@odilotid.es

³ ORCID [0000-0003-3918-986X](https://orcid.org/0000-0003-3918-986X). Archivero, España.
tblazquez2980@gmail.com

Resumen. La variedad de estándares de descripción archivística dificulta enormemente la materialización en aplicaciones que cumplan con todas las directrices. Esta es la situación que nos encontramos en el camino de adaptar la solución Odilo A3W, a un conjunto de estándares, lo que nos ha llevado a estudiar los diferentes modelos para hacerlos compatibles entre sí y con la evolución tecnológica de A3W. El objetivo es explicar cómo se ha definido el modelo conceptual de A3W en lo que se refiere a descripción archivística: entidades, relaciones y atributos; es decir, cómo se han compatibilizado RiC, NEDA, EAD y e-EMGDE y cómo se ha adaptado la aplicación para incorporarlos. Siendo ésta la primera fase de un proyecto de mayor calado, en esta etapa nos hemos centrado en el análisis de la entidad «documento de archivo», vital para la estructuración de la información archivística, y en sus atributos, en un contexto de convivencia de gestión de la documentación en papel y electrónica. Esto implica el estudio pormenorizado de las especificaciones de e-EMGDE, NEDA, RiC y EAD. Como resultado, se han mapeado los atributos de descripción archivística asociados con la entidad «documento de archivo» de odiloa3w con los definidos para los modelos indicados anteriormente.

Palabras clave: odiloa3w; e-EMGDE; RiC; NEDA; EAD

Abstract. The variety of archival description standards makes difficult the materialization in applications that comply with all requirements. We find ourselves in this situation on the way of adapting the Odilo A3W solution to a set of standards. That has led us to study the different models to make them compatible with each other and with the technological evolution of A3W. The objective is to explain how the A3W conceptual model has been defined in terms of archival description: entities, relationships and attributes; that is, how

the RiC, NEDA, EAD and e-EMGDE standards have been made compatible and how the application has been adapted to incorporate them. This communication constitutes the first phase of a major project. So, at this stage we focused on the analysis of the entity "archival document" and its attributes, in a context of coexistence of paper and electronic document management. This implies the detailed study of the e-EMGDE, NEDA, RiC and EAD specifications. As a result, the archival description attributes associated with the "file document" entity of odiloa3w have been mapped with those defined for the models indicated above.

Keywords: odiloa3w; e-EMGDE; RiC; NEDA; EAD

1 Modelos conceptuales, esquemas de metadatos y estándares de codificación de información en archivística: una aproximación

Muchos son los cambios que se han producido en materia de descripción archivística durante la última década. Uno de ellos es la aparición de estándares de descripción archivística basados en el modelo entidad-relación y de vocabularios especializados. El motivo de esta proliferación hay que buscarlo en la propia evolución de la disciplina, con el objetivo de adaptarse a los cambios tecnológicos y nuevas situaciones.

Así, era necesario construir un modelo conceptual que asentara los fundamentos de la descripción archivística más allá de los modelos o guías existentes hasta la fecha, que estaban basados en casos prácticos, como ocurre con ISAD (G). Este hecho tiene especial importancia en un contexto que tiende hacia la digitalización de actividades que tradicionalmente eran humanas (entendiendo digitalización en su concepción más amplia) y en el que la proliferación de nuevas realidades tecnológicas susceptibles de archivo está a la orden del día. En consecuencia, es lógico que, en entornos altamente tecnológicos, surjan modelos conceptuales que permitan sistematizar y guiar el diseño, desarrollo e implantación de aplicaciones informáticas orientadas a la gestión y difusión de documentación en fase de archivo.

Pero no es la única justificación. El surgimiento de estos estándares y vocabularios también se ve influenciado por la realidad que impone la administración electrónica y por una búsqueda permanente de la estandarización, entre otras cuestiones.

En resumen, nos encontramos con diferentes estándares que es necesario identificar y compatibilizar, también con herramientas precedentes: no se trata de empezar de nuevo o de realizar nuevas creaciones conceptuales. Como decimos, es fundamental identificar los diferentes modelos y estándares que deben ser tenidos en cuenta en la construcción de aplicaciones informáticas. Entre estos destacan:

1. Modelos conceptuales de descripción archivística, como las NEDA (Normas Españolas de Descripción Archivística) y RiC (Records in Context).
2. Esquemas de metadatos, utilizados para la representación de aspectos formales, administrativos y de contenido de documentos, muy vinculados a

la Web, como e-EMGDE (Esquema de Metadatos para la Gestión del Documento Electrónico).

3. Estándares de codificación de información, con una larga trayectoria, pero aún vigentes y en convivencia con los nuevos modelos. Estos son EAD (Descripción Archivística Codificada), EAC (Contexto Archivístico Codificado) y EAG (Guía de Archivo Codificada).
4. Estándares de descripción precedentes que, hasta la actualidad, han sido la base de la descripción archivística y están llamados a mantener su compatibilidad (en cuanto a guías para realizar descripciones) con los nuevos modelos. Encontramos: ISAD (G) (Norma Internacional General de Descripción Archivística); ISAAR (CPF) (Norma Internacional sobre los Registros de Autoridad de Archivos relativos a Instituciones, Personas y Familias); ISDIAH (Norma Internacional para la Descripción de Instituciones que Custodian Fondos de Archivo); e ISDF (Norma Internacional para la Descripción de Funciones).

2 Un baño de realidad: el desarrollo de aplicaciones de gestión de archivos y el caso de Odilo A3W

Ahora bien, la proliferación de estas herramientas no siempre se ha traducido en una materialización real en las aplicaciones informáticas de gestión de archivos o su implantación ha sido más lenta de lo que se esperaba. La situación se complica aún más con la llegada del expediente electrónico (Popovici, 2016). El resultado es que las organizaciones orientadas a los servicios archivísticos se encuentran (nos encontramos) con una variedad de estándares, vocabularios y directrices que no siempre se orientan al desarrollo de las propias aplicaciones y que, en ocasiones, no son fáciles de alinear. Además, hay que recordar que los estándares no son únicos, se han elaborado por diferentes comisiones o grupos de trabajo y, aunque persigan fines similares, cuentan con variaciones, lo que dificulta enormemente la materialización en aplicaciones prácticas que cumplan con todas las directrices existentes.

Como indica la Comisión de Normas Españolas de Descripción Archivística - CNEDA- (2017, pp. 7), existen esquemas y estándares aplicados a la descripción archivística, incluso con propósitos distintos a los de las NEDA, pero que no son incompatibles con los elementos reconocidos por este modelo. Efectivamente, no puede ser incompatible con otros modelos existentes o en desarrollo, o con guías que han venido utilizándose tradicionalmente en los archivos (como ISAD-G). Sin embargo, una mera declaración del tipo, «no son incompatibles» no necesariamente se traduce en facilitar el diseño, desarrollo e implantación de aplicaciones informáticas.

Esta es la situación que nos hemos encontrado en el camino de adaptar una solución especializada, Odilo A3W, a un conjunto de estándares existentes, a través de un perfil de aplicación. Esto implica estudiar los diferentes modelos y aunarlos para hacerlos compatibles entre sí y con la evolución tecnológica del producto. Al mismo tiempo, debemos asegurar que el resultado final sea eficiente, dando respuesta

a las necesidades de los archivos. Esto es, que la nueva aplicación A3W no se transforme en un engendro inutilizable, que el cumplimiento de los estándares no sea incompatible con funcionalidad. Por lo que respecta a la aplicación, A3W fue inicialmente desarrollada para la gestión de documentación en soporte físico. El producto ha venido evolucionando para incorporar otras realidades como la documentación electrónica digitalizada (asociada a los documentos en soporte físico descritos en la aplicación) y, más recientemente, para adaptarse a la gestión de documentación generada en el contexto de la administración electrónica y que cuenta con unas características muy concretas. A3W abarca todos los procesos necesarios con la fase de archivo de la documentación, desde el ingreso, descripción y transferencia, hasta la conservación definitiva, preservación digital, difusión o destrucción. Su eficacia se basa en el tratamiento integrado de toda la documentación (independientemente de su formato físico o digital), permitiendo la configuración y gestión completa de las tareas de gestión documental relacionadas con el archivo y la custodia de la documentación.

3 Objetivos del proyecto y de la comunicación

Inicialmente este proyecto se orientó a la creación de un perfil de aplicación. Sin embargo, dadas las características de algunos de los estándares y herramientas contemplados, la estructura generalmente adoptada y la no definición en todos los casos de un esquema fijo de metadatos, nos ha llevado a ampliar el alcance del trabajo.

De este modo, el objetivo del proyecto es definir el modelo conceptual de la aplicación Odilo A3W en lo que se refiere a descripción archivística: entidades, relaciones y atributos implicados. Es decir, compatibilizar e-EMGDE, RiC, NEDA, EAD, EAC y EAG entre sí e incorporarlos al programa Odilo A3W. Todos estos son elementos reconocidos por la comunidad, altamente comprometida con la estandarización, especialmente en el sector público, y están siendo objeto de análisis por distintas organizaciones y colectivos para su empleo en archivos, que apuestan por el ámbito electrónico, al que se encaminan los archivos.

Además del modelo, se persigue la definición de un perfil de aplicación, odiloa3w, que debe facilitar la migración de datos y de tecnologías, así como la visualización de información archivística desde distintas perspectivas, teniendo en cuenta los distintos propósitos de cada uno y su utilización simultánea por organizaciones archivísticas.

En cuanto a los objetivos de la comunicación, debido a la amplitud del proyecto y por cuestiones de extensión del texto, el propósito es mapear la entidad «documento de archivo», de acuerdo con la denominación empleada en las NEDA, y sus atributos con los distintos modelos que se analizan.

4 Metodología

En esta fase del proyecto, se han analizado e identificado las entidades requeridas para la estructuración de la información archivística y sus atributos, en un contexto de gestión en papel y electrónica. Se toma como referencia la experiencia de los autores, que trabajan con el software de gestión de archivos Odilo A3W; el análisis de la documentación de archivo de clientes; las especificaciones de los modelos estudiados, que se describen brevemente en el apartado siguiente; las equivalencias definidas por CNEDA (2017, p. 127-133). Después, se analizan e identifican los atributos de la entidad «documento de archivo» en los distintos modelos analizados: e-EMGDE, NEDA y RiC y EAD; esto es, cómo se estructuran, qué fines persiguen y cómo se relacionan unos con otros.

Posteriormente, se mapean con los principales componentes empleados en A3W. En este mapeo, se tiene en cuenta el propósito de cada elemento que se mapea y su grado de relación con el metadato empleado por Odilo.

Finalmente, se definen «vistas» especializadas por cada uno de los modelos, que son utilizados como formatos de salida en la aplicación.

5 Resultados

El estudio se centra en el análisis de modelos conceptuales, esquemas de metadatos y estándares de codificación, sin tener en cuenta los estándares de descripción. Así es, ya que en esta primera fase del proyecto, los estándares de descripción (como ISAD-G) no afectan de manera relevante a la definición de la aplicación y, además, dichos estándares ya habían sido incorporados a A3W. De modo que los modelos estudiados fueron los siguientes:

- 1) e-EMGDE, que «incluye los metadatos mínimos obligatorios, definidos en las normas técnicas de interoperabilidad (NTI) de Documento y Expediente electrónicos, así como otros metadatos complementarios pertinentes en una política de gestión y conservación de documentos electrónicos» (Centro de Transferencia de Tecnología, [2015]).
- 2) NEDA, «un modelo de referencia teórico basado en el esquema de modelado entidad-relación y supone la consolidación de los fundamentos adoptados por la CNEDA en cuanto a descripción archivística en los documentos aprobados en junio de 2012 y mayo de 2017. Este modelo se ha desarrollado con la intención de identificar los requisitos funcionales que deben tenerse en cuenta en el diseño y configuración de los sistemas de descripción archivística» (Ministerio de Cultura y Deporte, 2017). Parte, además, de un análisis muy completo del resto de normas conceptuales y estándares archivísticos, nacionales e internacionales, que también tiene en cuenta las normas de descripción archivística autonómicas (Quílez Mata, 2017, p. 22-23).
- 3) RiC, con unas características similares a NEDA, es «una norma para la descripción de documentos basada en los principios archivísticos» (International

Council on Archives, 2016). Se trata de un modelo conceptual en fase de elaboración que está llamado a asentarse como el estándar internacional de referencia.

- 4) EAD, EAC y EAG. Son estructuras de codificación basadas en XML, diseñadas para el intercambio de registros e información archivística de diferentes tipos: descripciones archivísticas, registros de autoridad y función, etc.

A continuación, nos centramos en el análisis y mapeado de la entidad «documento de archivo».

5.1 La «entidad documento de archivo»

En el contexto del modelo de Odilo A3W hay tres entidades generales y subtipos de entidades. En la Figura 1 se recogen las principales entidades y relaciones utilizadas en la aplicación. La entidad «documento de archivo», en el centro, es una de ellas.

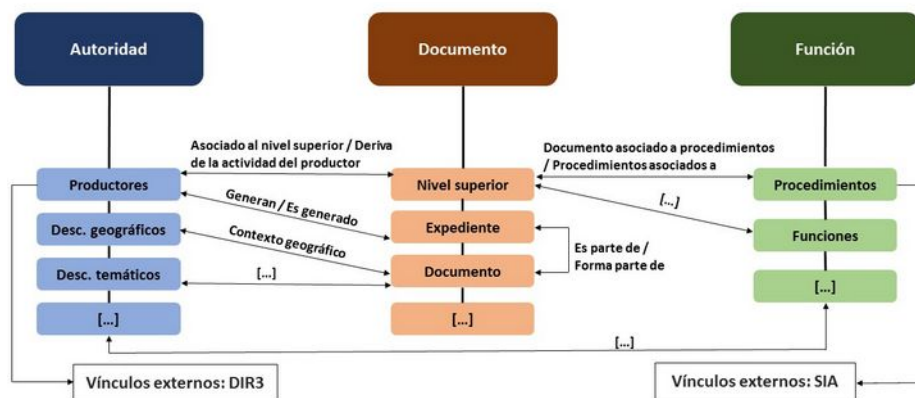


Figura 1. Entidades disponibles en A3W y algunas de las relaciones existentes entre ellas.

Bajo la denominación «documento de archivo» se engloba cualquier nivel del cuadro de clasificación, independientemente de su denominación, las agregaciones, las unidades documentales compuestas y las simples. Esta distinción no genera, sin embargo, entidades diferentes por cada «tipo de documento», como ocurre con RiC. Entre los atributos que pueden asociarse a esta entidad, destacan los asociados a políticas de disposición, políticas de acceso y de preservación digital.

Las entidades identificadas para las especificaciones analizadas son:

1) e-EMGDE – Este vocabulario de metadatos incorpora las siguientes 5 entidades: documento; agente; actividad; regulación; relación. (DTIC, 2016).

2) RiC – El modelo consta de 14 entidades: record (unidad documental compuesta); record component (documento), record set (niveles superiores), agent (agente); occupation (ocupación); position (puesto); function (función del agente); function-abstract (función); activity (actividad); mandate (mandato); documentary

form (tipo documental); date (fecha); place (lugar); concept/thing (conceptos o elementos asociados (ICA, 2016).

3) NEDA – El estándar consta de 6 entidades: documentos de archivo; agente; función; norma; concepto objeto o acontecimiento; lugar. (CNEDA, 2017).

4) EAD – Este estándar XML cuenta con varias entidades, de las cuales, dos pueden considerarse como referencias a documento de archivo.

Las equivalencias encontradas para la entidad «documento de archivo» se recogen en la siguiente tabla.

Tabla 1. Mapeado de la entidad “Documento de archivo”.

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Niveles superiores		E3 Record Set		<archdesc >
Expediente	Document	E1 Record	Documento de archivo	<c>
Pieza		E2 Record Component		<c>

En realidad, pese a las discrepancias, los modelos de entidad-relación de los diferentes estándares son compatibles, estando unos incluidos en otros o representados mediante atributos. Así, el elemento «tipo documental» de RiC sería un atributo del «documento de archivo» de NEDA o de la entidad «documento» en el caso de e-EMGDE. De modo que nuestro objetivo, llegados a este punto, es el de buscar un encaje entre las diferentes entidades en el modelo cuya implantación se trasladará al software A3W. Algunas de las entidades han sido transformadas en atributos aplicables a una o varias de las entidades de A3W.

5.2 El mapeo de odiloa3w con los modelos analizados

En el anexo de esta comunicación, se muestran las entidades de «documento de archivo» y sus atributos, mapeados para la descripción de la documentación de cualquier organización archivística. El mapeo no siempre se define en régimen de equivalencia entre una propiedad y otra, sino que se aprecian relaciones de distintos tipos y grados. Además, algunas de las propiedades son específicas de uno o más subtipos de la entidad «documento de archivo», pero no de todas.

También se han definido diferentes «vistas», de acuerdo con las especificaciones analizadas. Esto facilita la visualización de la información desde varias perspectivas.

6 Conclusiones y propuestas de futuro

Los perfiles de aplicación se presentan como una alternativa que permite aprovechar los elementos constitutivos de diferentes estándares, con el objetivo de facilitar la interoperabilidad entre sistemas y mejorar la descripción de documentos para una

recuperación de información precisa. Para ello, es necesario analizar en profundidad esos estándares, de acuerdo con las necesidades de las organizaciones implicadas y mapear sus distintos componentes. Ahora bien, realizar un mapeo conlleva la armonización de distintas realidades, que no siempre es fácil de llevar a la práctica. En esta comunicación se muestra la primera fase del proyecto de definición del modelo de Odilo A3W y el mapeo con e-EMGDE, NEDA, RiC y EAD y del perfil de aplicación odiloa3w, comenzando por la entidad «documento de archivo» y sus atributos.

Aun existiendo una relación entre las entidades y, sobre todo, entre los atributos analizados, existen evidentes diferencias entre las propiedades empleadas por unas especificaciones y otras. Probablemente, el propósito y orientación de cada modelo esté relacionado. Así, por ejemplo, que la descripción sea para documentación en papel o electrónica influye de manera decisiva. El tipo y grado de relación entre los distintos elementos analizados será objeto de un trabajo posterior.

Si las organizaciones especializadas están utilizando (o esperan utilizar) distintos modelados y vocabularios para sus archivos, las entidades que proporcionan el software para su gestión deben ser parte activa de la evolución técnica y tecnológica que se produce en el área y brindar soluciones que se adapten a las necesidades de sus clientes. El proyecto continuará con el análisis de otras entidades utilizadas en la aplicación y el mapeo de éstas y de sus atributos con las especificaciones indicadas.

7 Referencias

- Dirección de Tecnologías de la Información y las Comunicaciones (DTIC). (2016): *Esquema de Metadatos para la Gestión del Documento Electrónico (e-EMGDE). Versión 2.0. Documentación complementaria a la Norma Técnica de gestión de documentos electrónicos*. Ministerio de Hacienda, 2016. Recuperado de: <https://administracionelectronica.gob.es/ctt/eemgde#.XDzQgs9Kj-Z>
- Centro de Transferencia de Tecnología (CTT) [2015]: *Sistema de Información Administrativa (SIA)*. Recuperado de: <https://administracionelectronica.gob.es/ctt/sia#.XMca-ej7Tcs>
- Comisión de Normas Españolas de Descripción Archivística (CNEDA) (2017). *Neda_MC. Modelo conceptual de descripción archivística. Entidades, Relaciones y Atributos*. Recuperado de: <https://sede.educacion.gob.es/publiventa/d/20886C/19/0>
- International Council on Archives (ICA). (2016). *Records in Context – A Conceptual Model for Archival Description*, 2016. Recuperado de: <https://www.ica.org/es/records-context-modelo-conceptual>
- Quílez Mata, J. *Integració d'arxius, museus i biblioteques amb tecnologies del web semàntic: de la modelització conceptual a la iniciativa Link Open Data*. LLIGAL. Revista catalana d'arxivística. Descripció i Tractament de Dades Obertes. Associació d'Arxivers·Gestors de

Documents de Catalunya, 40, 2017, pp. 16-57 Recuperado de:
<https://www.raco.cat/index.php/liligall/article/download/340577/431359>

Ministerio de Cultura y Deporte. *NEDA-Voc: Vocabulario para la descripción archivística*, 2017. Recuperado de: <http://www.culturaydeporte.gob.es/cultura/areas/archivos/mc/cneda/documentacion/normas/neda-voc.html>

Popovici, B. F. (2016). *RECORDS IN CONTEXTS. Towards a New Level in Archival Description?* Pokrajinski arhiv Maribor. Recuperado de:
http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci_2016/013-031_popovici_2016.pdf

8 Anexo: el mapeo

Tabla 1. Mapeo realizado de odiloa3w con los distintos modelos

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Nivel de descripción	0. Tipo de entidad 1. Categoría	P23 Type (record set)	<DOC_TIPOENT>	<archdesc>@level <c>@level <separatedmaterial>
Identificadores	2. Identificador	-----	-----	-----
Identificador normalizado del expediente o del documento, N° informático	2.1. Secuencia de identificador	P2 Local Identifier	<DOC_ID>	<recordid> <unitid>
Número del expediente	2.1. Secuencia de identificador	P2 Local Identifier	<DOC_IDCOMP>	<otherrecordid>
Esquema del identificador	2.2. Esquema del identificador	-----	-----	-----
Identificador Persistente Único	-----	P1 Global Persistent Identifier	-----	-----
Clasificación	22. Clasificación	-----	-----	-----
Código de clasificación	22.1. Código de clasificación	P17 Classification (record) P27 Classification (record set)	<DOC_CLASIF>	<fileplan>
Título de la clasificación	22.2. Denominación de clase			-----
Tipo de clasificación	22.3. Tipo de clasificación			-----
Nota de clasificación	-----			-----
Información para la descripción	-----	-----	-----	-----
Instrumentos de descripción (niveles)	-----	-----	<DOC_IDD>	<otherfindaid>
Estado del nivel de clasificación (niveles)	-----	-----	<DOC_COMPLET>	-----
Títulos	3. Nombre	-----	-----	-----
Título	3.1. Nombre natural	P3 Name	<DOC_NOMBRE>	<unittitle>
Nombre del fichero adjunto	3.2. Nombre del fichero	-----	-----	-----
Fechas	4. Fechas	-----	-----	-----
Fecha de apertura del expediente	4.1. Fecha inicio	E12 Date	<DOC_FECHA>	<unitdate>
Fecha de creación del documento	4.1. Fecha inicio	E12 Date	<DOC_FECHA>	<unitdate>

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Fecha de alta del documento en el sistema	4.1. Fecha inicio	E12 Date	<DOC_FECHA>	<unitdate>
Fecha de cierre del expediente	4.2 Fecha fin	E12 Date	<DOC_FECHA>	<unitdate>
Fecha de ingreso en la institución	4.1. Fecha inicio	E12 Date	<DOC_FECHA>	<unitdate>
Fecha de firma de exp. o documento electrónico	4.2. Fecha fin	E12 Date	<DOC_FECHA>	<unitdate>
Fecha de transferencia	-----	-----	<DOC_FECHA> <DOC_INGRESO>	<unitdate>
Fecha de expurgo	-----	-----	<DOC_FECHA>	<unitdate>
Datos de descripción	5. Descripción	-----		
Descripción o resumen del expediente	5. Descripción	P9 Scope and Content (record) P31 Scope and Content (record set)	<DOC_SINTESIS>	<abstract> <scopecontent>
Datos de seguridad y acceso	8. Seguridad + 9. Derecho de acceso, uso y reutilización	-----	-----	-----
Nivel de acceso	8.1. Nivel de acceso	P18 Conditions of Access (record)	<DOC_ACCESO>	<accessrestrict>
Texto informativo sobre seguridad	8.2. Advertencia de seguridad			
Permisos de visualización, modificación y/o eliminación de la información	8.3. Permisos			
Nivel de protección de los datos LOPD	8.4. Sensibilidad datos de carácter personal			
Tipo acceso a doc. por grado de confidencialidad	8.6. Nivel de confidencialidad de la información			
Tipo acceso a doc. según tipo de información	9.1. Tipo de acceso			
Código de limitación del acceso a la información	9.2. Código de la causa de limitación			<legalstatus>
Causa de limitación de acceso a la información	9.3. Causa legal/normativa de limitación			
Condiciones de uso y reproducción	9.4. Condiciones de reutilización	P19 Conditions of Use (record)	<DOC_REPRO>	<userrestrict>
Nota de difusión			<DOC_DIFUS>	<bibliography>
Origen de la documentación (ingreso)	-----	-----	-----	-----
Forma de ingreso	-----	-----	<DOC_INGRESO>	<acqinfo>
Previsión de ingreso	-----	P25 Accrual Status (record set)	<DOC_PREVIS>	<accruals>
Procedencia de la documentación	-----	P24 Accrual Note (record set)	-----	<acqinfo>
Idioma	11. Idioma	P11 Language (record)	<DOC LENGUA> <DOC_ESCRITURA>	<langmaterial>
Transferencias y expurgos (niveles)	13. Calificación	-----	-----	-----
Valoración del documento	13.1. Valoración	-----	-----	-----

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Valor primario	13.1.1. Valor primario	-----	<DOC_VALORA>	<appraisal>
Tipo de valor	13.1.1.1. Tipo de valor	-----		
Plazo estimado	13.1.1.2. Plazo	-----		
Valor secundario	13.1.2. Valor secundario	-----		
Dictamen	13.2. Dictamen	-----	-----	-----
Acción de disposición que ejecutar	13.2.1. Tipo de dictamen	-----	<DOC_VALORA>	<appraisal>
Descripción de la política de disposición	13.2.2. Acción dictaminada	-----		
Plazo de disposición	13.2.3. Plazo de ejecución de la acción	-----		
Datos de transferencia	13.3. Transferencia	-----		
Archivo de origen	13.3.1. Fase de archivo	-----	<DOC_INGRESO>	<acqinfo>
Archivo de destino	13.3.1. Fase de archivo	-----		
Plazo de transferencia	13.3.2. Plazo de transferencia	-----		
Documento esencial	13.4. Documentación esencial	-----		
Datos de expurgo	-----	-----	-----	-----
Plazo de expurgo	-----	-----	<DOC_VALORA>	<appraisal>
Motivo de expurgo o eliminación	-----	-----		
Muestreo	-----	-----		
Datos de conservación y preservación	-----	-----	-----	-----
Estado de conservación del documento (pieza)	-----	P8 Quality of Information (record)	-----	<phystech>
Notas de conservación	-----	P16 Physical Characteristics Note (record)	<DOC_CONSERV>	-----
Medidas de preservación (nivel)	-----	-----	<DOC_PRESERV>	-----
Características del documento adjunto	14. Características técnicas 15.1. Ubicación	-----	-----	<physdescstructured>
Tipo de contenido	-----	P6 Content Type (record)	<DOC_TIPOCONT>	<genreform>
Formato del fichero adjunto	14.1. Formato	P10 Encoding Format (record)	-----	<physfacet>
Versión del formato	14.2. Versión del formato	P10 Encoding Format (record)	-----	<physfacet>
Resolución de la imagen	14.3. Resolución	-----	-----	<physfacet>
Técnica de producción	-----	P13 Production Technique (record)	-----	-----
Medio de representación del fichero	-----	P12 Media Type (record)	<DOC_TIPOMED>	<phystech>
Tamaño	14.4. Tamaño + 15.1	-----	-----	-----

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Dimensiones físicas	14.4.1. Dimensiones + 14.4.4. Unidades	P7 Content Extent (record)	<DOC_EXTEN>	<dimensions>
Tamaño del fichero adjunto	14.4.2. Tamaño lógico + 14.4.4. Unidades	P15 Physical or Logical Extent (record)		<quantity>
Número de elementos	14.4.3. Cantidad	P29 Content Extent (record set) P30 Physical or Logical Extent (record set)		
Profundidad del color	14.5. Profundidad de color	-----	-----	<physfacet>
Soporte físico	15.1. Soporte	P14 Medium (record)	<DOC_TIPOSOPORTE>	<unittype>
Localización	15. Ubicación	-----	-----	-----
Responsable de la documentación	-----	-----	-----	<repository>
Localización o signatura	15.2. Localización	-----	<DOC_LOCALIZ>	<physloc>
Signatura anterior (tras transferencia)	-----	-----	-----	-----
Integridad del expediente / documento electrónico	16. Verificación de integridad	-----	-----	-----
Algoritmo	16.1. Algoritmo	P5 Authenticity and Integrity Note (record)	-----	-----
Valor huella	16.2. Valor	P22 Authenticity and Integrity Note (record set)	-----	-----
Datos de firma (electrónica)	17. Firma	-----	-----	-----
Tipo de firma	17.1.1. Formato de firma	-----	-----	-----
Rol del firmante	17.2. Rol de firma	-----	-----	-----
Firma CSV	17.3. Valor del CSV	-----	-----	-----
Regulación Generación CSV	17.4. Definición generación CSV	-----	-----	-----
Datos del firmante	17.5. Firmante	-----	-----	-----
Nombre	17.5.1. Nombre y apellidos o razón social	-----	-----	-----
Identificador del firmante	17.5.2. Número de identificación de los firmantes	-----	-----	-----
Calidad del firmante	17.5.3. En calidad de	-----	-----	-----
Nivel de firma	17.5.4. Nivel de firma	-----	-----	-----
Otra información del firmante	17.5.5. Información adicional	-----	-----	-----
Tipo de documento	18. Tipo documental	E11 Documentary Form	<DOC_TIPODOC>	<descriptivenote>
Prioridad del documento (pieza)	19. Prioridad	-----	<DOC_CARACESP>	-----
Estado de elaboración del documento (pieza)	20. Estado de elaboración	P21 Record state (record)	<DOC_TRAD>	-----
Versión NTI	-----	-----	-----	-----
Versión NTI del expediente	23. Versión NTI	-----	-----	-----
Versión NTI del documento (pieza)	23. Versión NTI	-----	-----	-----

odiloa3w	e-EMGDE	RiC	NEDA	EAD
Órgano DIR3	-----	-----	-----	-----
Código DIR3	24. Órgano	P32 Type	-----	<origination>
Descripción del órgano DIR3			-----	
Origen del documento (pieza)	25. Origen del documento	-----	-----	
Identificación del documento de origen (pieza)	26. Identificación del documento de origen	P2 Local Identifier	-----	<originalsloc>
Localización de originales	-----	-----	-----	<originalsloc>
Localización de copias	-----	-----	-----	<altformavail>
Estado del expediente (expediente)	27. Estado del expediente	-----	-----	-----
Asiento registral	29. Asiento registral	-----	-----	-----
Tipo de asiento registral	29.1 Tipo de asiento registral	-----	-----	-----
Código de la oficina de registro	29.2. Código de la oficina de registro	-----	-----	-----
Fecha de asiento registral	29.3. Fecha de asiento registral	E12 Date	<DOC_FECHA>	<unitdate>
Número de asiento registral	29.4. Número de asiento registral	P2 Local Identifier	-----	<otherrecordid>
Notas adicionales	-----	-----	-----	-----
Nota de historia archivística	-----	P20 History (record) P28 History (record set)	<DOC_HISTORIAL> o <DOC_HISTARCHIV>	<custodhist>
Nota de ordenación (niveles)	-----	P26 Arrangement (record set)	<DOC_ORGANIZA>	<arrangement>
Otras notas	-----	P4 General Note	<DOC_NOTA>	<didnote> <descriptivenote> <odd>

Instrumentos De Classificação Em Arquivos: Uma Análise Comparada Na Administração Pública Brasileira

Paola Rodrigues Bittencourt

ORCID: 0000-0002-5946-8121. Universidade de Coimbra (Coimbra, Portugal).
Arquivo Nacional (Rio de Janeiro, Brasil)
paolarb@gmail.com

Resumo

O presente trabalho compara dois dos instrumentos de classificação de documentos de arquivo. Um para a administração pública estadual de São Paulo e outro na administração pública estadual do Rio de Janeiro. Trata-se de uma pesquisa que se caracteriza como descritiva e utiliza o método comparativo para análise do objeto. Os instrumentos foram comparados em duas etapas. A primeira etapa a partir dos textos de apresentação e da introdução dos documentos, considerando as informações de contexto e elaboração. A segunda etapa considerando as unidades de classificação propriamente ditas, desde os níveis mais gerais até os mais específicos de classificação. Com a análise foi possível perceber que há muito mais semelhanças do que diferenças entre os instrumentos apresentados. Ambos os instrumentos tendem a ter sua aplicação facilitada pelo uso das tipologias documentais constantes dos dois instrumentos. A principal diferença está na forma de elaboração do nível mais específico para o mais geral apresentado no instrumento de classificação do Estado do Rio de Janeiro. Tal diferença pode significar, na prática de uso dos instrumentos, uma representação mais fidedigna da realidade representada uma vez que a elaboração dos instrumentos começa na identificação e agrupamento das tipologias documentais oriundas de cada ação administrativa.

Palavras-chave: classificação de documentos; instrumento de classificação; arquivo público

Abstract

The present work compares two of the instruments of classification of archival records. One for the state public administration of São Paulo and another for the state public administration of Rio de Janeiro. It is a research that is characterized as descriptive and uses the comparative method for object analysis. The instruments were compared in two steps. The first step from the presentation texts and the introduction of the documents, considering the context and elaboration information. The second step considering the classification units themselves, from the most general to the more specific grading levels. With the analysis it was possible to perceive that there are much more similarities than differences between the presented instruments. Both instruments tend to have their application facilitated by the use of the records typologies contained in the two instruments. The main difference is in the form of elaboration of the most specific level for the most general presented in the classification instrument of the State of Rio de Janeiro. Such difference may mean, in the practice of using the instruments, a more reliable representation of

the reality represented since the elaboration of the instruments begins in the identification and grouping of the documentary typologies originating from each administrative action.

Keywords: records classification; classification scheme; public archive.

1. Introdução

Classificar é uma ação inerente à natureza humana e muitas vezes ocorre sem que sequer seja percebida. Essa ação surge a partir da necessidade do homem em compreender as coisas e os fenômenos do mundo no qual estão inseridos (Pombo, 1988)

Muito há o que se explorar sobre o potencial uso da classificação nos mais variados campos do saber. Entretanto, a este trabalho interessa debruçar-se sobre a ótica da classificação nos arquivos, local cujo acúmulo de informações demonstra a quantidade e a complexidade das ações humanas no exercício de suas atividades.

Sabe-se que, nos arquivos, a classificação tem papel fundamental para todos os processos de gerenciamento e processamento de documentos. Todas as demais ações empreendidas nos arquivos dependem da classificação dos documentos, pois esta é a função que permite preservar e compreender o vínculo existente entre os documentos de um arquivo. Porém, a produção científica voltada às questões teórico-metodológicas da classificação em arquivos muito pouco é explorada pelos profissionais da área (Sousa, 2005)

Como parte integrante da gestão de documentos, que deve controlar todo o ciclo de vida dos documentos, a classificação é a função arquivística responsável por preservar a relação existente entre os documentos de um arquivo, atributo reconhecido como organicidade ou vínculo arquivístico. Esta função tem como referenciais dois princípios arquivísticos fundamentais: a proveniência e a ordem original (Rodrigues, 2008; Sousa, 2005).

A função de classificar, para além de outras possíveis acepções, em arquivos significa distribuir os componentes de um conjunto em estruturas hierárquicas cujas relações são guardadas entre si. A classificação parte das semelhanças e diferenças entre os elementos do conjunto (Pombo, 1988; Simões & Freitas, 2013).

Pela classificação, o arquivo – o todo – é decomposto em uma estrutura hierárquica lógica e pré-definida. Em que pese as aplicações e práticas diferenciadas de gestão de documentos em diferentes contextos administrativos, a classificação representa uma função primordial para execução das demais ações empreendidas nos arquivos. (Indolfo, 2013; Rodrigues, 2008; Sousa, 2005). A avaliação, e a consequente definição dos documentos que irão integrar o patrimônio arquivístico nacional dependem essencialmente da construção de um instrumento de classificação consistente para os arquivos.

Para elaboração de um instrumento de classificação há um conjunto de atividades que pressupõe a pesquisa e coleta de dados sobre a instituição, suas funções e atividades e os documentos produzidos em decorrência dessas atividades (Sousa,

2005, 2012). Esse conjunto de atividades que antecedem a elaboração do plano de classificação é reconhecido por alguns autores como uma nova função arquivísticas: a identificação (Rodrigues, 2002, 2008).

Sousa (2005) analisa as bases teórico-práticas fundamentais à classificação até concluir com as bases da classificação na arquivística, onde destaca que

É claro que há diferenças entre os propósitos da classificação das ciências e da classificação de documentos arquivísticos, por exemplo. Enquanto as primeiras têm um caráter geralmente especulativo, a segunda tem objetivos pragmáticos, pois o valor dela está ancorado na sua **utilidade prática, rapidez, exatidão e facilidade de uso** (grifo nosso)

Uma vez que somente em um cenário mais recente a classificação de documentos em arquivos tem sido analisada com maior aprofundamento teórico e metodológico, pretende-se com a análise compreender as vantagens e desvantagens sobre a adoção de uma ou outra opção metodológica na elaboração de instrumentos de classificação.

A análise comparativa entre dois instrumentos de classificação para documentos de arquivo relativos às atividades-meio adotados pela administração pública no Brasil é parte de uma pesquisa desenvolvida no âmbito do doutoramento e contempla dois instrumentos estaduais brasileiros: 1. Plano de classificação de documentos da Administração Pública do Estado de São Paulo: atividades-meio (São Paulo, 2005) e 2. Plano de Classificação de documentos das atividades-meio do Estado do Rio de Janeiro (Rio de Janeiro, 2012).

O Plano de Classificação de Documentos da administração pública do Estado de São Paulo relativo às atividades-meio foi elaborado pelo Arquivo Público do Estado de São Paulo em parceria com as Comissões de Avaliação de Documentos de Arquivo das instituições integrantes do Sistema de Arquivos do Estado de São Paulo. Teve sua primeira versão aprovada em 2005 e é a mesma versão que continua em vigor desde então. O instrumento possui como embasamento legal o Decreto nº 48.897, de 27 de agosto de 2004 e o Decreto nº 48.898, de 27 de agosto de 2004 e devem ser aplicados por todas as organizações governamentais da administração pública estadual de São Paulo.

O Plano de Classificação de Documentos das Atividades Meio do Estado do Rio de Janeiro foi elaborado no âmbito do Programa de Gestão de Documentos do Estado do Rio de Janeiro (PGD). Foi elaborado pelo Departamento de Gestão de Documentos do Arquivo Público do Estado do Rio de Janeiro (APERJ), com o apoio das Secretarias de Estado da Casa Civil, Planejamento e Gestão e da Fazenda e foram aprovados e pelo Decreto Estadual nº 43.992, de 14 de dezembro de 2012.

A análise comparada foi realizada em três etapas e teve como referência elementos de elaboração, a estrutura e a aplicação dos três diferentes instrumentos.

Na primeira etapa foram coletados e relacionados dados constantes da apresentação e introdução dos três instrumentos, para compreender o contexto e a forma de elaboração e estruturação dos documentos. Na segunda etapa foram coletadas e elencadas informações relativas aos níveis mais gerais de classificação em

cada instrumento. Na terceira etapa, para se chegar aos níveis mais específicos de classificação, foi selecionada a atividade de gestão de pessoal para comparar as subdivisões dos níveis de classificação. A subdivisão relativa às ações de gestão de pessoas foi selecionada por representar grande parte da produção de documentos na atividade-meio dos órgãos da administração pública brasileira.

2. Análise comparada

Nesta seção serão comparados os elementos indicados na apresentação, introdução e metodologia.

No primeiro quadro (quadro 1) o objetivo foi sistematizar informações constantes nos textos de apresentação e introdução, relativas à existência de embasamento legal, política arquivística, programa de gestão de documentos, responsáveis pela elaboração do instrumento e nível de participação das organizações governamentais.

Quadro 1: Apresentação e introdução dos instrumentos de classificação

	Plano de Classificação de Documentos da administração pública do Estado de São Paulo relativo às atividades-meio	Plano de Classificação de Documentos das Atividades Meio do Estado do Rio de Janeiro
Aprovação legal	Decreto nº 48.898, de 27 de agosto de 2004	Decreto Estadual nº 43.992, de 14 de dezembro de 2012
Política arquivística	Não possui	Não possui
Programa de gestão de documentos	Não possui	Possui
Manual de Gestão de Documentos	Não possui	Possui
Histórico institucional e de elaboração do instrumento	Possui	Consta do Programa de Gestão de Documentos
Instituição responsável pela elaboração	Arquivo Público do Estado de São Paulo	Arquivo Público do Estado do Rio de Janeiro
Participação das organizações governamentais	Sim, por meio da Comissão de Avaliação Documentos de Arquivo em cada instituição	Sim, por meio da Comissão de Gestão de Documentos em cada instituição
Ano de aprovação	2004	2012

Fonte: elaboração própria utilizando como referência os instrumentos de classificação da administração pública estadual de São Paulo e da administração pública estadual do Rio de Janeiro.

O segundo quadro (quadro 2) organiza as informações constantes dos instrumentos relativas às opções metodológicas adotadas para elaboração de cada documento.

Quadro 2: Ferramentas e aspectos metodológicos de elaboração dos instrumentos de classificação

	Plano de Classificação de Documentos da administração pública do Estado de São Paulo relativo às atividades-meio	Plano de Classificação de Documentos das Atividades Meio do Estado do Rio de Janeiro
Funcional ou estrutural	Funcional	Funcional
Levantamento da produção documental	Realizada	Realizada
Codificação	Duplex	Duplex
Níveis de classificação	4 níveis	4 níveis
Instruções de aplicação	Possui um manual de aplicação publicado.	Possui instruções no manual de gestão de documentos

Fonte: elaboração própria utilizando como referência os instrumentos de classificação da administração pública estadual de São Paulo e da administração pública estadual do Rio de Janeiro.

Os próximos quadros têm como objetivo sistematizar um comparativo entre os instrumentos a partir dos níveis de classificação utilizados em cada instrumento. Para isso, o primeiro quadro (quadro 3) compara os níveis gerais: competência, funções e subfunções. Enquanto o quadro seguinte (quadro 4) compara uma das ações governamentais: gestão de pessoas (ou gestão de recursos humanos).

Quadro 3: Níveis gerais dos instrumentos de classificação

	Plano de Classificação de Documentos da administração pública do Estado de São Paulo relativo às atividades-meio	Plano de Classificação de Documentos das Atividades Meio do Estado do Rio de Janeiro
Competência	Não utiliza	Planejamento, orçamento e coordenação da ação governamental
Funções	<ul style="list-style-type: none"> - Organização Administrativa - Comunicação institucional - Gestão de Recursos Humanos - Gestão de bens materiais e patrimoniais - Gestão orçamentária e financeira - Gestão de documentos e informações - Gestão de atividades complementares 	<ul style="list-style-type: none"> - Acompanhamento da gestão orçamentária, financeira, patrimonial e industrial - Gestão dos recursos logísticos - Gestão das atividades de serviços gerais - Gestão das políticas e diretrizes orçamentárias estaduais - Gestão de planejamento das ações governamentais - Gestão de políticas de administração de recursos humanos - Gestão do patrimônio imobiliário

Fonte: elaboração própria utilizando como referência os instrumentos de classificação da administração pública estadual de São Paulo e da administração pública estadual do Rio de Janeiro.

O quadro a seguir (quadro 4) organiza as informações de que tratam as subdivisões relativas às ações de gestão de pessoas e recursos humanos.

Quadro 4: Níveis específicos de classificação - as subclasses relativas aos recursos humanos

	Plano de Classificação de Documentos da administração pública do Estado de São Paulo relativo às atividades-meio	Plano de Classificação de Documentos das Atividades Meio do Estado do Rio de Janeiro
	-----	Planejamento, orçamento e coordenação da ação governamental (Competência/Classe)
	Gestão de Recursos Humanos (Função/Classe)	Gestão de políticas de administração de recursos humanos (Função/Subclasse)
Subdivisões relativas às ações de gestão de pessoas	Subclasses: <ul style="list-style-type: none"> - Planejamento e formulação de políticas de recursos humanos - Seleção e desenvolvimento de recursos humanos - Expediente de pessoal - Controle de frequência - Pagamento de pessoal - Segurança e medicina do trabalho - Promoção do desenvolvimento social de funcionários e servidores - Coordenação de atividades de convivência infantil 	Grupos: <ul style="list-style-type: none"> - Acompanhar utilização do sistema de gestão de recursos humanos - Administrar processamento das folhas de pagamento - Administrar sistemas informatizados de pessoal - Analisar direitos e vantagens de servidores - Analisar irregularidades no serviço público - Analisar solicitações e propor criação e reestruturação de carreiras e remuneração - Atender decisões judiciais relativas a pessoal - Consolidar legislação de pessoal das administrações direta e indireta - Controlar atendimento ao servidor - Controlar cadastro de pessoal e frequência de servidores - Controlar o provimento de cargos, designação de servidores e apostilamento de atos oficiais - Controlar regras de parametrização do sistema de gestão de recursos humanos - Credenciar e recredenciar entidades consignatárias - Desenvolver estudos para elaboração de programas de capacitação

		<ul style="list-style-type: none"> - Gerenciar realização de concurso público - Implementar e controlar a capacitação e reciclagem dos usuários do sistema de gestão de recursos humanos - Propor diretrizes relativas a estágio probatório e avaliação de desempenho individual - Providenciar pagamento de pessoal - Responder consulta de aplicação da legislação e normas relativa à pessoal
--	--	---

Fonte: elaboração própria utilizando como referência os instrumentos de classificação da administração pública estadual de São Paulo e da administração pública estadual do Rio de Janeiro.

3. Os instrumentos de classificação: o que apontam os dados

O Arquivo Público do Estado de São Paulo possui um documento publicado denominado “Política Pública de Arquivos e Gestão de Documentos do Estado de São Paulo”. Entretanto, este documento compreende uma reunião da legislação arquivística federal e estadual que fundamenta as atividades no setor. Não foi encontrada nenhuma publicação com as diretrizes específicas de uma política arquivística ou de gestão de documentos para o Estado de São Paulo.

Sobre programas de gestão de documentos apenas no Estado Rio de Janeiro foi identificado que a elaboração e aplicação do instrumento de classificação é parte de um programa de gestão de documentos. O plano de classificação do Estado de São Paulo, apesar de citar uma política e um programa de gestão de documentos, não disponibiliza nenhum programa de gestão de documentos em seu sítio eletrônico.

Ambos os instrumentos, do Arquivo Público do Estado de São Paulo e o do Estado do Rio de Janeiro, contaram com a colaboração de organizações abrangidas pelo instrumento.

Na estruturação dos instrumentos percebeu-se que as competências, funções, subfunções e atividades são identificados e categorizados de formas distintas em cada instrumento. No instrumento elaborado pelo Arquivo do Estado de São Paulo foram identificadas sete funções, sendo: Organização Administrativa; Comunicação institucional; Gestão de Recursos Humanos; Gestão de bens materiais e patrimoniais; Gestão orçamentária e financeira; Gestão de documentos e informações; e Gestão de atividades complementares. No plano de classificação do Estado do Rio de Janeiro também foram identificadas sete funções: Acompanhamento da gestão orçamentária, financeira, patrimonial e industrial; Gestão dos recursos logísticos; Gestão das atividades de serviços gerais; Gestão das políticas e diretrizes orçamentárias estaduais; Gestão de planejamento das ações governamentais; Gestão de políticas de administração de recursos humanos; e Gestão do patrimônio imobiliário.

No caso dos níveis hierárquicos o instrumento do Arquivo do Estado de São Paulo possui maior representação de níveis hierárquicos entre a função e o nível mais específico – que é sempre o a tipologia documental. Embora no instrumento do Arquivo Público do Estado do Rio de Janeiro, as tipologias documentais também compreendam unidades de classificação específicas.

Na perspectiva da elaboração as duas principais questões distanciam os instrumentos são a forma de elaboração e o nível de especificidade dos instrumentos.



Figura 2: Elaboração do instrumento de classificação da administração pública do Estado de São Paulo

O instrumento elaborado pelo Estado do Rio de Janeiro partiu da identificação das tipologias documentais (nível mais específico) para identificação das funções (nível mais geral):



Figura 3: Elaboração do instrumento de classificação para a administração pública do Estado do Rio de Janeiro

4. Discussão e considerações finais

Percebe-se que os instrumentos possuem muitas semelhanças e poucas diferenças. Entretanto, essas divergências pesam principalmente nos aspectos de compreensão e da aplicação dos instrumentos.

Os instrumentos analisados guardam algumas similaridades na identificação das funções e atividades, variando, no entanto, a forma de perceber o que é competência,

o que é função e o que é subfunção, assim como a forma de alocar entre os níveis. A ausência de níveis intermediários que explicitem as ações pode influenciar na compreensão do contexto de produção dos documentos e deixar lacunas sobre as ações específicas que produzem cada conjunto de documentos.

A forma de aplicar os instrumentos é, sem dúvidas, facilitada nos instrumentos em que as tipologias documentais estão explicitadas. Entretanto, é preciso reconhecer que para relacionar as tipologias documentais no instrumento é preciso ter uma administração uniforme e sem perspectivas de alterações na produção de documentos. Na perspectiva da aplicação do instrumento de classificação, constar as tipologias documentais apresenta grande vantagem à utilização, permitindo uma classificação mais rápida, segura e precisa.

Foram identificados pontos como a utilização de princípios distintos para definição de unidades de classificação distintas em um mesmo nível de classificação na análise preliminar realizada no plano de classificação do Estado de São Paulo, onde também são usados termos que não indicam a atividade para designar uma unidade de classificação no nível de atividade. No plano de classificação do Estado do Rio de Janeiro em todos os níveis verificados não foram identificadas divergências.

Mais do que perceber as diferenças e semelhanças em instrumentos de classificação de documentos em arquivos, a análise realizada neste trabalho permite perceber como os conjuntos documentais irão compor, no futuro, o patrimônio arquivístico custodiado pelos arquivos públicos brasileiros.

Nas próximas etapas da pesquisa pretende-se, ainda, aprofundar na análise de alguns elementos destes instrumentos e verificar a partir destas estruturas quais documentos irão compor o patrimônio arquivístico das instituições arquivísticas públicas do Brasil.

Referências

- Indolfo, A. C. (2013). *Dimensões político-arquivísticas da avaliação de documentos na administração pública federal (2004-2012)*. Instituto Brasileiro de Informação, Ciência e Tecnologia. Universidade Federal do Rio de Janeiro.
- Pombo, O. (1988). Da classificação dos seres à classificação dos saberes. *Revista da Biblioteca Nacional de Lisboa*, 2, 19–33.
- Rio de Janeiro, A. P. do E. (2012). Plano de Classificação de documentos das atividades-meio do Estado do Rio de Janeiro. Arquivo Público do Estado do Rio de Janeiro (Brasil). Recuperado de <http://www.rj.gov.br/web/casacivil/exibeConteudo?article-id=2998121>
- Rodrigues, A. C. (2002). *Tipologia documental como fundamento para a gestão de documentos em arquivo: um manual para o município de Campo Belo (MG)* (Mestrado). Universidade de São Paulo.
- Rodrigues, A. C. (2008). *Diplomática contemporânea como fundamento metodológico para identificação da tipologia documental em arquivos* (Doutoramento). Universidade de São Paulo, São Paulo (Brasil).

- São Paulo, A. P. do E. (2005). Plano de classificação e tabela de temporalidade de documentos da Administração Pública do Estado de São Paulo: atividades-meio. Arquivo Público do Estado de São Paulo. Recuperado de <http://www.arquivoestado.sp.gov.br/site/publicacoes/tecnica/ver/plano-de-classificacao-e-tabela-de-temporalidade-da-administracao-publica-do-estado-de-sao-paulo-atividades-meio>
- Simões, M. da G., & Freitas, M. C. V. de. (2013). A classificação em arquivos e bibliotecas à luz da teoria da classificação. *Revista Ponto de Acesso*, 7, n. 1.
- Sousa, R. T. B. de. (2005). *Classificação em arquivística: trajetória e apropriação de um conceito* (Doutoramento). Universidade de São Paulo, São Paulo (Brasil).
- Sousa, R. T. B. de. (2012). A classificação como função matricial do que-fazer arquivístico. In *Arquivística: temas contemporâneos* (3^o ed). Brasília: Senac Distrito Federal.

Memória socioeconômica do ecossistema de startups de Florianópolis

Priscila Machado Borges Sena¹, Ursula Blattmann², José Antonio Moreiro González³ y Jéssica Bedin⁴

¹ ORCID [0000-0002-5612-4315](https://orcid.org/0000-0002-5612-4315). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil.
priscilasena.ufsc@gmail.com

² ORCID [0000-0002-8834-0987](https://orcid.org/0000-0002-8834-0987). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil.
ursula.blattmann@ufsc.br

³ ORCID [0000-0002-8827-158X](https://orcid.org/0000-0002-8827-158X). Universidad Carlos III de Madrid, Getafe, Madrid, Espanha.
jamore@bib.uc3m.es

⁴ ORCID [0000-0003-2324-4246](https://orcid.org/0000-0003-2324-4246). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina. Universidade Comunitária da Região de Chapecó – Unochapecó
jessicabedin@unochapeco.edu.br

Resumen. El objetivo buscado consiste en recomendar las acciones necesarias para la organización y representación del conocimiento en el ecosistema de startups de Florianópolis, con miras a construir su memoria socioeconómica. Se obtuvieron las siguientes acciones: 1) Entrevistar mediante un guión semiestructurado a emprendedores de startups, con el fin de identificar cuál es el origen del conocimiento que aplican en sus empresas y sus fuentes de información en el ecosistema en el que actúan; 2) relacionar las fuentes de información y los tipos de información pertinentes a cada etapa del proceso innovador del ecosistema de startups de Florianópolis; 3) elaborar un modelo de relación entre las fuentes y los tipos de información con ese proceso innovador. Resulta que las posibilidades enumeradas no son únicas y que es detallar las especificidades de sus desarrollos e, incluso, indicar otros desarrollos subsecuentes. Lo que se refleja en la aplicación teórica y práctica de la organización del conocimiento, con la búsqueda de lagunas, soluciones propuestas en estudios precedentes y puesta en uso por profesionales de la información.

Palabras clave: Ecosistema de startups; Memoria socioeconómica; Organización del conocimiento; Representación del conocimiento; SOC.

Abstract. This research aimed to indicate the actions necessary for the organization and representation of knowledge in the Florianópolis' startups ecosystem, with a view to building their socioeconomic memory. As a result,

the following actions were obtained: 1) Interviewing startups entrepreneurs through a semi-structured script, with a view to obtaining the origin of the knowledge they apply in their companies and in the ecosystem, that is, their information sources; 2) relate the information sources and types of information pertinent to each stage of the innovative process of the Florianópolis' startups ecosystem; 3) to elaborate a relationship model between the sources of information and the types of information with the innovative process of the Florianópolis' startups ecosystem. It is seen that the possibilities listed are not unique, and from their developments it will become possible to detail their specificities and indicate other developments. This is reflected in the theoretical and practical extension of the knowledge organization, with the search for gaps, proposed solutions based on studies already carried out, and application through information professionals.

Keywords: Startups ecosystem; Socioeconomic memory; Knowledge organization; Knowledge representation; SOC.

1 Introdução

Um ambiente propício a interação mútua de distintos atores, com fins de desenvolvimento de startups, foi como o ecossistema de startups foi definido por Sena, Blattmann e Teixeira (2017). Deste modo, é possível inferir que nesses aglomerados circula inúmeros e distintos conhecimentos, advindos das mais diversas formações de seus envolvidos. Por envolver um processo inovador dinâmico a entrada e saída de conhecimento também o é.

Neste contexto, a retenção e preservação do conhecimento tornam-se atividades complexas que poderiam ser facilitadas por meio da organização e representação, com a finalidade de construir a memória socioeconômica do ecossistema de startups de Florianópolis. Uma vez que, geralmente o desenvolvimento socioeconômico é considerado como um importante meio de amenizar as desigualdades de gênero e socioeconômicas no acesso aos serviços de saúde, no uso de serviços e, em última análise, nos resultados (Caldwell, 1986, Evans et al., 1994).

Posto isso, objetiva-se neste trabalho indicar as ações necessárias para a organização e representação do conhecimento no ecossistema de startups de Florianópolis, com vistas a construção de sua memória socioeconômica. Para tal, realiza-se uma pesquisa exploratório descritiva e posterior indicação das ações.

2 Por que organizar e representar o conhecimento em um ecossistema de startups?

Entende-se que a organização do conhecimento é "[...] o processo de modelagem do conhecimento que visa a construção de representações do conhecimento" (Brascher &

Café, 2008, p. 8). E a representação do conhecimento como um modelo de abstração do mundo real, construído para determinada finalidade.

Pode-se considerar que a organização da informação é uma representação da informação, que tem como objeto os documentos, estes podem ser representados por partes ou no todo, de forma individual ou em conjunto, abrangendo tanto os aspectos descritivos como os temáticos (Mucalan, 2015).

De acordo, com Brascher e Café (2008) ao realizar o processo de organização do conhecimento, implica também em representar, estabelecer estruturas conceituais que tem como objetivo representar os domínios de conhecimento, bem como a compreensão de fenômenos do mundo.

Já os Sistemas de Organização do Conhecimento (SOCs) ou Knowledge Organization Systems (KOS) podem ser compreendidos como sistemas conceituais, ou tipos de representações do conhecimento, frutos do processo de organização do conhecimento. (Brascher & Carlan, 2010).

Como exemplos desses sistemas, pode-se citar os tesouros e ontologias que podem ser considerados como linguagens comunicativas, que têm como finalidade representar o conhecimento de um certo domínio, oferecendo controle terminológico, minimizando a ambiguidade e melhorando a comunicação especializada (Soares & Maculan, 2016).

A polissemia e a ambiguidade são características intrínsecas da linguagem natural, por isso que o uso de SOC's é uma alternativa viável para dar suporte a interpretação adequada das entidades textuais dos documentos e nos termos de busca utilizados pelos usuários (Soares & Maculan, 2016).

Desta maneira, acredita-se que a organização e representação do conhecimento possam contribuir para os posteriores processos de gestão, e consequentemente para a construção de uma memória que mantenha a dinâmica mais eficaz e eficiente do desenvolvimento do processo inovador existente no ecossistema de startups de Florianópolis. Bem como, disponibilize informações úteis para a sociedade em geral.

3 Ações necessárias para organização e representação do conhecimento no ecossistema de startups de Florianópolis

Neste trabalho entende-se a memória associada ao desenvolvimento socioeconômico como coletiva. A qual, segundo Halbwachs (2013), trata-se do fato de que a recordação e localização das lembranças não pode ser efetivamente analisado se não for levado em consideração os contextos sociais que atuam como alicerce para o trabalho de reconstrução da memória.

A partir do registro do conhecimento produzido pela organização se torna possível a construção da memória organizacional, definida por Spiller e Pontes (2007, p. 99) como a soma das experiências que uma organização acumula ao

longo do tempo, seja nas pessoas, na cultura, nos processos e em seus registros documentados.

Para Menezes (2006, p. 31) a memória organizacional refere-se ao acúmulo de informação, conhecimentos e práticas, reunidos e mantidos por uma organização no decorrer de sua existência, “utilizados para o suporte às suas atividades, seus processos decisórios e para a preservação do seu capital intelectual”.

É possível inferir que a memória organizacional é um diferencial competitivo. Para Nascimento et al. (2016), o diferencial competitivo é o resultado do valor que as pessoas que formam a organização atribuem ao conhecimento e a informação. Isso garante que a história da organização seja registrada e usada ao longo do tempo, para fins decisórios, possibilitando que os gestores embasem suas decisões nas ações que foram tomadas no passado.

Assim, as ações consideradas necessárias para a organização e representação do conhecimento no ecossistema de startups de Florianópolis são:

- 1 Entrevistar por meio de um roteiro semiestruturado os empreendedores de startups, com vistas a obter a origem dos conhecimentos que aplicam em suas empresas e no ecossistema, ou seja, suas fontes de informação.
- 2 Relacionar as fontes de informação e tipos de informação pertinentes a cada etapa do processo inovador do ecossistema de startups de Florianópolis.
- 3 Elaborar modelo de relacionamento das fontes de informação e os tipos de informação com o processo inovador do ecossistema de startups de Florianópolis.

Entende-se por fontes de informação, o conceito de Rodrigues e Blattmann (2014, p. 10) que define “como tudo o que gera ou veicula informação”. Pode-se dizer que é toda forma que atende as necessidades de informação, incluindo serviços e produtos e informação, pessoas, meio digital, como sites e portais (Rodrigues & Blattmann, 2014).

No contexto organizacional para fins de gestão, Choo (2006) entende que as fontes de informação podem ser classificadas em externas e pessoais, externas e impessoais, internas e pessoais, e internas e impessoais, presentes em ações desenvolvidas por todos os colaboradores. É a partir do uso contínuo e efetivo das fontes de informação que se alcança sucesso na pesquisa, no desenvolvimento e na inovação (Cunha, 2016).

Espera-se que por meio das ações descritas se obtenha a organização e representação do conhecimento, consequentemente a categorização dos tipos de informação científica e tecnológica mais utilizadas no ecossistema de startups. Bem como, a possibilidade do estabelecimento de folksonomias e taxonomias, que de acordo com Moreiro-González e Bolaños-Mejías (2015) contribuem para a gestão da informação e do conhecimento.

É a taxonomia que possibilita o funcionamento dos sistemas de representação, organização e recuperação de informações em um vocabulário controlado, que em algumas situações isso pode limitar-se a pesquisa do usuário, quando o resultado das buscas é definido com base em uma lista de termos relacionados com o assunto. A

taxonomia é construída por pessoas especializadas em organizar informações, e não por profissionais de cada área (Dreyfus, 2001).

Já a folksonomia se diferencia por ser um sistema de classificação criado por usuários individuais, pode-se dizer que é construída pelo povo (Guy & Tonkin, 2006). A grande diferença está pautada nas variadas formas de representação para um único conceito ou documento, com uma imensidão de significados atribuídos por diferentes indivíduos oriundos de contextos diversos (Lara, 2009).

Deste modo, o uso institucional e empresarial de folksonomias e taxonomias para a representação da informação científica e técnica pode se tornar uma ferramenta útil para a gestão estratégica das atividades, dos procedimentos, dos produtos e dos recursos humanos das empresas, instituições, e entidades de diversas origens (Moreiro-Gozález & Bolaños-Mejías, 2015).

4 Considerações Finais

Enxerga-se que as possibilidades elencadas não são únicas, e a partir de seus desenvolvimentos torna-se possível detalhar suas especificidades e indicar outros desdobramentos. O que reflete na extensão teórico prática da organização do conhecimento, com a busca de lacunas, proposta de soluções pautadas em estudos já realizados, e aplicação por intermédio dos profissionais da informação.

Nesse sentido, acredita-se que o objetivo estabelecido para este trabalho, indicar as ações necessárias para a organização e representação do conhecimento no ecossistema de startups de Florianópolis, com vistas a construção de sua memória socioeconômica foi alcançado. Além disso, conjecturou-se as implicações da organização e representação do conhecimento no ecossistema de startups, refletindo sobre outros desdobramentos, como o estabelecimento de folksonomias e taxonomias.

Vislumbra-se, o ecossistema de startups de Florianópolis em um campo vasto para estudos pautados nas teorias e práticas da organização do conhecimento, devido aos inúmeros e distintos conhecimentos que nele circulam.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) do Brasil por financiar a bolsa de estudos do doutorado em andamento.

Referências

- Brascher, M.; Café, L. (2008). Organização da informação ou organização do conhecimento? In *Anais do Encontro Nacional de Pesquisa em Ciência da Informação*, 9.
- Brasher, M. & Carlan, E. (2010). Sistemas de organização do conhecimento: antigas e novas linguagens. In Robredo, J.; Brascher, M. (Org.). *Passeios pelo bosque da informação: estudos sobre a representação e organização da informação e do conhecimento*. Brasília DF: IBICT.
- Caldwell, J. C. (1986). Routes to low mortality in poor countries. *Pop. Dev. Rev.*, 12(2), 171-220.
- Choo, C. W. (2003). *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. São Paulo: Senac São Paulo.
- Cunha, M. B. D. (2001). *Para saber mais: fontes de informação em ciência e tecnologia*. Brasília: Brinquet de Lemos.
- Dreyfus, H. L. (2001). *On the internet*. Routledge.
- Guy, M., & Tonkin, E. (2006). Folksonomies: *Tidying up tags?*. *D-lib Magazine*, 12(1).
- Halbwachs, M. (2013). *A memória coletiva* (2a ed.). Tradução de Beatriz Sidou. São Paulo: Centauro.
- Lara, M. L. (2009). *Linguística documentária: seleção de conceitos*. 2009. 178 f (Doctoral dissertation, Tese (Livre-docência) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo).
- Menezes, E. M. (2006). *Estruturação da memória organizacional de uma instituição em iminência de evasão de especialistas: um estudo de caso da CONAB*. Mestrado. Universidade Católica de Brasília.
- Moreiro-González, J. A., & Bolaños-Mejías, C. (2015). Papel de los vocabularios semánticos en la economía en red. Uso institucional y empresarial de folksonomías y taxonomías. *PRISMA. COM*, (29), 121-138. Recuperado em 5 janeiro, 2019, de <http://pentaho.lettras.up.pt/ojs/index.php/prismacom/article/view/1834>
- Maculan, B. C. M. S. (2015). *Estudo e aplicação de metodologia para reengenharia de tesauro: remodelagem do THESAGRO*.
- Nascimento, N. M., de Souza, J. S. F., Valentim, M. L. P., & Moro-Cabero, M. M. (2016). Gerenciamento dos fluxos de informação como requisito para a preservação da memória organizacional: um diferencial competitivo. *Perspectivas em Gestão & Conhecimento*, 6(1), 29-44.

- Rodrigues, C., & Blattmann, U. (2014). Gestão da informação e a importância do uso de fontes de informação para geração de conhecimento. *Perspectivas em Ciência da Informação*, 19(3), 04-29.
- Sena, P., Blattmann, U., & Teixeira, C. (2017). Ecossistema de startups em Florianópolis: possibilidades para profissionais da Biblioteconomia. *RBBB. Revista Brasileira de Biblioteconomia e Documentação*, 13, 2571-2588. Recuperado em 5 janeiro, 2019, de <https://rbbd.febab.org.br/rbbd/article/view/864/957>
- Soares, F. M. & Maculan, B. C. M. S. (2016). Organização e representação do conhecimento em tesauros: uso do sistema e-terms. *Seminário em Ciência da Informação: fenômenos emergentes na Ciência da Informação*.
- Spiller, A., & Cunha Pontes, C. C. (2007). Memória organizacional e reutilização do conhecimento técnico em uma empresa do setor eletroeletrônico no Brasil. *Revista Brasileira de Gestão de Negócios*, 9(25).

Estudo de tipologia documental em arquivos pessoais: contribuições para a organização do conhecimento arquivístico

Mabel Meira Mota¹ y Francisco José Aragão Pedroza Cunha²

¹ ORCID [0000-0003-4175-3669](https://orcid.org/0000-0003-4175-3669). Programa de Pós-graduação em Ciência da Informação, Universidade Federal da Bahia, Brasil.

Mabelmmotagmail.com

² ORCID [0000-0003-2770-7818](https://orcid.org/0000-0003-2770-7818). Programa de Pós-graduação em Ciência da Informação, Universidade Federal da Bahia, Brasil.

pedrozaici@gmail.com

Resumen. Se presenta el desarrollo y resultados parciales de la investigación de maestría desarrollada en la Línea 1 - Políticas y Tecnologías de la Información, del Programa de Postgrado en Ciencia de la Información, de la Universidad Federal de Bahía, acerca de la identificación de los tipos de documentos textuales existentes en el archivo personal de Thales de Azevedo. En el marco de la organización del conocimiento arquivístico, la identificación de tipologia documental es fundamental para el desarrollo de las demás operaciones físicas e intelectuales para que se efectúa la organización de archivos, entre las cuales destacan la clasificación y la descripción. Se pretende, por lo tanto, abordar cuestiones que involucran la identificación de tipologia documental en los archivos personales, evidenciando sus características en lo que se refiere a la producción documental ya la diversidad de tipologías resultantes de la praxis historiográfica emprendida por el titular. Para ello, presentamos una revisión de literatura sobre el estatuto de los archivos personales en la Archivística, y un breve histórico, conceptualización y análisis de la Identificación Documental en la literatura nacional e internacional. Por último, destacamos el papel de la identificación de tipologia documental en el tratamiento de la información en los archivos personales como proceso esencial de la práctica archivística, que comporta un conjunto de actividades orientadas a posibilitar el acceso a los documentos y a la memoria en ellos registrada.

Palabras clave: Organización del conocimiento arquivístico; tipologia documental; archivos personales.

Abstract. The development and partial results of the master's research developed in Line 1 - Policies and Information Technologies, of the Postgraduate Program in Information Science, Federal University of Bahia, about the identification of the types of textual documents in the personal archive of

Thales de Azevedo. In the context of the organization of archival knowledge, the identification of documentary typology is fundamental for the development of other physical and intellectual operations in order to organize the archives, among which stand out the classification and the description. The objective is, therefore, to address issues that involve the identification of documentary typology in personal archives, evidencing its characteristics in relation to documentary production and the diversity of typologies resulting from the historiographical practice undertaken by the owner. To do so, we present a review of the literature on the status of personal archives in Archives, and a brief history, conceptualization and analysis of Documentation Identification in national and international literature. Finally, we highlight the role of identification of documentary typology in the treatment of information in personal archives as an essential process of archival practice, which includes a set of activities aimed at providing access to documents and the memory in them.

Keywords: Organization of archival knowledge; documentary typology; personal archives.

1 Introdução

Origem institucional e processo natural foram determinantes para delimitação do conceito de arquivo e para a disseminação de uma “matriz pública” que designou um lugar periférico para os arquivos fora dos padrões governamentais e institucionais e implicou na utilização de abordagens metodológicas distintas, de acordo com as filiações institucionais e disciplinares dos responsáveis por sua custódia (HEYMANN, 2012).

Inseridos no conjunto dos arquivos privados, os arquivos pessoais foram, e continuam sendo, em grande parte, medidos de acordo com o valor probatório e informativo, próprios aos documentos de natureza institucional, único modo de garantir-lhes uma abordagem arquivística, conforme se pode observar em Belloto (1991), a conceituação de arquivos pessoais está embutida na definição geral de arquivos privados, ao inseri-los no rol de papéis produzidos/recebidos por entidades ou pessoas físicas de direito privado. Para essa autora, não são documentos funcionais e administrativos no sentido que possuem os de gestão de organização pública ou privada, mas registros de vida familiar, civil e profissional e vinculados à produção política e/ou intelectual, científica, artística de estadistas, políticos, artistas, literatos, cientistas etc. (BELLOTO, 1991, p. 254).

A crítica pós-moderna tem demandado novos posicionamentos dos profissionais responsáveis pelo tratamento dos arquivos pessoais, considerados, agora, a partir de sua própria urdidura e especificidade. Contemporaneamente, destaca-se uma postura relativista sobre a prática arquivística, que considera o impacto da instabilidade do sujeito moderno na produção de sentidos nos arquivos (FOUCAULT, 1995). A

observação do arquivamento proposto pelo titular do arquivo interessa, portanto, por facultar uma lógica própria para organizar os registros do que foi vivido no âmbito público e/ou no privado, conforme o uso que deles poderia ser feito. Assim, conforme Heymann (2012), “pensar os arquivos e sua produção implica, portanto, pensar em representações e práticas” (HEYMANN, 2012, p. 36).

Na contramão de uma vertente que enxerga e considera a “dimensão funcional” dos arquivos pessoais como preponderante no arranjo desses acervos, Catherine Hobbs afirma que “these intimate elements are reflected not only in the content and organization of personal records, but come into play in the archivist’s direct interaction with the creator/donor during appraisal, acquisition, and subsequent management of personal archives” (HOBBS, 2001, p. 127). Para essa autora, se os arquivistas se concentrarem apenas na preservação do contexto transacional dos registros, vão estar valorizando-os somente pelo seu uso imediato em determinado contexto, como prova da atividade ou da finalidade para a qual ele foi criado, desconsiderando que os registros quando advindos da instância mais íntima e privada delineiam a inscrição de vários sujeitos numa mesma identidade. Nessa perspectiva, Hobbs (2001) conclui que a melhor alternativa deva ser a observação da personalidade do indivíduo que se arquivava em vez dos comportamentos condicionados socialmente.

Assim, a abordagem sugerida por Hobbs (2016) propõe que proveniência e a ordem original sejam encontradas justamente se forem observados os atos de vontade e de negligência dos produtores, se considerados os diversos elos entre os documentos da esfera pessoal e da esfera pública, se lidas as materialidades dos documentos relacionando-as a sua carga emocional e a sua utilidade prática, se entendermos que o ato de documentar e o comportamento arquivístico são derivados da condição humana em vez de limitá-los à construção de algo memorável – pois nem sempre o é –, enfim, se percebermos que “a documentação é ligada a e parte da experiência dos criadores” (HOBBS, 2012, p. 322, grifos da autora).

Os arquivos pessoais precisam, portanto, ser reintegrados à descontinuidade que lhes é inerente. Inserido no complexo e multifacetado quadro composto pelos arquivos pessoais encontram-se os documentos produzidos e acumulados por Thales de Azevedo. As peculiaridades de sua urdidura são investigadas na etapa metodológica da Identificação Arquivística, a qual torna possível a instauração de rasuras no paradigma probatório, ao afirmarem que o valor transacional é apenas um dos valores encontrados nos arquivos pessoais. Diante disso, apresenta-se o desenvolvimento e resultados parciais da pesquisa de mestrado desenvolvida na Linha 1 – Políticas e Tecnologias da Informação, do Programa de Pós-graduação em Ciência da Informação, da Universidade Federal da Bahia, acerca da identificação dos tipos de documentos textuais existentes no arquivo pessoal de Thales de Azevedo.

2 Metodologia de Identificação

A terminologia “Identificação Arquivística” surge num contexto marcado, principalmente, pelo “boom da memória” decorrente de um rápido processo de mudanças que fragmentaram as paisagens identitárias, culturais, sociais, econômicas, dentre outras: a pós-modernidade. Construída na década de 1980, reflete o diálogo anterior estabelecido entre os arquivos, a História, o Direito e a Diplomática, principalmente entre os séculos XVII e XIX, que, através do estudo crítico dos elementos intrínsecos e extrínsecos dos documentos escritos, conferiam a autenticidade e fidedignidade dos fatos neles registrados. A Diplomática, ao postular conceitos e métodos para a crítica de documentos medievais, demonstrou a força de um documento escrito, que passou a ser sentida na Arquivística a partir de meados do século XX, quando Robert-Henri Bautier (1962) identifica os documentos de arquivo como o objeto de estudo da Diplomática (TOGNOLI, 2013).

A Identificação, no domínio da Arquivística, trata-se de uma investigação, de caráter intelectual, que visa recuperar a relação de indiciabilidade necessária entre os documentos e as funções responsáveis por sua existência. Conforme Rodrigues (2008), a metodologia de Identificação Arquivística é definida como um levantamento de dados acerca do documento de arquivo e dos vínculos que mantém com o órgão ou pessoa que o produziu, seja em fase de produção ou de acumulação. Neste sentido, é um trabalho de pesquisa e de crítica sobre a gênese documental (RODRIGUES, 2008, p. 22).

Tendo como base o Princípio da Proveniência ou Respeito aos Fundos, nessa etapa recolhem-se informações relativas a “quem?”, “quando?”, “como?” e, principalmente, “por quê?” os documentos foram produzidos e acumulados como registros probatórios do fazer específico de cada organização (LOPEZ, 2012), ou seja, reconhece-se a gênese documental. Rodrigues (2008), inserida no âmbito da documentação de matriz organizacional, explica que essa fase do tratamento técnico, contempla, primeiramente, o entendimento profundo do organismo produtor (evolução orgânica e competências administrativas), seguido do elemento funcional (competências, funções e atividades) e seu registro sob a forma de tipo documental (procedimento administrativo e tramitação) e, por último, a delimitação da série documental (RODRIGUES, 2008, p. 68-69).

Nesta fase também são estudados os elementos internos e externos dos documentos, que tornam possível a delimitação dos vínculos e das atividades que tornaram possível o tipo documental, para sua adequada denominação. O tipo documental vem a ser a “divisão da espécie documental que reúne documentos por suas características comuns no que diz respeito à fórmula diplomática, natureza de conteúdo ou técnica do registro” (DICIONÁRIO, 2005, p. 163). Conforme Bellotto (2002, p. 20), o tipo documental é “a configuração que assume uma espécie documental, de acordo com a atividade que a gerou”.

O arquivo pessoal reflete uma justaposição de intenções, acontecimentos e práticas. Por esse motivo, tais arquivos demandam uma ampliação do aporte teórico e

metodológico da Arquivística também no sentido de entendê-los como “evento cultural e social”.

3 Resultados

Thales de Azevedo (1904-1995) ocupa um lugar de destaque na cena intelectual baiana e brasileira, não apenas pelo papel que desempenhou como professor e pesquisador, mas por ter explorado o ousado e o inovador na sua área de atuação, as Ciências Sociais. Voltando-se para os rituais do cotidiano, analisou a sociedade de sua época em suas regras e ritos. Considera-se, assim, o seu arquivo como uma construção inserida nos rituais do cotidiano do seu produtor, ou seja, como uma arte de fazer forjada numa prática de resistência própria ao arquivamento de si e, ao mesmo tempo, atravessada por outras subjetividades. Sendo assim, apresenta-se um breve relato sobre a experiência de aplicação da metodologia arquivística da Identificação ao arquivo pessoal dessa personalidade baiana.

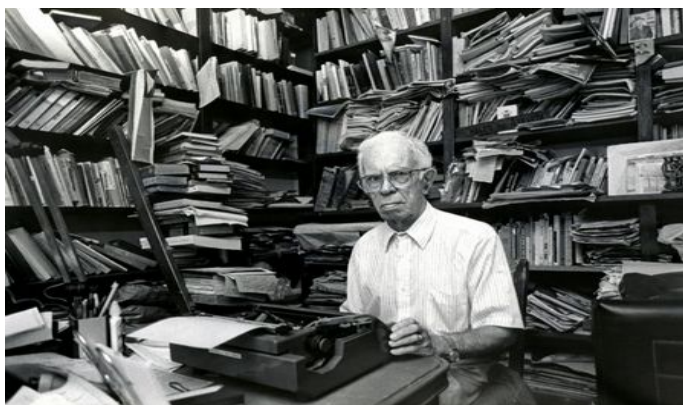


Figura 1. Thales de Azevedo em seu gabinete.

Trata-se de conjunto documental que reúne, em média, 50 mil documentos textuais, divididos entre produzidos e recebidos pelo titular; e uma biblioteca composta, atualmente, por cerca de 300 livros, de um montante anterior superior a 5 mil livros, catalogados em dois cadernos manuscritos. A biblioteca de Thales de Azevedo foi doada, em parte, pela família. Atualmente podem ser encontrados livros com o “carimbo do titular” na biblioteca do Colégio Estadual Thales de Azevedo, em Salvador; e na biblioteca da Faculdade de Filosofia e Ciências Humanas, na Universidade Federal da Bahia. No primeiro, é possível também encontrar alguns dos 21 quadros pintados por Thales de Azevedo de que temos conhecimento.

O volume de caixas, pastas, dossiês e empilhamentos, dão a ler múltiplos fragmentos biográficos. Observa-se que a multiplicidade de campos em que atuara Thales de Azevedo encontra-se representada em 52 caixas-arquivo, 40 pastas, 23

dossiês e 12 empilhamentos, nos quais se destacam os cargos ocupados, principalmente, como presidente da ABA (Associação Brasileira de Antropologia) e também do Instituto Geográfico e Histórico da Bahia (IGHB) – do qual fora presidente em 1978 – possui maior volume nesses períodos – em torno de 8 caixas-arquivo –, principalmente no que tange à correspondência passiva. Encontra-se pouco documentada, contudo, sua atuação como médico no interior da Bahia.

A TARDE		CADERNO	PÁGINA	LAUDA
REPORTER	ASSUNTO		DIAGRAMAÇÃO	
RETRANCIA DO TÍTULO	ALTURA DA MATÉRIA		PIQUE	COMP.
BANCO	ARQUIVO (S)	OBSERVAÇÕES	REDAZ	INICIAL/PRIMEIRO

1	2	3	4	5	6
PARA CONHECER A BAHIA					
Thales de Azevedo					
<p>Acredito que paira na Bahia algo de novo na curiosidade por nossas coisas. Indaga-se sobre a história, o desenvolvimento, o decorrer da nossa história. Algo é isto parecido com o que se passou a partir da fundação de nossa Faculdade de Filosofia sob a inspiração de Isaias Alves, dando lugar a um novo ensino das nossas coisas a começar pela história que tinha por exemplo, um precursor científico em Braz do Amaral e recebeu, a partir de então, a contribuição de Frederico Edelweiss entre alguns. Respeito a respeito de seu trabalho renovador escreveu o prof. Consuelo de Sena de Sena. Essa revisão de nossa historiografia, em que releva o nome de Pedro Calmon, teve um momento de interesse que necessita agora ser atualizando. Inaugurando uma visão melhor documentada e atualizando dessas disciplinas, surgiram nomes e obras para as quais é preciso novamente chamar a atenção. Foi um tempo propício, em que se destacou Wanderley Pinho e se fixou muito de sua influência sobretudo nos cursos da Faculdade de Filosofia, órgão da Universidade Federal.</p> <p>É hora de fazer uma análise do que então se produziu de renovação de métodos e de atualização de dados arquivísticos. Mas resta ainda bem a empreender sobretudo de material primário dos arquivos e de outras fontes. Precisamos de nova contribuição dos cursos de história e de outras ciências sociais, em que se destaca ultimamente o prof. J. Reis e muito o polígrafo mestre Wladimir Freitas Oliveira.</p> <p>Uma necessidade a ser preenchida é a de organizar um acervo de obras substanciais sobre nossa gente. Precisamos saber bem como se formou e se continua formando a Bahia e quem melhor abordou essa temática.</p> <p>Os cursos de nossa história e de saberes parelhados devem dar agora renovada contribuição a tais indagações. Uma coisa a fazer é reunir a bibliografia de um mestre como Pedro Calmon cuja obra tem segura base. Dados dessa ordem aparecem nos livros por exemplo, de Luis Henrique Dias Tavares, de Wladimir Pinho.</p>					

Artigo datilografado e corrigido em formulário de A Tarde

Figura 2. Produção jornalística de Thales de Azevedo (Minuta)

A dimensão privada ganha destaque pelos dossiês de pesquisa, que remetem para o exercício constante da escrita e da leitura. Há vasta correspondência com intelectuais nacionais e estrangeiros, com amigos, colegas, colaboradores e, em menor volume, com os familiares. Além disso, destaca-se a correspondência com intelectuais e docentes de Universidades Americanas de Miami, da Flórida, da Geórgia e da Califórnia. O arquivo inclui também uma vasta produção jornalística de Thales de Azevedo, principalmente em *A Tarde* (Cf. Figura 2), no qual atuara desde 1922, inicialmente como auxiliar e, posteriormente, atuando como articulista de produção semanal até a semana de sua morte. Trata-se de mais de 1400 artigos publicados,

principalmente a partir de 1975, quando ele passa a fazer parte do grupo de colaboradores permanentes do jornal (BRANDÃO, 1993).

Documentos de processo advindos dos diferentes ciclos de pesquisa possuem lugar de destaque no arquivo desse baiano, além de trabalhos de terceiros, enviados seja para Thales de Azevedo emitir opinião, seja para prefaciar uma publicação. Através dos primeiros é possível adentrar a dimensão privada da escritura, observando a relação entre o que pesquisador lia e escrevia, pois em alguns dos livros de sua biblioteca estão registradas anotações, marcações. Os livros lidos tornam-se, assim, dossiês (Cf. **Figura 3**), devido ao acréscimo de notas de leitura, de recortes e algumas análises. Fatos, versões e ficções convivem nesse arquivo.



Figura 3. Livro de terceiro com marcações e anotações de Thales de Azevedo

Um volume expressivo de documentos arquivados em 29 caixas-arquivo, sem dúvida o maior, dá a ler a sistematização de uma intensa pesquisa dedicada a diversos ciclos temáticos, dos quais se destaca um conjunto de dez mil documentos – entre notas de campo, recortes de jornais e versões manuscritas de textos inéditos e daqueles publicados – referentes ao ciclo de pesquisa sobre as variações culturais inter-regionais e a migração, que remetem para diversas publicações acerca da presença de italianos no Rio Grande do Sul, das quais são expressivas as obras *Gaúchos: notas de antropologia social* (1943), *Italianos e gaúchos: os anos pioneiros da colonização italiana no Rio Grande do Sul* (1975); *Os italianos no Rio Grande do Sul: cadernos de pesquisa* (1994), esse último revisto e publicado um ano antes de seu falecimento; e do estudo de italianos na Bahia, expressos em *Italianos na Bahia e outros temas* (1989) e no conto-memória, *A Filha do Alferes: nas redondezas das Guerras do Sul* (1993).

A variedade de temas pesquisados por Thales de Azevedo, para além da grande acumulação e produção documental, reflete, em geral, a coincidência temporal dos ciclos temáticos, muitas vezes estudados e documentados simultaneamente, como é o caso daquele sobre a imigração de Italianos no Rio Grande do Sul e na Bahia, e aqueles voltados para os temas do cotidiano, o ciclo da vida e seus ritos. Conforme Maria de Azevedo Bandão (2004, no prelo), num texto inédito, provável ampliação da primeira edição de *Dados de uma Assinatura* (1993), o cotidiano como tema de interesse de Thales de Azevedo já aparece em 1941, no primeiro ensaio sobre “O rancho de gaúchos brasileiros e uruguaios” e nas notas de uma viagem ao Uruguai, publicadas em *Gaúchos* (1943).

No arquivo pessoal do médico, antropólogo e professor Thales de Azevedo é possível perceber práticas ativadas em decorrência de atividades cotidianas, como é o caso da preservação dos documentos de identificação comuns e até mesmo um cartão que comunica à sociedade o seu nascimento (Cf. **Figura 4**), comprovantes de pagamento de contas, exames médicos, diplomas e históricos acadêmicos, dentre outros.

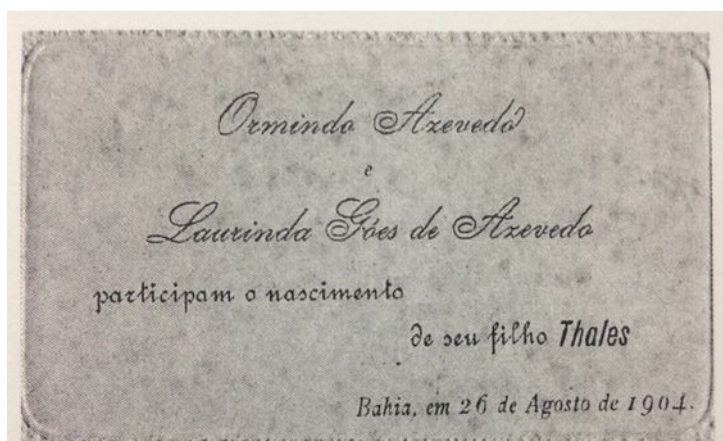


Figura 4. Cartão que comunica nascimento de Thales de Azevedo em 1904

Práticas ativadas por motivos e interesses diversos são observadas no seu arquivo, e revelam, antes de tudo, a função instrumental do arquivo para o seu produtor. Essa função instrumental pode ser lida, principalmente, pelo processo de documentação dos densos ciclos de pesquisa por ele sistematizados em torno de vários temas, como o cotidiano. Dessa tentativa de compreender as simbologias e significados próprios de práticas rotineiras, resultou um conjunto formado por: anotações de campo e de pesquisa documental, notas avulsas, folhas soltas, agendas pessoais, fotos, desenhos, listas de endereços, recortes de jornais sobre variados temas, resenhas e comentários de textos lidos, manuscritos inéditos e manuscritos de textos publicados, roteiros de aulas, seminários, apostilas, entrevistas e reportagens referentes ao tema estudado, manuscritos de terceiros, separatas recebidas de colaboradores e amigos, correspondências, bibliografias e fichas bibliográficas, resumos analíticos de cada

leitura empreendida, almanaques consultados, além de livros em cujas margens observam-se as marcas de um leitor atento.

O conjunto representativo de tais ciclos dá a ler processos de elaboração de empreendimentos intelectuais densos, nos quais não se pode estabelecer o início ou o fim, pois o tema do cotidiano, por exemplo, no âmbito dos ciclos temáticos de pesquisa empreendidos pelo titular, perpassa diversas produções intelectuais e artísticas que suplementam umas às outras e fazem uso de dados e informações muitas vezes comuns, mas que são ressignificadas a cada retomada do tema pelo autor.

4 Considerações finais

A pesquisa de mestrado desenvolvida na Linha 1 – Políticas e Tecnologias da Informação, do Programa de Pós-graduação em Ciência da Informação, da Universidade Federal da Bahia, tem demonstrado que, durante a vida, as pessoas produzem documentos por razões muito diversas, nem sempre objetivando a comprovação de uma atividade, função ou cargo desempenhado. O arquivamento nem sempre é consciente ou voluntariamente construído para o futuro, podendo ser motivado por dar uma resposta à injunção social de comprovar o vivido; assim como pelo cultivo de emoções e experiências.

O trabalho de Identificação ainda não foi concluído, mas revela-se, desde já, importante para o futuro arranjo, descrição e difusão. Percebeu-se que para Thales de Azevedo o arquivo para além de lugar para guardar suas lembranças, tinha uma função instrumental, pois ali estava tudo que precisava para alimentar o seu ininterrupto fluxo de criativo, dado a ler em textos sempre inacabados – versões de versões –, sempre passíveis de retomadas e contínuas experimentações.

Referencias

- BELLOTTTO, Heloisa Liberalli. (1991). *Arquivos permanentes: tratamento documental*. Rio de Janeiro: Editora FGV.
- BRANDÃO, Maria de Azevedo. (1993). *Thales de Azevedo: dados de uma assinatura*. Salvador: ABA /UFBA.
- BRANDÃO, Maria de Azevedo. (2004). *Thales de Azevedo: dados de uma assinatura*. Salvador: UFBA. No prelo.
- FOUCAULT, Michel. (1999) *Nietzsche, Freud e Marx: theatrum filosoficum*. Tradução Jorge Lima Barreto. São Paulo: Princípios.
- HEYMANN, Luciana. Indivíduo, memória e resíduo histórico: uma reflexão sobre arquivos pessoais e o caso Filinto Müller. (1998). *Estudos Históricos*, 19, 41-66.

- HEYMANN, Luciana (2012). *O lugar do Arquivo: a construção do legado de Darcy Ribeiro*. Rio de Janeiro: Editora Contra Capa.
- HOOBS, Catherine (2001). The Character of Personal Archives: Reflections on the Value of Records of Individuals. *Archivaria*, 52, p.126-135.
- HOOBS, Catherine. (2016). Vislumbrando o pessoal. Reconstituindo traços de vida individual (pp. 303-342). En *Correntes atuais do pensamento arquivístico*. Belo Horizonte: Editora UFMG.
- LOPEZ, André Porto Ancona. (2003). Arquivos pessoais e as fronteiras da arquivologia. *Gragoatá*, Niterói, 154, p. 1-140.
- RODRIGUES, Ana Célia. (2008). *Diplomática contemporânea como fundamento metodológico da identificação de tipologia documental em arquivos*. Tese (Doutorado em História Social) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.
- RODRIGUES, Ana Célia; PENHA, Noemi Andreza da. (2004). A noção de identificação arquivística na Espanha e no Brasil nas décadas de 1980-1990. *Prisma.com* – Revista do Centro de Estudos das Tecnologias e Ciências da Comunicação, 21, p. 1-28.
- TOGNOLI, Natália Bolfarini. (2014). *A construção teórica da Diplomática: em busca de uma sistematização de seus marcos teóricos como subsídio aos estudos arquivísticos*. São Paulo: Editora Unesp.

Propuesta teórica para la categorización y el etiquetado semántico de los archivos familiares a través del Fondo Kati.

Joseba Martínez de Lahidalga Santillana

ORCID [0000-0003-2750-9432](https://orcid.org/0000-0003-2750-9432). Archivero en la Fundación Sancho el Sabio Fundazioa, España.
josebalm07@gmail.com

Resumen: La reciente tendencia que está incrementando la visibilidad y la preservación de los archivos familiares está produciendo un nuevo desafío para los archiveros, ya que además de procurar la gestión y conservación física de esta tipología de archivos, tienen que introducirlos y administrarlos en las aplicaciones digitales. Siendo necesaria la constitución de una nueva metodología que no solo ordene sino que también corrija las carencias de los ficheros digitales. En consecuencia, por medio de un caso real, el archivo familiar Kati, se plantean dos procedimientos, la categorización y el etiquetado semántico, enfocados a resolver algunas de las dificultades que se generan tras la digitalización y que facilitan la preservación y divulgación de los archivos familiares en entornos digitales.

Palabras Clave: Archivo familiar Kati; Categorización digital; Etiquetado semántico.

Abstract: The recent trend that is increasing the visibility and preservation of family archive is producing a new challenge for archivists. Because as well as they ensure the management and physical conservation of this type of archives they have to introduce and manage them in digital apps. It is for these reasons that it is absolutely necessary the set up of a new methodology that not only arranges the digital archives but also corrects their deficiencies. Consequently, through a real case, the Kati family archive, two procedures are proposed, precisely the categorization and semantic tagging. Which there are focused on solving some of the difficulties that are generated after digitalization and facilitate the preservation and dissemination of the files in digital environments.

Keywords: Kati family archive; Digital categorization; Semantic tagging.

1 Introducción

A causa del mayor interés y estudios realizados en torno a los archivos familiares, principalmente desde la última década del siglo XX, actualmente se está dando una tendencia que está impulsando la preservación de los fondos documentales reunidos por familias, que no tienen que ser obligatoriamente de origen nobiliario. Este factor está generando la formación de un patrimonio documental bastante rico que está facilitando no solo el estudio de los acontecimientos más conocidos, con los archivos nobiliarios, sino que también los ocurridos en lugares concretos que hasta ahora han pasado desapercibidos.

Esta situación está produciendo un nuevo desafío para los archiveros ya que además de procurar la organización, clasificación y conservación física de esta tipología de archivos, tienen que introducirlos a los repositorios digitales por medio de la digitalización. Siendo necesaria la constitución de una nueva metodología que no solo ordene sino que también corrija las carencias de los ficheros digitales, entre otras, la pérdida de relaciones y vínculos que poseen los documentos en soportes tradicionales.

En consecuencia, durante los siguientes apartados a través de un caso real, en concreto el archivo familiar Kati, se proponen dos técnicas enfocadas a facilitar la preservación y divulgación de los archivos familiares en entornos digitales. En particular, la categorización por medio del cuadro de clasificación y la identificación y vinculación a través del etiquetado semántico.

2 El archivo familiar Kati o Fondo Kati

Un ejemplo de archivo familiar es el patrimonio documental formado desde principios del siglo XVI por los descendientes de *Ali b Ziyad al-Quti* y su hijo *Mahmud Kati*. Este patrimonio familiar se denomina archivo familiar Kati o Fondo Kati porque tal y como se ha indicado lo crearon los miembros del clan Kati, principalmente en Tombuctú (Mali), a partir de los manuscritos que generaban y también con los que compraban o conseguían. Si bien, antes de que *Ali* se estableciera en Tombuctú a finales del siglo XV, la familia Kati vivió en los reinos de la Península Ibérica donde empezó a reunir la documentación que relataba su historia, labores y los acontecimientos vividos.

Desde un punto de vista europeísta, la importancia de este archivo y el principal motivo del interés para su preservación radican en su contenido tan singular y en su valor para reconstruir la historia de la cultura andalusí y judaica durante la Edad Media en la Península Ibérica. No obstante, también se considera una rica fuente de información sobre la historia del África Subsahariana, en concreto del Imperio *Songhay*. En definitiva el Fondo Kati no solo es el legado histórico de una familia sino que también es aconsejable considerarlo como bien de interés nacional e

internacional por ser una de las pocas fuentes directas que trata sobre los asuntos indicados.

A consecuencia de los sucesos ocurridos durante los pasados siglos, en particular en el siglo XIX y principios del XX, todo el patrimonio reunido por los Kati se dispersó entre los miembros de la familia, cada cual guardándolo en lugares secretos que habitualmente eran sitios donde no estaba garantizada su preservación, agravando de esta manera sus problemas de conservación (Diadié Haidara y Pimentel, 2017). Hasta que *Ismäel Diadié Haidara*, actual responsable del Fondo, decidió continuar con la misión comenzada por su padre, fundamentada en la recuperación del patrimonio familiar, para poder darlo a conocer al mundo académico y conseguir ayudas para mejorar su situación y garantizar su preservación. Estas ayudas, y también colaboraciones, se han sucedido de dos formas diferentes. Por un lado, aquellas que fomentan la conservación del Fondo en su ubicación original, es decir en Tombuctú. Cabe subrayar la subvención valorada en 120.000 dada por el Junta de Andalucía en 2002 para la construcción en Tombuctú de un edificio que albergara la documentación de la familia (BOJA 125, 2002). Por el contrario, las que con acuerdos o tratados de colaboración mueven de Tombuctú el patrimonio familiar con fines específicos. Por ejemplo, al Instituto Andaluz del Patrimonio Histórico, en Sevilla, para la restauración de cinco manuscritos andalusíes, o asimismo, a la ciudad de Cádiz para su conservación temporal, exposición y estudio (La Fundación Fondo Kati tendrá sede en la provincia de Cádiz antes de que termine el año 2019, 2018).

Resumidamente y dejando a un lado los aspectos más físicos, en la actualidad el archivo familiar Kati es un archivo histórico que consta de aproximadamente 12.000 documentos. Están escritos principalmente por autores de origen árabe y en árabe aunque también se identifican autores y lenguas no arábigas, como el romance castellano. Asimismo, a pesar de que la documentación no solo procede de la familia Kati, ya que como es habitual en esta clase de archivos los Kati obtuvieron los patrimonios documentales de otros linajes o familias, se perciben tipologías documentales similares, por ejemplo: títulos de propiedad, actas de matrimoniales y de pleitos, correspondencia personal y comercial, etc. En concreto, estas semejanzas en la tipología documental facilitan los procesos que en el próximo apartado se describen.

Pese a la labor realizada por *Ismäel Diadié*, la preservación del patrimonio de los Kati sigue estando en peligro debido principalmente a dos motivos. Por un lado, las amenazas externas intentan localizar el Fondo Kati bien para destruirlo o bien para robarlo y venderlo en el mercado negro. Estas situaciones han generado la detención de las ayudas recibidas para afianzar el Fondo en Tombuctú y la vuelta al anonimato. Por otro lado, si continúan los problemas ocasionados por el depósito de la documentación en espacios

inapropiados se puede llegar a producir la pérdida parcial o completa de la información que contienen.

Reflexionando sobre la acción más eficaz y legalmente admisible para mejorar la continuidad del patrimonio documental Kati, resulta que la opción más clara es su digitalización y posterior preservación digital. Puesto que además de conseguir una copia digital del fondo se mejoraría el estado físico de los documentos y se conocería la extensión y el valor exacto del Fondo, gracias a los procedimientos que se llevarían a cabo previos a la digitalización. No obstante, esta opción tiene sus problemas o limitaciones, como el lugar donde realizar la captura digital o las autorizaciones para su gestión digital, que se tienen que solucionar antes de poder efectuarlo.

3 Técnicas para categorizar y vincular los manuscritos digitales

Entre los diversos procedimientos a realizar durante la digitalización y posterior administración digital de las copias digitales, cabría subrayar el diseño del cuadro de clasificación, junto con la ordenación de la documentación según este esquema, y el etiquetado semántico. Debido a su importancia en el proceso para la recuperación de las relaciones y vínculos que posee la documentación original, los cuales se pierden en la captura digital.

3.1 La categorización del fondo

En primer lugar, el cuadro de clasificación es la herramienta por la cual se basa la organización del fondo documental. Se suele optar entre dos tipologías, o bien la funcional basada en la finalidad u objetivos con los que se generaron los expedientes, o bien por materias fundamentado en la tipología documental o en la temática que posee la documentación (Mastropierro, 2006).

En concreto, para el archivo familiar Kati se decide emplear un cuadro de clasificación por materias asentado en las tipologías documentales que se distinguen en el fondo, creando tantas secciones como productores se identifican. Es decir, por cada productor se utiliza las mismas unidades documentales diseñadas en el cuadro de clasificación. Un ejemplo de cuadro de clasificación válido para el Fondo Kati, siempre y cuando se realicen algunas modificaciones, especialmente en aspectos culturales y religiosos, es el cuadro propuesto por Borja Aguinalgalde (Aguinalgalde, 1992):

Tabla 1. Cuadro de clasificación por materias.

Sección	Serie
Genealogía-Heráldica	1.1. Certificación de Escudos de Armerías 1.2. Certificación de nobleza, genealogía y armas 1.3. Probanza de nobleza 1.4. Cuadros genealógicos 1.5. Obras genealógicas, manuscritos o impresas 1.6. Noticias o dibujos sobre armerías
Pleitos	2.1. Autos originales 2.2. Real Carta Ejecutoria 2.3. Memoriales ajustados 2.4. Informaciones de testigos 2.5. Informes jurídicos y memoriales de abogados
Transmisión de bienes	3.1. Contrato matrimonial 3.2. Dotes matrimoniales y religiosas 3.3. Testamento 3.4. Codicilo 3.5. Fundación de Mayorazgo 3.6. Donaciones 3.7. Cesiones 3.8. Inventario de bienes 3.9. Partición de bienes 3.10. Tutela o curaduría 3.11. Renuncias 3.12. Poderes 3.13. Carta de pago 3.14. Mejora de tercio y quinto
Administración de patrimonio	4.1. Almoneda 4.2. Apenamiento 4.3. Arrendamiento 4.4. Carta de pago 4.5. Compra 4.6. Contrato 4.7. Correspondencia 4.8. Cuentas 4.9. Hipoteca 4.10. Imposición de censo 4.11. Juro 4.12. Libramiento 4.13. Libros de administración 4.14. Obligación

	4.15. Permuta 4.16. Planos 4.17. PoderesRedención de censo 4.18. Recibo 4.19. Traspaso 4.20. Venta
Actividad pública	
Correspondencia personal	
Honores y privilegios	7.1.Mercedes diversas 7.2.Cartas de sucesión o confirmación 7.3.Nombramiento para cargo público
Relación con la Iglesia	8.1.Certificación de Partida sacramental 8.2.Documentos de Patronato 8.3.Fundación de capellanía u Obra Pía 8.4.Reparto de dotaciones de Obras Pías
Varios	

Además de organizar, a través del cuadro de clasificación se categorizan todos los documentos que forman el Fondo Kati siendo posible agruparlos en las mismas unidades documentales, en otras palabras secciones y series. Es decir, gracias al emparejamiento de documentos con similares características, en este caso basado en la tipología documental y en un segundo nivel en la temática, se consigue agrupar en un único nivel toda la documentación que comparte la misma tipología documental producida por el correspondiente productor. Facilitando de esta manera su recuperación y relación con el resto de los expedientes de la unidad. Puesto que se agiliza su localización al poder ser buscado a través de la clase documental y mediante el lenguaje que se crea a partir del cuadro de clasificación. A modo de ejemplo, cualquier documento que trate sobre cesión de bienes entre familiares o personas ajenas a la familia se estructura en torno a la sección de transmisión de bienes, separándose según su tipología en las diferentes series como las formadas por los testamentos, contratos matrimoniales o la división de bienes. Esta última clase agrupa tanto notas, cuentas y acuerdos.

3.2 La etiquetación semántica

Por otro lado, a diferencia del etiquetado identificativo, es decir las firmas que otorgan valores únicos e irrepetibles, el etiquetado semántico facilita las relaciones entre ficheros por medio de palabras claves compartidas. Asimismo, aumenta el valor del patrimonio al simplificar la recuperación y visualización de los documentos que lo forman. Este etiquetado se genera a través de la creación de una estructura digital compuesta por nodos generales (concepto genérico) y específicos (concepto

preferido), relacionados entre ellos (García Marco, 2016). Estos nodos están formados por palabras claves que definen la descripción correspondiente a una propiedad identificada del fondo. Estas etiquetas son más eficaces si se originan desde un vocabulario específico que esté fundamentado en un lenguaje especializado con las formalidades y fortalezas que le aporta la estructura de un lenguaje documental tan coherente como un tesoro (Marzal García-Quismondo, 2016).

En el caso que nos concierne, los nodos o términos se crean a partir de la terminología utilizada en el diseño del cuadro de clasificación del Fondo Kati, si bien con el fin de completarlos se añaden descriptores correspondientes a otras cualidades de los documentos, tal y como se observa en la siguiente tabla:

Tabla 2. Ejemplo de descriptores al estilo de un tesoro

C. Específico	C. Genérico	C. Asociados	Grupo	Descripción
Codicilo	Testamento	Hijuela Poderes Herencia	Transmisión de bienes	Última voluntad que no contiene la institución del heredero y que puede otorgarse en ausencia de testamento o como complemento de él (RAE).
Hijuela	División de bienes	Codicilo Poderes Herencia	Transmisión de bienes	Documento donde se reseñan los bienes que tocan en una partición a cada uno de los partícipes en el caudal que dejó un difunto (RAE).

A modo de aclaración, los dos términos, codicilo e hijuela, pertenecen a los grupos formados con los conceptos vinculados con las locuciones testamento y división de bienes pertenecientes al nodo de transmisión de bienes. Y a pesar de que no coincidan en el concepto genérico comparten locuciones relativas ya que tanto codicilo e hijuela como poder y herencia están asociadas con la tipología documental de herencias y sucesiones.

En definitiva, al emplear las etiquetas semánticas además de proporcionar al Fondo Kati de las ventajas propias que genera el tesoro y de recuperar las carencias causadas por la digitalización, se proporciona una relación más segura entre los documentos que forman las unidades documentales y facilita también su vinculación con aquellos expedientes que aunque no se clasifican en la misma unidad comparten propiedades. Del mismo modo, las etiquetas agilizan

la localización y recuperación de los expedientes ya que los sistemas de administración digital y los buscadores se basan o permiten la búsqueda mediante descriptores.

4 Conclusiones

Tal y como se indica, el procedimiento más eficaz y menos dañino para la preservación de los archivos, no solo los familiares, es la digitalización y su posterior gestión mediante aplicaciones tecnológicas. No obstante, para evitar cometer daños tanto en la documentación original como en las copias digitales es necesario, por un lado, adoptar ciertas pautas y técnicas basadas en la metodología archivista pero ajustadas a las características del archivo, y por otro lado, respetar las necesidades de cada procedimiento y la legislación. Con especial relevancia en el método por el cual la documentación se deposita en la institución que va a llevar a cabo la digitalización, sobre todo con los archivos familiares que poseen su origen en el extranjero, como es el caso del Fondo Kati.

Entre todos los procesos a realizar durante la digitalización se han destacado y explicado la categorización a través del cuadro de clasificación y el etiquetado semántico porque gracias a la aplicación y buen uso de ambas herramientas es posible paliar la mayoría de los inconvenientes o riesgos que se producen en la digitalización de cualquier archivo familiar. Si bien, ambos procedimientos se deben empezar a realizar antes de proceder con la digitalización ya que el diseño tanto del cuadro de clasificación como de los nodos requieren del conocimiento completo del alcance del fondo.

Bibliografía

- Aguinagalde Olaizola, F. B. (1992). Los Archivos de familia definición, estructura, organización lección de ingreso en la R.S.B.A.P. *Nuevos extractos de la R.S.B.A.P.*, 5, 9-35.
- Diadié Haidara, I.; Pimentel, M. (2017). *Diario de un Bibliotecario en Tombuctú*. Córdoba: Almunzara.
- García Marzo, J. (2016). Normas y estándares para la elaboración de tesauros de patrimonio cultural. En *El lenguaje sobre el patrimonio. Estándares documentales para la descripción y gestión de colecciones* (pp. 29-46). Ministerio de la Educación, Cultura y Deporte.
- La Fundación Fondo Kati tendrá sede en la provincia de Cádiz antes de que termine el año 2019 (7 de agosto de 2018). *Europa Press*. Recuperado de [https://www.europapress.es/andalucia/cadiz-00351/noticia-fundacion-fondo-kati-tendra-sede-provincia-cadiz-antes-terminate-ano-2019-20180807143952.html](https://www.europapress.es/andalucia/cadiz-00351/noticia-fundacion-fondo-kati-tendra-sede-provincia-cadiz-antes-termine-ano-2019-20180807143952.html).

- Marzal García-Quismondo, M. A. (2016). Tesoros, vocabularios y listas de control terminológico. Fuentes para la construcción de tesauros culturales. En *El lenguaje sobre el patrimonio. Estándares documentales para la descripción y gestión de colecciones* (pp. 93-104). Ministerio de la Educación, Cultura y Deporte.
- Mastropierro, M.C. (2006). *Archivos privados: análisis y gestión*. Buenos Aires: Alfagrama.
- Real Academia de la Lengua. (2018). *Diccionario de la lengua española*. Recuperado de <http://dle.rae.es/index.html>.
- Resolución de 4 de octubre de 2002, de la Secretaría General Técnica, por la que se hace pública la concesión de la subvención de carácter excepcional que se cita. *Boletín Oficial de la Junta de Andalucía*. Sevilla, 26 de octubre 2002, núm. 125, pp. 20.796.

El archivo y la biblioteca de la familia Belmonte-Chico de Guzmán: una propuesta de gestión conjunta como Fondo Documental Patrimonial Familiar

Camino Sánchez Oliveira

Departamento de Ciencias de la Documentación e Historia de la Ciencia, Universidad de Zaragoza, España.

camsanch@unizar.es

Resumen. Se presenta parte de una investigación en curso, cuyo objetivo es definir y reflexionar sobre los elementos que fundamentan un sistema de gestión de lo que se ha denominado como un Fondo Documental Patrimonial Familiar (FDPF), compuesto por un fondo bibliográfico (una biblioteca), un fondo archivístico (un archivo familiar) y, con bastante frecuencia, una colección de objetos materiales relacionados con las actividades de los generadores del FDPF. El objetivo es reflexionar sobre la definición de un sistema de gestión conjunto en base a su conceptualización patrimonial, a su procesamiento documental y a su funcionalidad, así como atender al tipo de centro a cargo de dicho patrimonio. Se expone la experiencia de gestión de un caso particular de titularidad privada: la biblioteca y el archivo de la familia Belmonte-Chico de Guzmán. Nos referimos a un fondo histórico, de carácter bibliográfico y archivístico, aportado y generado por los miembros de las distintas generaciones, ramas y casas que componen esta familia castellana de origen nobiliario- hidalgo, a lo largo de varios siglos (desde finales del s. XV hasta inicios del siglo XX).

Palabras clave: Fondo Documental Patrimonial Familiar, Patrimonio Bibliográfico, Patrimonio Cultural, Patrimonio Documental, Sistema de Gestión.

Abstract. The paper presents a current research and examines the design elements of a Familiar Heritage Documentary Collection, composed by bibliographical (a library) and archival (a family archive) collections, and, with enough frequency, a collection of objects related to the activities of the generators of these kind of heritage. The objective is to reflect on the definition of an integrator management system, with particular reference to patrimonial conceptualization and organization documental systems. It is exposed the experience of managing a particular case of private ownership the library and archive of the Belmonte-Chico de Guzmán family. We refer to a historical collection, of bibliographic and archival character, contributed and generated by

the members of the different generations, branches and houses that compose this Castilian family of noble origin, throughout several centuries (from the end of the 15th century to the beginning of the 20th century).

Keywords: Familiar Heritage Documentary Collection, Bibliographic heritage, Cultural heritage, Documentary heritage, Management system.

1 Presentación

Este trabajo forma parte de la investigación que se está llevando a cabo para la elaboración de la tesis doctoral, todavía en curso, en la que se está estudiando lo que hemos denominado como un Fondo Documental Patrimonial Familiar (FDPF). Con este concepto se ha abordado el análisis de un caso concreto: el de la documentación privada de la familia Belmonte-Chico de Guzmán. Nos referimos a un fondo histórico, de carácter bibliográfico y archivístico, aportado y generado por los diferentes miembros de distintas casas y ramas de una “familia de familias”, un linaje, originaria de Castilla La Mancha y de carácter nobiliario-hidalgo, a lo largo de varios siglos que comprenden desde la Edad Moderna (finales del s. XV) hasta la Edad Contemporánea (s. XX).

Las fuentes utilizadas han sido los propios fondos documentales patrimoniales familiares, que están custodiados en la casa solariega del s. XVIII, propiedad de la familia, que hasta el día de hoy habían permanecido desconocidos y sin haber recibido ningún tipo de análisis o tratamiento documental científico. En concreto estamos hablando de un fondo bibliográfico, una biblioteca familiar, compuesta por más de 600 ejemplares, entre monografías y publicaciones seriadas, todos ellos impresos entre el siglo XVI y el siglo XIX. Y un fondo archivístico, un archivo familiar, que consta de unos 9.000 documentos aproximadamente, del s. XV al XX, distribuidos en 88 cajas o unidades de instalación, en el que encontramos una variada documentación de tipo notarial y judicial, así como de tipo contable. En él se encuentran todo tipo de inventarios, tasaciones, almonedas, escrituras de dote, de poder, testamentos, documentación judicial, así como todo ese tipo de material de naturaleza económica: facturas, recibos, libros de cuentas, cartas de pago, etc. Se conserva además abundante correspondencia. Al uso de estas fuentes primarias, se le debe sumar la consulta de la bibliografía pertinente para poder conocer exhaustivamente los fondos, la familia y para aplicar un adecuado sistema de organización de manera científica.

El FDPF, que podemos llamar un fondo de fondos, recibe la denominación del último generador-propietario-productor del mismo, heredero y ostentador del apellido principal de la Casa familiar troncal: Carlos Belmonte Chico de Guzmán (1862-1899).

2 Definiciones y conceptos empleados

De manera general, con la definición del concepto de Fondo Documental Patrimonial Familiar (FDPF) lo que se propone es contemplar el acervo existente en bibliotecas y archivos bajo una perspectiva en la que su condición de patrimonio cultural como bien cultural sea transversal a su definición y tratamiento documental, y desvele una serie de cualidades y características que permitan una visión y gestión de conjunto. Se ofrece de este modo una idea y una denominación con la que poder aunar la diversidad de documentos y fondos –bibliográficos y archivísticos– que puede abarcar un conjunto documental de esas características: privado, familiar, histórico y patrimonial. Por ello, a cada pieza documental que compone e integra ese FDPF se verá designada, de manera indistinta, como “objeto documental patrimonial”.

La expresión “Fondo Documental Patrimonial Familiar” deriva de la unión o nexo de varios conceptos y definiciones que se complementan.

Se emplea el término de “fondo” por encima del de colección –entendido como conjunto documental creado artificialmente con unos fines específicos–, con una acepción fundamental y vertebradora que lo define como un “conjunto documental que llega hasta el presente y se caracteriza por poseer un origen común que lo dota de sentido, por lo que debe mantenerse reunido” (Pedraza, 2014, p. 46). Al hablar de ese origen común nos referimos, a su generador, propietario, usuario y productor: la familia, que se entiende como una entidad, de carácter orgánico y funcional, que de manera natural vincula e integra a los diferentes miembros que son conocidos como los generadores-propietarios-usuarios de los objetos patrimoniales materiales, documentales y no documentales.

Esta definición de “fondo” se complementa con las acepciones dadas por otros autores al disertar sobre “fondo antiguo”, entendido como un “conjunto de objetos bibliográficos y documentales que atesora la riqueza, la esencia y el espíritu de la cultura, cuyos caracteres esenciales son el valor histórico, el estético y el valor del conocimiento” (Mora, 2003, p. 202).

Con “documental” se alude, tal y como indica su acepción adjetivada, a lo “perteneciente o relativo a los documentos”, entendiendo el documento con su significado más genérico y amplio: cualquier tipo de escrito, en cualquier soporte, de cualquier época, que testimonia un hecho o informa de él. En concreto, se acude a la definición de “patrimonio documental” en su sentido más cultural y antropológico que se explica como: “las expresiones artísticas, históricas, culturales, folklóricas, educativas, intelectuales, científicas, entre otras, que han sido producidas para ser testigo fiel del desarrollo de las sociedades; y que han sido objetivadas en manuscritos, impresos, medios audiovisuales, documentos electrónicos, etc., cuya finalidad es almacenar,

transmitir, preservar, conservar, comunicar y difundir la suma de conocimientos que contienen” (Palma, 2013, p. 34).

Para alcanzar esa consideración patrimonial como bien cultural debe darse una condición clave: pertenecer a esa parte del legado que debe ser salvado y protegido. De manera general, internacional y nacionalmente, están dadas las bases para la definición, conceptualización y regulación del tratamiento que debe recibir el patrimonio documental como bien cultural, aunque estas sean consideradas insuficientes o “en clara desventaja y atención” respecto a las dadas a otro tipo de bienes culturales. Por todo ello, este trabajo parte del marco distintivo que se hace en España del patrimonio bibliográfico y documental (no bibliográfico), tanto a nivel legislativo y teórico como práctico, cuyas definiciones, acciones y actuaciones a realizar aparecen claramente definidas en la *Ley de Patrimonio Histórico Español* (Ley 16/1985), cuyos fines últimos patrimoniales están también claramente delimitados: conservarse y difundirse.

Para finalizar, y retomando lo apuntado unas líneas más arriba, al referirnos a ese origen común se está haciendo alusión al generador, propietario, usuario y productor: a la familia, de ahí su acepción última de “Familiar”. Con ella nos referimos al concepto más amplio de familia posible, al de “familia de familias” o mejor dicho, a un linaje, formado por varias familias, casas, ramas, generaciones... que acaban constituyendo una misma memoria, una unidad identitaria que a lo largo del tiempo se ha ido definiendo por medio de la herencia y la condición genealógica. Los linajes o familias entendidos como una entidad, constituyeron en la España del Antiguo Régimen el eje de la vida social, política y economía, que mantuvieron toda su influencia y razón de ser hasta bien entrado el siglo XIX.

De esta manera, cuando hablamos de un FDPF, nos hallamos, de manera unísona y cohesionada, ante lo que se puede denominar un “fondo de fondos”: una biblioteca o fondo bibliográfico (que se ha denominado Biblioteca Patrimonial Familiar Belmonte-Chico de Guzmán, BPF-BCH) y un archivo o fondo archivístico (que se ha denominado Archivo Patrimonial Familiar Belmonte-Chico de Guzmán, APF-BCH). Ambos son el reflejo del *status*, condición, funciones y actividades de sus diferentes generadores-propietarios en un momento determinado. Sin olvidar, que en numerosas ocasiones estos fondos los encontramos vinculados a una gran cantidad de objetos materiales relacionados con esas mismas actividades y sus generadores. Y no solo eso, pues también están conectados con el espacio físico que los alberga, el inmueble en el que han sido encontrados y han permanecido custodiados durante generaciones.

Según lo expuesto y contemplado como punto de partida, se propone definir un sistema de gestión para la documentación patrimonial familiar, que se derive por igual tanto de sus funcionalidades como patrimonio o bien cultural, como por la esencia misma de los documentos que lo integran y del tratamiento aplicado para procesarlos. De tal manera que los retos de la práctica patrimonial se integren con los de la práctica documental, gestionando el legado científica y adecuadamente, aunando las funciones tradicionales de investigación y preservación, hacia una difusión más mayoritaria, incluso –si cabe– como agente social.

3 La gestión del FDPF Belmonte-Chico de Guzmán

Se quiere dar a conocer la experiencia de gestión llevada a cabo en este fondo, contemplando una forma de organización documental diferenciada, interrelacionada y de conjunto, que tenga cabida bajo una perspectiva de gestión también como patrimonio cultural. Se presenta el trabajo realizado, sus características, los problemas o límites encontrados, las medidas adoptadas, etc. así como se ofrecen las reflexiones teóricas alcanzadas.

3.1 Bases teóricas y ejecución práctica del trabajo de campo realizado

El cometido que se ha llevado a cabo se traduce en la aplicación de un eficaz, pertinente y adecuado tratamiento documental a este FDPF teniendo en cuenta que estamos hablando de fondos diferenciados y complementarios (Sánchez, 2018a). Para llevar a cabo dicho proceso se ha establecido y definido una secuencia de tareas de tratamiento documental, aplicadas al fondo bibliográfico y al archivístico, permitiendo un tratamiento específico a cada naturaleza documental.

La ejecución de un tratamiento documental, se traduce en la aplicación de las diferentes tareas que conlleva el proceso de organización, expresado en su sentido más amplio. Hablamos de las tareas fundamentales de clasificación, ordenación y descripción. Para ello, se ha desarrollado un trabajo de campo in situ en el lugar en el que está custodiada la documentación, en la casa solariega dieciochesca propiedad de la familia, y “objeto documental en mano”. Se ha establecido una secuencia de tareas de tratamiento documental, concretas y diferenciadas según se tratase el fondo bibliográfico o el archivístico, tal y como mandan las ciencias pertinentes: la biblioteconómica o bibliotecológica y la archivística. Para ello, se han elaborado las herramientas de control, clasificación y descripción pertinentes y necesarias para cada caso y fase de trabajo. Sin olvidar la elaboración de todas aquellas otras herramientas de apoyo que permiten un conocimiento detallado y exhaustivo del fondo y de su generador-propietario-productor.

La dualidad confrontada del archivo y biblioteca se muestran de una manera determinante a la hora de realizar estas tareas.

El archivo requiere de una organización programada y jerarquizada, podemos decir piramidal, donde se debe establecer un orden indispensable de prioridades y una secuencia estructurada de realización de tareas y herramientas. En el caso de las bibliotecas, la organización es lineal, podemos decir incluso horizontal, se establece documento a documento, libro a libro,

donde las tareas se pueden realizar de manera paralela y no necesariamente de manera secuencial, aunque siempre es lo recomendable por evitar la reiteración de trabajo y esfuerzos.

Debemos además tener en cuenta que estamos ante dos realidades muy específicas y concretas dentro de cada ámbito, el archivístico y el bibliotecario, que definen e influyen claramente a la hora de determinar su forma de organización y tratamiento documental.

Como medida previa a cualquier tipo de organización o descripción, se han llevado a cabo las tareas de limpieza pertinentes, y se han aplicado las medidas cautelares necesarias para la correcta preservación de cada una de las piezas u objetos documentales patrimoniales que conforman el FDPF.

El estado inicial del APF-BCH se ajustaba a la forma en la que se suelen encontrar este tipo de fondos: descuidado, desorganizado y muy fragmentado, pese a averiguar que había poseído una ordenación anterior decimonónica de carácter territorial y por tipología documental, del que quedan indicios y se han localizado inventarios previos. El tratamiento dado se ha basado en aplicar uno de los principios fundamentales que deben tenerse en cuenta para abordar cualquier archivo: separar e identificar los diferentes fondos existentes, en este caso de carácter familiar, manteniendo así el principio de procedencia. El orden de realización de tareas y las herramientas que se han elaborado se han implementado en función, fundamentalmente, del trabajo de campo inicial que debía efectuarse. La finalidad principal ha sido poder hacerse una primera idea del contenido del archivo, tipología, composición, ordenación y estado de conservación, para poder otorgarle una inicial estructura organizativa y descriptiva. Este trabajo se ha trazado elaborando un detallado inventario topográfico en Excel de cada unidad de instalación, describiendo cada unidad documental, simple o compuesta, que contenía. Se han identificado y asignado los fondos familiares principales (el troncal, los principales agregados directos, algunos indirectos e incluso, también, ajenos); y se ha establecido un inicial cuadro de clasificación, basado en el ofrecido por García Aser y Lafuente Urién (2000) para el Archivo Histórico de la Nobleza, que ha sido de gran utilidad para clasificar e identificar la variada naturaleza y tipología documental existente en el archivo, así como para poder hacer una primera valoración del mismo. En ese mismo Excel se han ido reflejando las diferentes formas o sistemas de organización original halladas previamente, para poder ser evaluadas y tenidas en cuenta. En este caso, la tarea de recomposición de expedientes no se ha podido aún llevar a cabo, -aunque siempre se han respetado y mantenido las composiciones documentales originales- debido al estudio en profundidad que exige de la documentación. Finalmente, como resultado de todo el proceso anterior, se ha elaborado una descripción completa y normalizada del archivo en su conjunto a nivel de fondo, distinguiendo los diferentes fondos familiares que lo componen o se han logrado identificar por el momento.

La BPF-BCH se define como una biblioteca histórica por el tipo de obras que alberga, impresas hasta el año 1899. Pese a poseer muchos volúmenes impresos en el siglo XIX, se caracteriza por contener un fondo bibliográfico bastante equilibrado de

libros modernos y antiguos, es decir, lo que en biblioteconomía se diferencia como fondo moderno y antiguo. El tratamiento documental aplicado se ajusta al establecido para las bibliotecas, que se materializa básicamente en la elaboración de un catálogo. Dicho catálogo incluye descripciones bibliográficas exhaustivas y normalizadas, así como pormenorizadas descripciones de los ejemplares, en las que se consignan los datos específicos y propios de cada volumen. Del mismo modo, estas descripciones han sido completadas con la aplicación de clasificaciones temáticas o de contenido de las obras a través de lenguajes documentales controlados. Al tratarse de una biblioteca con una importante parte de fondo antiguo, se han incluido, además, las referencias bibliográficas pertinentes.

Ambos fondos, el bibliográfico y el archivístico, han sido ordenados físicamente según su orden topográfico original y alojados en muebles y estanterías, situados en dos estancias diferentes de la casa. Estas estancias han sido escogidas por sus pertinentes condiciones dentro de la vivienda, y en el caso de la biblioteca, por tratarse además de la habitación original en la que se localizaron la mayoría de los volúmenes. En su inicio, el fondo bibliográfico, no estaba agrupado, sino que se hallaba disperso físicamente por diversas habitaciones de la vivienda, por lo que se tuvo que realizar una primera tarea de búsqueda y localización de volúmenes. El fondo archivístico estaba alojado, desde el principio, en una misma habitación que es en la que en la actualidad se encuentra.

Paralelo a todo este trabajo se ha realizado un estudio pormenorizado y laborioso —aunque no ha podido ser completo— de la historia de la familia-linaje, de las principales casas y ramas familiares que la componen, así como de los personajes que se integran en ellas. El objetivo ha sido tener una idea clara de la estructura general de la familia en la que poder "ubicar" familiar, geográfica e históricamente cada fondo y objeto documental patrimonial en el contexto de la historia de la familia-linaje, y así del FDPF. De ambos fondos, el bibliográfico y el archivístico, se han realizado estudios sobre el origen de su formación y analizado su tipología y características por medio de análisis cuantitativos y cualitativos.

La experiencia obtenida a la hora de llevar a cabo este "tratamiento documental dual" ha sido compleja, e incluso se pudo afirmar que desigual, aunque no desequilibrada. A ambos fondos se les ha aplicado un tratamiento documental científico, acorde a la ciencia correspondiente, que garantiza su control, localización y conservación.

Hay que tener en cuenta que el trabajo de investigación se realiza sobre la base de un marco temporal y unos objetivos concretos, por los que se han tenido que establecer prioridades y limitaciones. De tal manera que se han debido adecuar las tareas y herramientas a realizar, ajustándolas al tiempo, recursos y a las circunstancias del estudio. El resultado ha sido positivo, de

manera rigurosa y efectiva –científica–, tal y como permite cada ciencia documental, se ha logrado conocer, definir y analizar el tipo de fondo que se tiene entre manos. A nivel de materialización del trabajo ejecutado, es decir, de elaborar las herramientas de control, organización y descripción de los fondos, el fondo bibliográfico ha debido llevarse a cabo de manera completa y exhaustiva, documento a documento –no hay más opciones en biblioteconomía– y se ha materializado en la elaboración de un catálogo detallado. El fondo archivístico, dado su mayor volumen y complejidad, ha podido ser organizado y descrito en un nivel más genérico y con un carácter más provisional, adecuado, eso sí, a la ciencia archivística y a los diferentes niveles de descripción y sistematización que permite de sus fondos.

El tratamiento dado es claramente dual y diferenciador, archivos por un lado y bibliotecas por otro. Sin embargo, eso no significa que no formen un todo. Esta es una cuestión objeto de debate que ha sido planteada en el mundo de la ciencia documental desde siempre. A modo de ejemplo, proponemos lo que recoge Heredia (1995, pp. 169-170) en uno de sus trabajos, donde indica que la IFLA, en el Congreso Internacional de Archivos de 1988, presentaba ya el debate en torno a la integración o creación de instituciones específicas para las bibliotecas y los archivos (en el que incluía también a los centros de documentación). En dicha reunión se pronunciaba por la integración, abogando por la unidad de sistemas de información, especificando, eso sí, que esa integración no podía nunca significar indiferenciación ni mezcla. Se hablaba de la ubicación en un mismo lugar de tres centros diferenciados, con una separación física independiente y un tratamiento diferente, aunque la información obtenida de cada uno fuese complementaria.

Los resultados obtenidos en el tratamiento documental desarrollado y los trabajos de investigación o estudios llevados a cabo de manera paralela, es decir, el estudio de la familia y de los fondos y objetos documentales que alberga este FDPF, reflejan una complementariedad informativa y material, que hace que se pueda abogar por una definición y una gestión integrada, más si contemplamos todo este acervo documental como parte de un legado cultural que hay que conservar y difundir.

3.2 Reflexiones acerca de los elementos definitorios de un sistema de gestión integrador documental, patrimonial y cultural

Veamos cuales son los elementos y/o principios que defienden la premisa de la integración (Sánchez, 2018b) en los que se puede sostener de manera teórica y práctica una definición conceptual y una gestión de conjunto, que englobe un tratamiento documental y patrimonial cultural, que defina un FDPF.

- Se guarda un principio común y fundamental ya comentado: el uso de una metodología científica, con la que se realiza un procesamiento documental exhaustivo y específico, ya sea sobre un documento bibliográfico o uno archivístico, sobre una ciencia u otra. Así mismo, esa científicidad se aplica al estudio del generador-propietario-productor, la familia, y a las investigaciones en torno al origen y tipología de fondos, que

fundamentalmente se realizan sobre la base de estudios historiográficos.

No olvidemos, además, que la gestión del patrimonio cultural se basa en ese mismo principio fundamental. “En materia de patrimonio cultural, el conocimiento científico disciplinar nos indica la manera de proceder para identificar correctamente los objetos que pasaran a formar parte del legado material, y guiarán sobre la forma de elaborar una base de datos que permita clasificar y reunir toda la información que se pueda extraer tales bienes” (Ballart y Juan, 2001, p. 23.). Según Tugores y Planas (2006, p. 92) “investigar tiene como objeto descodificar toda la información de la que son portadores los objetos patrimoniales, tanto como piezas individuales como en su conjunto, para incrementar el conocimiento sobre ellos y las sociedades”. En este caso concreto, ese entorno social queda circunscrito, al menos inicialmente, a la familia.

La investigación se revela como un proceso esencial y transversal a cualquier intervención que ha de servir de fundamento a los niveles restantes.

- Los recursos informacionales, los objetos documentales patrimoniales, son documentos primarios y únicos: el documento de archivo lo es por su propia naturaleza y materialidad, pero el libro antiguo posee también ese carácter de *unicum*, porque en este caso no se define solamente como un impreso de una tirada determinada de una edición concreta, sino porque cada volumen posee unas características propias de ejemplar. Esas singularidades son el reflejo de un devenir por las que debe ser tratado y considerado como un documento “sin igual”. Por eso podemos hablar de manera unificada como objetos documentales patrimoniales.

Esos documentos primarios y únicos, forman un conjunto, un *unica*, agrupados por su tipología específica –por el que elevan su valor patrimonial e informativo–, que a su vez, reunidos por su pertenencia a un FDPF, conforman un *unica* superior que los dota de completo sentido y significado. De esta manera, nos hallamos ante un conjunto singular, exclusivo, inigualable, que hace referencia a unas personas y a unas circunstancias –tiempo y espacio– irrepetibles, pues solo esos documentos conservan la memoria y la identidad de una familia-linaje en un lugar y momento determinado.

- La visión de conjunto y el concepto de fondo son los axiomas indiscutibles, de tal manera que los objetos documentales patrimoniales deben ser estudiados, siempre y en última instancia, de manera conjunta, como parte de un todo indisoluble para poder ser procesados adecuadamente, por lo que deben mantenerse reunidos.

Tanto el documento de archivo como el libro antiguo (fondo antiguo), no son vistos –aunque así lo sean también por su condición de *unicum*– como piezas aisladas sino que requieren del conocimiento del contexto, grupo o

conjunto del que son originarios y en el que se integran. Según afirma Pedraza (2014) “esta particularidad es inherente a los elementos que constituyen el patrimonio cultural, cuyo estudio precisa, para su perfecta comprensión, el conocimiento del entorno para el que se concibieron, se seleccionaron o en el que alcanzaron su utilidad” (p. 41).

Estos fondos, derivan y están conectados por su origen común, por el “context” en el que fueron generados-seleccionados-utilizados: la familia. En este sentido, conocer, identificar, localizar a la familia generadora se desvela tan fundamental como el trabajo realizado de tratamiento de los fondos documentales. Es más, identificar al propietario-productor es la pieza clave que debe ser conocida en el caso de la teoría archivística, pues la delimitación y distinción de fondos es primordial para mantener uno de los principios fundamentales de dicha ciencia: el principio de procedencia. En el caso del fondo bibliográfico antiguo, conocer al “antiguo poseedor” no es fundamental para darle al fondo un adecuado tratamiento organizativo. Es una tarea, incluso, podemos decir, ajena al proceso de organización bibliográfica. Se concibe como una tarea documental añadida –a priori, no olvidemos, no organizativa– que consiste en la descripción de las características del ejemplar en concreto. Sin embargo, en el caso de los fondos antiguos, si se trata de un cometido trascendental e igual de sustancial que la descripción bibliográfica, pues conocer a los antiguos poseedores se revela como un factor fundamental de la trayectoria vital del libro, es lo que le da –la mayoría de las veces– esa condición de singularidad, de *unicum*, además, obviamente, de por su carácter histórico. Es más, a su vez, es lo que desvela en muchas ocasiones su pertenencia a un conjunto que lo dota de todo su sentido y significado. Podemos decir, incluso, a varios conjuntos: a una biblioteca familiar y, a su vez, a un FDPF.

Bajo el paraguas conceptual de FDPF la contextualización familiar, la identificación de las familias, se hace clave para los dos fondos documentales. Ambos tienen cabida y cobran significado dentro de la estructura genealógica familiar que los alberga y genera. Es más, solo conociendo ese origen familiar se comprende no solo el fondo archivístico, sino también el fondo bibliográfico, que adquiere un sentido completo como bien patrimonial cultural, como reflejo de una sociedad en un momento histórico determinado, que debe conservarse –valorarse– reunido y darse a conocer en su conjunto.

Ofreciendo solo un esbozo genérico del linaje con el que se trabaja y de las casas, familias y ramas que lo componen, hablamos de la historia del linaje de la casa troncal Belmonte, originaria de la localidad conquense que lleva ese mismo nombre, que a lo largo de casi seis siglos (s. XV-XX) emparentó con numerosas casas y familias de localidades cercanas. De momento, se han identificado casi una veintena de casas o familias, entre la troncal y las agregadas de manera directa e indirecta. Del mismo modo se han localizado dos fondos ajenos de personajes y familias no vinculadas familiarmente a la casa Belmonte sino por cuestiones de oficio o profesión. En el archivo se halla –lógicamente en distinto grado o importancia– documentación de todas las casas o familias del FDPF, y en la biblioteca encontramos

que la mayoría de los libros pertenecieron o fueron usados por miembros de las más recientes ramas o generaciones de la Casa troncal Belmonte (siglos XVIII y XIX), así como de alguna familia principal o secundaria, y algún que otro personaje ajeno a la familia.

Se adelantaba que la máxima potencialidad informativa de los documentos, se extrae junto al estudio y tratamiento de conjunto como fondo, asociado a su valor patrimonial cultural, que solo se desvela mediante la investigación. Y esto se debe a que estos fondos están conectados por la complementariedad de información que contienen, por la propia documentación que albergan y por la ubicación física en los que los encontramos en su génesis.

De esta manera, y siguiendo el hilo discursivo trazado, se plantea que la realización de estudios-investigaciones paralelos sobre la familia y los fondos, el archivo y la biblioteca (origen y tipología), deben ser parte indispensable del trabajo de tratamiento documental y patrimonial de un FDPF, y deben considerarse, de manera transversal, en la toma de decisiones y actuaciones en torno a la gestión de ese patrimonio documental familiar.

En este sentido, por ejemplo, el estudio en profundidad del origen y la formación del fondo bibliográfico, así como de su tipología, ha ayudado, en gran medida, a localizar ese vínculo entre el libro y su propietario-usuario-lector, que de otra manera no hubiese podido ser comprendido. Al realizar un estudio-investigación de esas características se ha podido conocer qué parte del fondo bibliográfico pertenece a qué familia o rama familiar y porqué. Se ha podido averiguar y deducir cómo pudo llegar gran parte de los libros a la biblioteca e inclusive definido el lugar o valor que se daba a los libros y a los “papeles” como bien familiar. Este trabajo de investigación sobre el fondo bibliográfico se ha realizado de manera conjunta, teniendo en cuenta la propia información que proporcionaban los libros: por medio del análisis de las diferentes marcas de propiedad, sellos comerciales, del tipo de encuadernación, del análisis de su contenido, etc.; y también gracias a la documentación –información– existente en el archivo relativa a los libros o la biblioteca: teniendo en cuenta todo tipo de inventarios de bienes, almonedas, cartas, etc. que en su interior se pronunciaban sobre algún aspecto relacionado con los libros que eran de interés/uso para la familia.

De la misma manera, durante el trabajo de tratamiento documental, se abren las páginas de los libros y en su interior se halla un importante número de testigos que deben ser identificados, interpretados y contextualizados. Con los testigos (Martínez de Sousa, 2004) se hace mención a un variado conjunto de documentos gráficos –impresos y manuscritos– e iconográficos de diferente tipología y cronología. Documentos de carácter personal acumulados por un individuo, generados por diferente casuística y que dan testimonio de algo, en este caso de variadas actividades de carácter privado y profesional. Nos referimos a cartas epistolares, naipes, telegramas, propaganda publicitaria,

dibujos o escritos manuscritos, almanaques, etc. y como no, algún documento de naturaleza puramente archivística. Su localización original y paralela consignación en las descripciones del ejemplar realizadas en el catálogo han sido fundamentales. El hallazgo y análisis de este tipo de materiales ha sido de gran utilidad y han aportado un gran valor al FDPF: por un lado, se ha podido incorporar al fondo materiales –de lectura o no– de difícil localización y conservación por su carácter efímero; y, por otro lado, han ayudado a corroborar y definir ese contexto que se estudia en torno a los documentos. Esos testigos han resultado ser un vivo testimonio de los familiares que fueron los principales propietarios y formadores del fondo bibliográfico, dando una información privilegiada de esos hábitos y modo de vida de carácter más privado y particular.

El problema radica en que si partimos de procesos de tratamiento documental diferenciados, hay que plantearse y definir cómo se mantiene –y refleja– la conexión y ese carácter indisoluble de la información. Este vínculo informativo e informacional hay que establecerlo por medio de las herramientas o instrumentos de organización y descripción, tan necesarios para la difusión documental e informacional: en las guías, cuadros de clasificación, inventarios, catálogos, etc., en los que de alguna manera hay que vincular y representar esa conexión entre fondos. Siendo en este ámbito donde el desarrollo de la tecnología informática –y digital– amplía claramente las posibilidades, y acorta distancias entre la dualidad existente como punto de partida. Es un asunto que ya está presente en los debates sobre las prácticas de ambas disciplinas de las que partimos. Por ejemplo, la ciencia archivística, en la elaboración de los cuadros de clasificación deja cabida a indicar como parte del fondo propiamente documental –archivístico–, en un genérico pero siempre presente apartado de “Colecciones”, esa otra documentación hallada en los archivos, los “libros, las publicaciones periódicas y la propaganda” como documentación impresa no propia de los archivos, pero que refleja los intereses de una familia, sus actividades –su mundo, en definitiva– y por ello deben incluirse y considerarse como parte del fondo (Gifre, Matas y Soler, 2002, pp. 34-35). Es más, se afirma que es tarea del centro que los agrupa y diferencia entre biblioteca, hemeroteca y archivo, dejar testimonio de su origen común, incluyendo siempre una relación entre esos materiales complementarios en la descripción de cada fondo.

- El FDPF, como parte del patrimonio cultural se debe a un sistema de gestión que se ve limitado –e implementado– para llevar a cabo sus dos funcionalidades patrimoniales principales: la conservación o preservación de originales, para lo que se deben aplicar las necesarias y adecuadas medidas cautelares; que, a su vez, deben ser compatibles con su otro leitmotiv: la difusión. Dos funcionalidades que se retroalimentan y complementan, y que se antojan imprescindibles para la gestión de un FDPF. En palabras de Tugores y Planas (2006): “con la difusión se cumple la función social y obligada del patrimonio, pero también se asegura su conservación: aquello que es conocido, comprendido y valorado difícilmente será destruido” (p. 95).

Estamos, sin duda, ante la patrimonialización de los bienes documentales de una familia-linaje. Bajo este prisma el FDPF es visto –y tratado– como parte de un legado material e inmaterial que se debe salvaguardar, pues constituye un testimonio que contribuye a definir la memoria e identidad de una comunidad concreta, una familia, e, inclusive, podríamos afirmar, de un entorno más amplio: de una sociedad y una cultura de una región determinada.

Tradicionalmente la difusión de este tipo de fondos documentales se ha visto asociada a una misión principal: hacerla accesible para la investigación. Es decir, se trataba de una difusión de carácter exclusivo y especializado, con fines científicos, ¿pero debemos detenernos ahí? Con la idea que se ofrece de FDPF, vinculada a una gestión también como bien cultural, se quiere plantear ampliar esa visión de difusión únicamente especializada, abriendo la posibilidad de hacerla compatible y extensible a una de carácter más divulgativo. ¿Dónde o cómo tendría cabida el enfoque de este patrimonio documental como bien cultural hacia una difusión más mayoritaria, incluso –si cabe– como agente social?

Tal y como se ha apuntado en las líneas anteriores, en un FDPF la visión de conjunto prima sobre la individual, y se debe tener en cuenta tanto la forma y la ubicación en que encontramos la documentación en su génesis, como su interrelación. Siguiendo ese enfoque, se quiere dar un paso más y poner en relación al documento, no solo con los otros documentos, sino con el resto de bienes muebles, que le rodean e, inclusive también, con el espacio físico que lo alberga, la vivienda en la que se localizan. Recordemos que este FDPF se encuentra en la casa original del siglo XVIII que aún posee la última generación heredera de la familia-linaje. De tal manera que los objetos documentales patrimoniales que conforman el archivo y biblioteca, se patrimonialicen junto a la casa y a los otros objetos u bienes muebles que se custodian en ella.

En este sentido, sin poder profundizar mucho en ello y salvando las distancias del caso concreto que ofrece Francesca Uccella al hablar de la gestión del patrimonio literario, abogamos por emplear la enfoque dado en su estudio y su definición de Casa-museo (2013, pp. 45-53). La autora apuesta por una visión patrimonial de conjunto para la gestión del patrimonio literario que gira en torno a las casas de los escritores: a los espacios que sirven para conservar y perpetuar la memoria, así como para mantener o recrear una situación original con el fin de acercarse a un personaje y al periodo histórico que le tocó vivir. El material documental (libros, manuscritos, cartas) son mostrados junto a otros objetos materiales gracias a su fuerza simbólica y a su innato poder narrativo. Todos ellos son presentados como pertenencias que dan a conocer de manera inequívoca la trayectoria vital de un individuo –en nuestro caso sería también de una familia– en una época concreta,

trasladándote, a menudo de manera muy real, al tiempo y a las circunstancias en las que le tocó vivir.

Bajo esta perspectiva, se propone una relación patrimonial tripartita –casa, mueble y documento– que se establece y estructura en torno, una vez más, a la familia-linaje, a los diferentes miembros que la conforman, y a la información complementaria e interrelacionada que contienen todos los objetos de la vivienda familiar, que solo se desvela mediante la investigación. Se deben realizar, por lo tanto, exhaustivos trabajos de investigación que nos permitan reconocer a las familias y a los miembros que conformaron ese linaje; que nos posibiliten, a su vez, identificar correctamente todas las piezas documentales y materiales que albergan los fondos y la casa; y que nos den la facultad de comprenderlos tanto de manera individual como en su conjunto, es decir, contextualizándolos adecuadamente y poniéndolos en relación con los otros materiales que les rodean.

De esta manera, hemos podido contemplar –a veces asombrados– como una ilustración a toda plana de un árbol genealógico de una Real una Carta ejecutoria de hidalguía del siglo XVII perteneciente a los miembros de una de las familias-fondos principales que se asociaron a la Casa troncal Belmonte –los Granero de Alarcón– ha servido como modelo –¿o fue al revés?– para realizar los retratos que amueblan el salón de baile. O como el escudo de armas de los Chico de Guzmán, que decora reposteros, vajillas y ex libris, nos dice que ese fue el apellido que les llevó a alcanzar el ansiado título de nobleza. No sin antes detenernos en los anteriores y numerosos escudos de armas que embellecen los documentos del archivo, las paredes y la fachada de la casa diciendo que desde tiempos inmemoriales ya residían en esa casa solar hidalgos de “nobleza reconocida por todos”.

3.3 Otros modelos de gestión integradores: el modelo sistémico portugués

Nos hemos querido preguntar por las formas de gestión de este tipo de casos por medio de otros modelos, propuestas o teorías. En ese sentido, nos queremos referir a la propuesta científica ideada por Armandho Malheiro da Silva y otros profesionales, que parten de situar a la archivística como una ciencia autónoma dentro del ámbito de las Ciencias de la Información y ofrecen unos planteamientos teóricos que dan como resultado la definición de un “modelo sistémico”. Infra este modelo, el archivo se define como un “sistema (semi)fechado de informação social, materializada em qualquer tipo de suporte, configurado por dois factores essenciais –a natureza orgânica (estrutura) e a natureza funcional (serviço/uso)– a que se associa um terceiro –a memória imbricado nos anteriores” (Silva, 1998, p. 214). Del mismo modo, se establece que con los archivos familiares y personales estamos ante “um Sistemas de Informação organizado ou operatôiro, cujo pólo estruturante e dinamizador é uma entidade - Família e Pessoa, cada qual com estrutura própria e

acção fixada sempre por objectivos diversos, uns perenes e outros mutáveis (...)" (Silva, 2004, p. 60).

Desarrollamos la explicación de este modelo siguiendo la reflexión y estado de la cuestión que traza Fernández (2018, p. 38) en su tesis doctoral: se pretende ofrecer un modelo aplicable a todo tipo de agrupaciones documentales, a las que considera un sistema de información con una estructura determinada y una funcionalidad dinámica, que puede ser de uso originario de los documentos como reutilización posterior. La aplicación práctica de este sistema obliga, en primer lugar, al análisis de la estructura orgánica del archivo en cuestión, en las que se pueden dar dos tipos de cuerpos: unicelulares (estructuras individuales) y pluricelulares (estructuras colectivas) ambas con diferentes sectores funcionales basados en la relación afectiva y física de individuos con continuidad genética mediante descendencia de generaciones y la acción de los miembros para garantizar la supervivencia colectiva y las estrategias de poder.

Esto se traduce y materializa en considerar a la familia como una institución y un sistema de información. La propia familia es vista como una institución que produce documentación y que constituye un cuadro orgánico, compuesto por el fundador/es y sus descendientes. Cada familia constituye un sistema que se puede dividir en subsistemas, tantos como casas o familias se hayan incorporado al linaje troncal. A su vez cada sistema o subsistema está formado por secciones identificadas por las generaciones de cada casa o familia, o por subsecciones, siendo la primera subsección el matrimonio, la segunda el marido, la tercera la esposa, la cuarta los hermanos y así sucesivamente. Cada subsección se organiza en torno a las fases de la vida (infancia, juventud y adulto) de donde se cuelgan las series y los diferentes documentos (informaciones) que se refieran al individuo a la matrimonio concreto.

No se pretende entrar a profundizar en el grado de aceptación que genera la aplicación del modelo sistémico como forma de organización de los archivos familiares, que, sin duda, daría para un artículo aparte. Lo que más interesa es analizar y reflexionar sobre la perspectiva que se ofrece, bajo ese modelo sistémico informacional, de contemplar una concepción más radical e integradora para la gestión de los Archivos, las Bibliotecas y los Museos. Concepción que tiene cabida en el caso de los archivos familiares o personales (lo que podríamos diferenciar entre Fondos Documentales Patrimoniales Familiares y Personales), que se ven acompañados muchas veces por una biblioteca, varios objetos materiales y que se gestionan –en ocasiones– en torno a un palacio o paço.

En concreto, queremos proferir del proyecto llevado a cabo en la Casa de Mateus con el que Armando Malherido da Silva (2007, pp. 305-317) puso a prueba las bases y los objetivos de ese modelo sistémico. En él se desarrolló un

modelo basado en la reconstrucción del contexto orgánico-funcional y temporal de la documentación, es decir, la familia, estudiando las diferentes generaciones y los miembros de la misma; y estableciendo una clasificación de la documentación (de tipo personal o profesional y administrativa) desde la óptica del productor/receptor/utilizador de la información. Se aplicó el concepto operatorio de familia como un sistema de información y se articuló un trabajo de gestión en torno a la “información”, más que a un tipo de documento u otro en concreto (de archivo, biblioteca o museo). De este modo, la concepción orgánico-funcional del archivo exigió que paralelamente al trabajo de organización y descripción de la documentación se hiciesen las pertinentes tareas de investigación, históricas y genealógicas, para reconstruir la estructura orgánico-funcional de la familia y con ello el origen de todos los conjuntos documentales-informacionales del archivo. El sistema aplicado al archivo permitió dar una perspectiva integradora de la información depositada no solo en el archivo, sino también a la biblioteca y al museo. De esta manera, se procuró asociar cada pieza del museo y de la biblioteca al cuadro orgánico-funcional genealógico familiar establecido en el archivo. Y esta tarea se pudo llevar a término, principalmente, gracias a la propia documentación del archivo, a la información contenida en él (en inventarios, documentos de pagos, etc.) que permitieron identificar el origen (productor /receptor /utilizador) de muchas de las piezas.

Con este modelo se deseaba encontrar una solución para lo que, en su día, el propio Armando Malheiro da Silva denominó como un Sistema patrimonial complejo (termino que luego desestimó, prefiriendo emplear el de Sistemas de Información). Con este concepto quería referirse a la dificultad y necesidad de establecer una conexión entre los diferentes fondos existentes en archivos, bibliotecas y muesos, en la importancia de mantenerlos reunidos e interrelacionados. La base inicial y principal de su propuesta o perspectiva defiende la importancia de estudiar el contexto de la documentación. Dicho de otra manera, de llevar a cabo un estudio y un conocimiento exhaustivo del su contexto original e histórico, o sea, de la familia, en su vertiente más orgánica y funcional como base estructural para la sistematización de los fondos documentales y materiales. De manera general, este enfoque apuesta por la relevancia de extraer la información de la documentación, pero mucho más por interpretarla y establecer las relaciones de información pertinentes. El problema principal de esta propuesta es el tiempo que conlleva el estudio pormenorizado de los fondos, y de la familia/s, conocer cada matrimonio en concreto, cada trayectoria vital, establecer relaciones, identificaciones, etc.

De esta breve presentación y síntesis del modelo sistémico portugués se extraen algunos aspectos claves que hacen reflexionar sobre las posibilidades de hallar puntos de encuentro entre el planteamiento luso y las bases propuestas para desarrollar un FDPF:

- La necesidad primordial del estudio del contexto de la documentación, es decir, de la familia como base del trabajo, que exige exhaustivos y laboriosos trabajos de investigación.

- La primacía de la visión orgánico-funcional, podemos decir incluso de carácter genealogista, en torno a la familia como entidad para definir la estructura y organización del archivo familiar.
- El hecho de que de alguna manera, los trabajos y planteamientos para darle una estructura al archivo o fondo archivístico, se establecen como fundamento para la gestión y concepción estructural de todo el conjunto de bienes patrimoniales familiares. El trabajo en torno al archivo es el hilo conductor y transversal para realizar una propuesta organizativa-informativa mayor, que agrupa, a los libros, a los objetos materiales y al inmueble. La familia, desde una perspectiva genealógica, es lo que ha permitido establecer los lazos entre las piezas del archivo, la biblioteca y del museo.
- La importancia dada a la vinculación e interrelación de la información complementaria que contienen todos los bienes –documentales o no– familiares.

Sin duda, esta es una perspectiva que guarda muchas similitudes con caso que presentamos del FDPF Belmonte-Chico de Guzmán. A través de conocer e identificar a las familias, las distintas generaciones familiares y/o los miembros que la integran, podemos ubicar y estructurar los documentos del APF-BCH, según los procedimientos y principios ya explicados. Esa contextualización es extensible a los libros de la BPF-BCH –identificando a los antiguos poseedores–, y además, permite dotar al conjunto bibliográfico de un sentido-visión de conjunto imposible de establecer por otro tipo de vía organizativa estructural –obviando la histórica-cronológica– que guarda además relación con la temática y funcionalidad que posee esa “biblioteca de uso” para las últimas generaciones de la familia-linaje. Es más, tal y como se ha ejemplificado anteriormente, esa perspectiva ha permitido también “situar familiarmente” varios objetos muebles e, inclusive, el inmueble.

Todos los objetos patrimoniales familiares del FDPF Belmonte-Chico de Guzmán, documentales o no, conllevan el nombre y el apellido de una persona o una generación, que solo se desvelan mediante trabajos de investigación. Si estos logran efectuarse con éxito, llevando a cabo óptimos trabajos de identificación, los objetos patrimoniales familiares son ubicados dentro de una estructura familiar, que les aporta una visión de conjunto única, inigualable e irrepetible. Encontrarlos reunidos de manera original es un privilegio informativo, científico y cultural, pues lo habitual es que este tipo de patrimonio se encuentre siempre fragmentado y disgregado.

4 A modo de conclusión

La concepción de un FDPF como propuesta de gestión integrada documental y patrimonial- cultural, obedece a la necesidad de gestión de un caso real, y por

ello se intentan establecer una bases teóricas y prácticas que de algún modo den respuesta todas esas necesidades.

El establecimiento de modelos organizativos propios y diferenciados por fondos tipológicos documentales –bibliográficos y archivísticos– afines y respetuosos a las ciencias y a los principios que los rigen no es incompatible con el establecimiento de una (macro)estructura familiar de conjunto, de carácter más genealogista, que lo dote de sentido y significado como patrimonio documental familiar y cultural de carácter único y singular.

Si contemplamos integrar en esa macro-estructura y concepción patrimonial cultural de los fondos documentales a los distintos bienes materiales-muebles, así como el inmueble, nos estamos acercando a concepciones y prácticas museísticas que no solo no tienen por qué dejar de contemplarse, sino que pueden dar respuesta a las dos funciones primordiales que garantizan la salvaguarda de estos fondos: conservarse y difundirse.

La concepción de FDPF se aproxima –podemos decir casi de manera natural– a la teoría o modelo sistémico luso orgánico funcional de carácter integrador, que de algún modo subyace a la hora de idear y defender el conjunto patrimonial de la Familia Belmonte-Chico de Guzmán.

En España no existe un modelo teórico integrador como el sistémico portugués y sería interesante iniciar un debate en ese sentido, que diese respuesta a lo que es ya una realidad y representa una práctica habitual en muchas instituciones de gestión documental y patrimonial, tanto de carácter público como privado. Pues tal y como se ha indicado al inicio de estas las conclusiones, todo este trabajo, teórico y práctico, obedece a una necesidad real, que no es otra que la de mantener vivo, a salvo y reunido este patrimonio familiar, que se ha tenido el privilegio de encontrar reunido de forma original y excepcional.

El archivo es, sin duda, el corazón del patrimonio familiar, es el que conserva la memoria e identidad de la familia. Se entiende que de él se proyecte toda una gestión y concepción patrimonial de conjunto. Y todo ese patrimonio se puede convertir en un valor que traspare la representatividad o interés familiar, y pasar a tener –por qué no–, un papel en el simbolismo de una identidad regional del que sentirse orgullosos y partícipes

Referencias

- Ballart, J. y Juan, J. (2001). *Gestión del patrimonio cultural*. Barcelona: Ariel.
- Fernández Valdivieso, José Luis (2018). *Estudio, organización y descripción del Archivo de los Marqueses de Corvera*. Tesis doctoral. Universidad de Granada.
- García Aser, Rosario y Lafuente Urién, Aránzazu (2000). *Archivos nobiliarios: cuadro de clasificación: sección nobleza del Archivo Histórico Nacional*. Madrid: Ministerio de Educación, Cultura y Deporte, Centro de Publicaciones.

- Gifre, Pere, Matas, Josep y Soler, Santi (2002). *Els arxius patrimonials*. Girona: Associació d'història rural de les comarques Gironines.
- Heredia Herrera, Antonia (1995). *Archivística general: teoría y práctica*. Sevilla: Diputación de Sevilla. Publicaciones de la Diputación Provincial de Sevilla.
- Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español (B.O.E. n. 155 de 29 de junio de 1985). <http://noticias.juridicas.com/base_datos/Admin/l16-1985.html>. [Consulta: 31/05/2019].
- Martínez de Sousa, José (2004). *Diccionario de Bibliología y ciencias afines*. Gijón: Trea.
- Pedraza Gracia, Manuel José (2014). Algunas reflexiones sobre bibliotecas históricas o patrimoniales: nuevo paradigma entre los centros y servicios de información. *Investigación bibliotecológica*, 64(28), 33-50.
- Núñez Fernández, Eduardo (1999). *Organización y gestión de archivos*. Gijón: Ediciones Trea.
- Palma Peña, Juan Miguel (2013). Valores sociales y valores patrimoniales: elementos para determinar la significación del patrimonio documental. *Biblioteca Universitaria: Revista de la Dirección de Bibliotecas de la UNAM*. 16(1), 33-45.
- Sánchez Oliveira, Camino (2018a). En busca de un pacto patrimonial: el Fondo Documental Patrimonial Familiar (FDPF) Belmonte-Chico de Guzmán. En *Actas de las II Jornadas de estudiantes de Ciencias de la Documentación: "tendiendo puentes"*: (pp. 36-43). Madrid: Facultad de Ciencias de la Documentación. Universidad Complutense de Madrid,
- Sánchez Oliveira, Camino (2018b). Propuesta de un sistema de gestión integrado para fondos documentales patrimoniales familiares. *Ibersid: revista de sistemas de información y documentación*, 12(1), 69-76.
- Silva, Armando B. Malheiro da (et. al) (1998). *Arquivística: teoria e prática de uma ciência da informação*, vol. 1. Port: Afrontamento.
- Silva, Armando Barreiros Malheiro da (2015). Arquivo, biblioteca, museu, sistema de informação: em busca da clarificação possível... *Cadernos BAD*, 1, 103-124.
- Silva, Armando Barreiros Malheiro da (2004). Arquivos familiares e pessoais: bases científicas par ala aplicação do modelo sistémico e interactivo. *Revista FLUP: ciências e técnicas do património*, 1(3), 55-84.
- Silva, Armando Barreiros Malheiro da (2007). Da memória ao acesso à Informação na Casa de Mateus: as bases e objectivos de um projecto sistémico. *Revista de Letras*, 2(6), 305-317.

- Tugores Truyol, Francesca y Planas Ferrer, Rosa (2006). *Introducción al Patrimonio cultural*. Gijón: Trea.
- Uccella, Francesca R. (2013). *Manual de patrimonio literario: espacios, casas-museo y rutas*. Madrid: Trea.

Organização do Conhecimento e Identificação Tipológica: Aportes para Projetos de Pesquisa em Arquivos Pessoais

Lucia Maria Velloso de Oliveira¹, Camilla Campoi de Sobral², Bianca Panisset³ e José Antonio da Silva⁴

¹ ORCID [0000-0003-0546-2602](https://orcid.org/0000-0003-0546-2602). Universidade Federal Fluminense e Casa Rui Barbosa, Rio de Janeiro, RJ, Brasil.
luciemarie@gmail.com

² ORCID [0000-0002-6725-2929](https://orcid.org/0000-0002-6725-2929). Universidade Federal Fluminense, Brasil, Rio de Janeiro, RJ, Brasil.
milla.campoi@gmail.com

³ ORCID [0000-0002-5568-5225](https://orcid.org/0000-0002-5568-5225). Universidade Federal Fluminense e Casa Rui Barbosa, Rio de Janeiro, RJ, Brasil.
biancapanisset@gmail.com

⁴ ORCID [0000-0002-7309-2319](https://orcid.org/0000-0002-7309-2319). Universidade Federal Fluminense e Casa Rui Barbosa, Rio de Janeiro, RJ, Brasil.
professor.jose.antonio@rb.gov.br

Resumo: O presente trabalho busca apresentar o desenvolvimento e os resultados parciais dos projetos de pesquisa sobre tipologia documental desenvolvidos na Fundação Casa de Rui Barbosa. O enfoque de nossas análises pautam-se na convergência da identificação do tipo documental como forma de representação das ações entre o indivíduo e a sociedade, e os conceitos elaborados no escopo da pesquisa como equivalentes das características e contextos que um documento pode representar.

Palavras-chave: Arquivo de pessoa. Representação Arquivística. Teoria do Conceito. Tipologia documental.

Abstract: This paper presents the development and partial results of research projects on documentary typology developed at the Fundação Casa de Rui Barbosa. The focus of our analyzes is based on the convergence of the identification of the documentary type as a form of representation of actions between the individual and society, and the concepts elaborated in the scope of the research as equivalents of the characteristics and contexts that a document can represent.

Keywords: Personal archives. Archival representation. Concept Theory. Documentary typology.

1 Introdução

Nos últimos vinte anos no Brasil, a produção de conhecimento sobre os arquivos de pessoas tem obtido destaque na agenda de pesquisas no campo da Arquivologia. Os estudos e pesquisas produzidos nos últimos anos têm buscado direcionar a discussão sobre esses conjuntos documentais para além da concepção da literatura, da construção do discurso ou da memória de si, que emplacaram ideias polêmicas na perspectiva do núcleo duro da Arquivologia. Os trabalhos recentes discutem abordagens teóricas e metodológicas pertinentes aos conjuntos de natureza pessoal, fundamentando as pesquisas no corpus teórico dos arquivos. Como fica evidente, o processo de reconhecimento desses conjuntos como objeto de estudo e interesse da Arquivologia historicamente não foi linear e a sua organização nem sempre respeitou os princípios teóricos da área. Dessa maneira, a pluralidade de possibilidades dos documentos produzidos e acumulados por indivíduos no ato de viver nem sempre foi entendida pelos profissionais da área como passível de reflexões e abordagens arquivísticas.

Nesse sentido, o reconhecimento dos arquivos de pessoas como patrimônio cultural e campo de interesse de diferentes usuários proporcionou a valorização desses conjuntos frente à teoria arquivística, e, com isso, incentivou a produção teórica e metodológica de conhecimento sobre arquivos produzidos por indivíduos.

Produzidos no âmbito da vida privada, os arquivos pessoais são compostos por conjuntos de documentos criados e mantidos por um indivíduo para instrumentalizar sua vida em sociedade e representam os relacionamentos, afetos, desafetos, expressões de sentimentos, acordos e compromissos assumidos pelo produtor do arquivo como sujeito social (Oliveira, 2016). Além de abrigar as muitas personas do produtor, um arquivo pessoal é produto dos contextos que permeiam a gênese documental. As nuances entre o público, o social e a intimidade são aspectos presentes nas cartas produzidas e recebidas por indivíduos no decorrer de sua trajetória e representam um desafio aos arquivistas no que concerne à representação destes documentos no arranjo e na descrição arquivística.

As particularidades do “pessoal” frente ao crescimento de estudos e metodologias de organização sobre arquivos pessoais permitiram que desde 2010 fossem desenvolvidas pesquisas nos arquivos privados custodiados pelo Serviço de Arquivo Histórico e Institucional da Fundação Casa de Rui Barbosa, onde foram realizados projetos de pesquisa de análise tipológica em arquivos pessoais. O projeto de pesquisa Análise tipológica dos documentos em arquivos pessoais: uma representação do código social foi o primeiro dentre os três desenvolvidos pela instituição na temática deste tipo de análise e, finalizado em 2012, identificou 41 tipos documentais em quatro grandes eventos da vida humana: nascimento, casamento, grandes viagens e

morte. Diante da necessidade de ampliar o campo experimental do projeto inicial, em 2013, deu-se início ao projeto de pesquisa Tipologia documental na família Barbosa de Oliveira, coordenado pela Dra. Lucia Maria Velloso de Oliveira, e que tem por objetivo principal identificar dos tipos documentais produzidos na esfera privada, além de analisar os costumes dos séculos XVIII e XIX. Articulado ao projeto base mencionado anteriormente, um segundo projeto foi concebido para conceituar os tipos documentais identificados no primeiro e produzir um glossário de termos. O projeto Tipologia documental na Família Barbosa de Oliveira: definição dos tipos documentais está intrinsecamente relacionado ao primeiro e ambos visam contribuir para o conhecimento das ações e atividades que dão origem ao documento em si no âmbito privado (Oliveira; Sobral, 2017).

Os projetos de pesquisa em desenvolvimento almejam contribuir para a representação dos documentos em arquivos pessoais. Para Geoffrey Yeo, “os documentos são a representação persistente das atividades, são produzidos pelos participantes ou observadores dessas atividades ou pelos seus representantes autorizados” (Yeo, 2007, p.337, tradução nossa). Segundo esse autor, uma representação persistente significa que os documentos persistem para as circunstâncias que vão além daquelas que motivaram a sua produção. Por exemplo, uma certidão de batismo: o motivo imediato da sua produção visa registrar, comprovar que uma criança ou um adulto foi batizado. Representa a iniciação de um indivíduo na igreja católica, para além do motivo imediato de produção desse documento a partir da compreensão da característica de representação persistente dessa certidão de nascimento, sendo possível perceber que ele também representa um modo de viver social relacionado à prática religiosa do batizado e também de sua família.

Produto do ato de viver de um indivíduo, os documentos produzidos e acumulados são consequências de atividades e refletem as práticas sociais de um determinado período. Dessa forma, o tipo documental, quando corretamente identificado, pode oferecer ao usuário o caminho mais específico para que encontre o que procura, na medida em que o tipo traduz a relação de causa e efeito que origina o documento (Oliveira & Oliveira, 2017).

Documentos, quaisquer que sejam a ação ou atividade que os originaram, são produzidos em um contexto específico. O lugar social que um sujeito ocupa, as relações sustentadas, os contratos e pactos estabelecidos são elementos passíveis de identificação nestes documentos. Para Theo Thomassen (2006, p. 11), o contexto arquivístico está além do contexto de produção do documento, pois engloba os contextos sócio-político, cultural e econômico que incidem sobre a produção documental, condicionando muitas vezes sua forma e conteúdo.

À medida que o estudo do tipo documental busca, a partir de uma nomeação adequada, encontrar a ligação existente entre os documentos e os contextos que os geraram, os estudos tipológicos contribuem para a compreensão do arquivo e suas relações. Para a autora alemã Angelika Menne-Haritz, “os arquivistas são os únicos especialistas que possuem as ferramentas teóricas e metodológicas para tornar a

evidência acessível e assim revelar o contexto exploratório da informação” (Menne-Haritz, 2005, p.34). Nesse sentido, as relações contextuais são fundamentais para a compreensão dos documentos, sendo o arquivista o profissional que deve representá-los no processo de organização da informação. Mas seria simplório pensar a aplicação da tipologia documental sem explorá-la à luz da definição de conceitos que traduzam efetivamente os tipos documentais, na perspectiva que estes representam atos e ações de um produtor de arquivo.

Hjorland defende que a “teoria da Organização do Conhecimento (KO) está relacionada a uma teoria dos conceitos” (Hjorland, 2008, p.1, tradução nossa). Assim, é a partir de uma compreensão pragmática dos conceitos que podemos entendê-los como formas pelas quais os atores estão escolhendo partes da realidade para poder se comunicar e agir em relação àquela parte da realidade (Hjorland, 2008, p.1). Nesse sentido, reconhecemos que a nomeação adequada dos documentos representa um esforço significativo ao representar em um termo as características físicas e a atividade que o gerou em um determinado contexto.

Destacamos que a identificação do tipo documental, visando garantir aos potenciais usuários o entendimento de um espectro de um termo, necessita da conceituação e apresentação das características que um documento pode possuir. A conceituação de termos, quando apresentada em um glossário, promove um efeito semelhante ao de um vocabulário controlado, “que pode ser visto como um conjunto de técnicas e procedimentos que são aplicados à linguagem para resolver problemas de entendimento, ambiguidade e o alcance e a relação entre termos que expressam conceitos, e/ou entre as denominações que expressam nomes de pessoas, lugares, produtos ou instituições” (Barité, 2014, p.97).

Assim, tais iniciativas na representação dos arquivos de pessoas caminham na contramão das usuais abordagens no tratamento desses conjuntos. A identificação tipológica tem se mostrado um coerente método no processo de organização do conhecimento e representa um dispositivo de potencialização do acesso ao usuário, em especial se associada à teoria do conceito. A discussão da representação no âmbito dos arquivos ainda necessita de mais estudos e tem sido timidamente trabalhada nas funções de arranjo e descrição. Os projetos de tipologia desenvolvidos na Fundação Casa de Rui Barbosa reconhecem o tipo documental como representação das ações e atividades vinculadas à gênese dos documentos, sendo a conceituação dos tipos uma forma de ampliar o escopo de estudos e pesquisas sobre a temática e oferecer, aos pesquisadores e profissionais da área, uma abordagem que promova o diálogo entre teoria e organização dos documentos.

2 Objetivos

O presente trabalho busca apresentar o desenvolvimento e os resultados parciais dos projetos de pesquisa sobre tipologia documental desenvolvidos na Fundação Casa de Rui Barbosa. Para tal, focaremos nossas análises na convergência da identificação do

tipo documental como forma de representação das ações entre o indivíduo e a sociedade, e os conceitos elaborados no escopo da pesquisa como equivalentes das características e contextos que um documento pode representar.

3 Metodologia

A metodologia empregada na pesquisa pressupõe uma etapa inicial de identificação dos tipos documentais. A identificação ocorre após a equipe de pesquisadores cumprir as seguintes etapas: reconstruir o contexto arquivístico; identificar as ações e efeitos dos documentos em análise; analisar o seu formato e conteúdo; analisar comparativamente as conclusões sobre a análise dos documentos e as informações coletadas em manuais de etiqueta, secretários, legislação e demais fontes; nomear preliminarmente o tipo documental; verificar se o termo para identificar o tipo do documental já foi validado; e verificar a prática arquivística consultando bases internacionais que disponibilizam instrumentos de pesquisa de consulta a acervos arquivísticos on-line (Biblioteca Nacional dos Estados Unidos, Arquivo Nacional do Reino Unido e Arquivo Nacional da França). Por fim, ocorre a validação do termo em reunião da equipe com a consultora do projeto, uma pesquisadora de arquivos pessoais e tipos documentais da Universidade de São Paulo/Brasil (Oliveira & Oliveira, 2017).

O processo de reconstrução ou identificação do contexto arquivístico é resultado de um percurso de pesquisa que considera as funções sociais do produtor e seus diferentes papéis, assim como busca reconstruir e compreender, por meio dos documentos que compõem o arquivo, as relações sustentadas pelo produtor com o Estado, familiares, amigos e instituições no percurso de sua vida.

Em um segmento paralelo ao processo de identificação dos tipos documentais, a pesquisa investiga quantitativamente os instrumentos de pesquisas, levantando ocorrências de termos relativos às espécies e tipos documentais em bases de dados de instituições de alta credibilidade internacional: Library of Congress (EUA), National Archives (Reino Unido) e Archives Nationales (França). O levantamento realizado nas bases de dados internacionais atua como método balizador de termos e ocorrências utilizados na área e as pesquisas sobre o tipo documental realizadas no projeto. (Oliveira & Sobral, 2017).

O processo de conceituação dos tipos documentais identificados e validados busca contemplar, na elaboração dos verbetes, por meio de estudos e pesquisas, as práticas sociais e a legislação da época da produção do documento e de que forma esses dispositivos incidem sobre a gênese documental, determinando possíveis elementos de forma e conteúdo.

Conforme destaca Hjørland (2008), definir conceitos na organização do conhecimento considera as implicações das diferentes formas de entendimento. Nesse sentido, conforme exemplifica o autor, o conceito de “hino”, por exemplo, está

relacionado à atividade de “cantar em igrejas” e isso o distingue de outras canções não entendidas como hinos e, portanto, não incluídas no livro de hinos.

A conceituação dos termos identificados no projeto é uma abordagem inovadora no âmbito da organização dos arquivos pessoais. O diálogo entre análise tipológica e a conceituação considera outras áreas de conhecimento, como a história social, política e cultural, na elaboração dos verbetes e, com isso, contribui para um relacionamento interdisciplinar na construção do conhecimento sobre esses documentos produzidos no âmbito da intimidade.

É importante ressaltar que a nomeação dos documentos tem se mostrado a abordagem mais coerente para representação arquivística. Segundo Oliveira, em sua tese de doutorado, “o processo e/ou programa descritivo implica necessariamente na identificação das espécies e dos tipos documentais individualmente encontrados num arquivo e/ou coleção” (Oliveira, 2012, p. 43). Ainda para autora, a identificação tipológica é uma importante ferramenta para o estudo das atividades e ações representadas nos documentos, permitindo, também, identificar as formas de comunicação entre o titular e os demais indivíduos mencionados nos documentos.

Para a pesquisadora Maria Celina Soares de Mello e Silva, o estudo tipológico em arquivos pessoais “representa um desafio para arquivistas, que, ainda, encontram poucas iniciativas para servirem de exemplos, mas que têm um futuro promissor pela frente, já que sua utilidade no tratamento das informações dos arquivos tem sido cada vez mais reconhecida” (Silva, 2013, p. 166).

Portanto, o ato de nomear documentos a partir da sua espécie e atividade geradora nos arquivos pessoais pode surpreender pela liberdade de produção de muitos documentos que dizem respeito à intimidade. Para além, caracterizar a natureza de algo é uma atividade complexa que visa identificar o nome das atividades representadas nos documentos. Isso nos remete a discussão proposta por Yeo ao definir documentos como representações persistentes de atividades.

E isso não significa persistência sem limite de tempo, mas que “documentos são persistentes no sentido que eles perduram além do término temporal das atividades que representam” (YEO, 2007, p. 338, tradução nossa). Uma vez que, uma atividade pode ocorrer em um período específico de tempo ou pode persistir por um longo período até o seu término, mas, independentemente dessa permanência, o documento como representação de dada atividade persiste. Mais do que sintetizar em uma fórmula nuclear a espécie documental acoplada da atividade que representa, nomear documentos nos permite conhecer a forma como diferentes atividades são representadas por meio deles.

Para Camargo (2015), a nomeação adequada dos documentos em relação à organização e descrição dos arquivos tem sido bastante negligenciada, em um cenário no qual os profissionais, entendem que, sendo todos da mesma espécie, basta reconhecê-los pela função ou assunto de que tratam. Na contramão desse cenário, “os estudos e usos da tipologia documental em arquivos pessoais representam esse esforço significativo de conferir ao documento o nome mais adequado de acordo com

as características físicas e a atividade que o gerou dentro de um determinado contexto” (Oliveira & Sobral, 2017, p. 1013).

A tipologia documental “pode ser uma importante ferramenta teórico-metodológica para os arquivos pessoais, pois busca encontrar a ligação existente entre os documentos e os contextos que os geraram” (Oliveira & Sobral, 2017, p. 1013). Os tipos documentais quando conceituados possibilitam a compreensão do espectro do termo por meio de sua definição. Os verbetes garantem a correta interpretação de um termo, uma vez que a polissemia de certas palavras podem ser um desafio ao entendimento do tipo documental.

4 Resultados

Na contramão da prática arquivística, que ainda não identifica a importância metodológica da conceituação dos tipos documentais como etapa essencial da compreensão dos arquivos, os projetos que temos desenvolvido demonstram que a associação do conceito ao tipo documental tem alta eficácia na produção de conhecimento sobre o arquivo estudado e viabiliza a construção de pontos de acesso relevantes para o processo de pesquisa do usuário dos arquivos, diminuindo o seu tempo de pesquisa e dando resultados de busca mais precisos.

A tipologia documental permite, como método de análise em arquivos pessoais, a compreensão do contexto documental e exprime em uma fórmula nuclear os elementos físicos e de gênese da produção documental. Os projetos Tipologia documental na família Barbosa de Oliveira e Tipologia documental na Família Barbosa de Oliveira: definição dos tipos documentais identificaram, até o momento, 203 espécies e 178 tipos documentais, num total de 749 documentos pertencentes ao campo empírico do projeto. Foram também produzidos 178 verbetes referentes à conceituação dos tipos identificados. Conforme já destacado em outros trabalhos:

Para dar conta da diversidade e complexidade da pesquisa e seu objeto, uma das formas selecionadas para a sistematização e divulgação do conhecimento produzido foi a elaboração de um glossário. Nesse instrumento, é possível não só encontrar os termos validados, respectivos conceitos e notas técnicas de pesquisa, mas também a representação dos relacionamentos de equivalência, por meio do Usado por ou Use e os demais relacionamentos pelo Ver também. Já foram conceituados 175 termos que identificam os tipos documentais (Oliveira & Oliveira, 2017, p.1576).

Em termos metodológicos, o glossário em produção busca dar conta não apenas da conceituação dos tipos documentais, mas também registra as notas de pesquisa e as relações entre os termos definidos, como forma de oferecer aos potenciais usuários outros pontos relevantes para a compreensão da definição.

Destacamos, a partir do desenvolvimento dos projetos citados a importância da definição dos tipos documentais como forma de garantir estabilidade de sentido dos termos utilizados para nomear documentos. Por mais que a fórmula espécie mais atividade seja reconhecida pelos profissionais da área, os limites do tipo documental pode ser melhor compreendido quando devidamente conceituado. Como forma de exemplificação do tipo documental, apresentamos:

Declaração de nascimento de escravo: documento obrigatório onde se é notificado, ao oficial do registro, o nascimento de um escravo para que proceda a matrícula do escravo junto ao órgão municipal.

Características típicas: A estrutura do documento está estabelecida no modelo D do Decreto n. 4835, de 01 de dezembro de 1871 que aprova o regulamento para a matrícula especial dos escravos e dos filhos livres de mulher escrava.

Na declaração deve constar o nome do proprietário, o dia do nascimento da criança, nome da mãe, sexo da criança, nome de batismo da criança e cor. Além disso, devem ser indicadas as duas matrículas da mãe (a feita pelo proprietário e a do município), bem como a atividade exercida.

Os documentos que compõem um arquivo pessoal refletem a sociedade na qual o produtor esteve inserido. Em sociedades escravagistas são comuns procedimentos burocráticos para o controle de nascimento, compra e venda, e deslocamento de escravos. A produção documental desse segmento é balizada pela legislação do período, que atua como instrução normativa aos proprietários de escravos. No conjunto documental estudado nos projetos de pesquisa da Fundação Casa de Rui Barbosa, destacamos a Declaração de Nascimento de Escravo e a Declaração de Chegada de Escravo, ambos tipos documentais só puderam ser nomeados e definidos a partir do estudo da legislação do período. Fundamental no processo de identificação e conceituação foi o Decreto nº 4835, de 01 de dezembro de 1871, que aprova o regulamento para a matrícula especial dos escravos e dos filhos livres de mulher escrava, o decreto explicita os procedimentos necessários para esse registro. A Declaração de Chegada de Escravo, também necessitou do entendimento dos procedimentos legais que orientam a mudança de residência de um escravo para ser definida.

É importante considerar que a pluralidade que representa a vida de um indivíduo, que como produtores de documentos podem estar inseridos em diferentes épocas, grupos, profissões e relações, torna os arquivos pessoais um *locus* de estudo ímpar sobre a produção e acumulação de documentos fora do âmbito institucional. O que não significa que os documentos acumulados não possam ser regulados pela legislação, procedimentos burocráticos e normas de conduta. Outro exemplo interessante é o Recibo de Pagamento de Tributo, conceituado como:

Documento padronizado comprobatório de pagamento compulsório ao Estado de impostos e taxas.

Como exemplo, temos o recibo de pagamento de arruamento, a décima urbana, de foro, de foro de sesmaria, de indústrias e profissões, imposto predial, de pena de água, metros corridos, propriedade rural, laudêmio e de taxa de águas e esgotos. Características típicas: Deve conter a instituição a receber o tributo compulsório. Deve conter o nome de quem realiza o pagamento e a data do pagamento. Deve conter a identificação do serviço e o valor da contribuição devida. Usualmente apresenta número e o no texto do documento overbo receber conjugado. Nota de pesquisa sobre o recibo de pagamento de tributo (pena de água): em 1840, o regulamento nº 39 estabelece a maneira de se concederem água dos aquedutos públicos, a concessão não podia exceder a duas penas de água. Em 1875 foi estabelecido um novo regulamento provisório para a execução da lei nº 2639, em relação à obrigatoriedade do suprimento de água para as construções existentes no perímetro da cidade determinada pelo governo. Em 1882, o decreto é aprovado. Em 1898, um novo regulamento para a arrecadação das taxas de consumo de água foi aprovado.

Nomear documentos também exige a percepção do quão próximo da atividade será o termo atribuído na identificação. Elucidativo desse quadro é o tipo documental Recibo de Pagamento de Tributo, uma vez que, adotamos o pagamento de tributo como atividade, invés de nomear cada um dos tributos como diferentes tipos documentais. Essa escolha baseia-se no horizonte de que um tipo documental deve ser capaz de representar a atividade geradora do documento, mas com vistas a também conseguir ser representativo de uma gama de documentos de uma mesma atividade. Como exemplar de atividades circunscritas a vida íntima, destacamos o tipo Carta de amor:

Documento em forma de carta em que são expressos sentimentos de interesse amoroso mútuo entre duas pessoas, pode expressar também a conquista, a sedução e a reconciliação. Pode ser escrita de forma platônica, unilateral, ou seja, a partir de um sentimento idealizado ou cândido.

Características típicas:

- Pode apresentar-se nos formatos, carta, cartão, bilhete, etc.
- O texto privilegia a manifestação e o vínculo amoroso.
- Os correspondentes normalmente utilizam apelidos e se tratam amorosamente.

A representação das cartas na prática arquivística usualmente atribuiu rubricas amplas para abrigar os documentos em forma de epístola sob o termo

correspondência. Nesse sentido, a tipologia documental potencializa a representação dessa espécie nas diferentes atividades que assume. No âmbito da intimidade e das relações pessoais às cartas podem assumir atividades relativas às emoções, como a manifestação de amor. Ressaltamos que, o tratamento amoroso em uma epístola não configura que a atividade que orienta a produção da carta seja o ato de manifestar o amor por outrem. Com isso, reforçamos a importância da conceituação dos tipos documentais, uma vez que, é pelo conceito que podemos vislumbrar o escopo de representação de um termo.

Consideramos que a partir dos exemplos de tipos documentais apresentados, é possível identificar a que no ato de viver um indivíduo pode produzir documentos a partir dos mais diferentes atos. Tais documentos acumulados podem vir a ser descritos por um profissional da informação, que deve orientar sua prática profissional pela garantia da estabilidade de sentido do documento em relação a sua gênese.

Portanto, devemos integrar cada vez mais em nossas agendas de pesquisas o estudo dos tipos documentais. Uma vez que, é por meio do conhecimento da produção documental, das formas e características que podemos contribuir para uma representação adequada dos arquivos e seus documentos.

Referências

- Bellotto, H. L. (2008). *Diplomática e tipologia documental em arquivos*. Brasília: Briquet de Lemos/Livros.
- Camargo, A. M. A. (2009, jul-dez.). Arquivos pessoais são arquivos. *Revista do Arquivo Público Mineiro*. Belo Horizonte, n. 2, pp. 26-39.
- Camargo, A. M. A. (2015) Sobre espécies e tipos documentais. In Camargo et al. (org.). *Dar nome aos documentos*. São Paulo: Instituto Fernando Henrique Cardoso, p.14-31.
- Hjørland, B. (2008). Deliberate bias in knowledge organization. *Advances in Knowledge Organization*, v.11, pp.256-261.
- Menne-Haritz, A. (2005). Avaliação ou Documentação: podemos avaliar arquivos através da seleção de conteúdos? *Registro: Revista do Arquivo Público Municipal de Indaiatuba*, 4(4), (pp. 18-34). São Paulo.
- Oliveira, L. M. V. (2005). *Tipologia documental na família Barbosa de Oliveira*. Disponível em: http://www.casaruibarbosa.gov.br/arquivos/file/Bolsistas13/aa%20_%20Tipologia%20na%20Familia%20Barbosa%20de%20Oliveira.pdf
- Oliveira, L. M. V. (2010). *Análise tipológica dos documentos em arquivos pessoais: uma representação do código social*. Disponível em: http://www.casaruibarbosa.gov.br/dados/DOC/bolsistas/2010/FCRB_Selecao_de_Bolsistas_2010_Analise_tipologica_dos_documentos.pdf.

- Oliveira, L. M. V. (2010). *Modelagem e status científico na descrição arquivística no campo dos arquivos pessoais*. Tese de Doutorado. Universidade de São Paulo.
- Oliveira, L. M. V. (2012). *Descrição e pesquisa: reflexões em torno dos Arquivos Pessoais*. Rio de Janeiro: Móbile.
- Oliveira, L. M. V. (2015). *Tipologia documental na Família Barbosa de Oliveira: definição dos tipos documentais*. Fundação Casa de Rui Barbosa.
- Oliveira, L. M. V. (2016). A propósito da Carta de Mário. In Freitas, L. S.; Fonseca, V. M. M.; Lima, M. H. T. F. (orgs.). *Ética e políticas da informação: agentes, regimes e mediações*. Niterói: EdUFF (Estudos da Informação, 4). No prelo.
- Oliveira, L. M. V. (2016). Os tipos documentais: a forma de registro de um modo de viver. In: *Seminário Hispano Brasileiro de Investigación en Información, Documentación y Sociedad*, 5, Madrid. Inédito.
- Oliveira, L. M. V. & Oliveira, I. C. B. (2016). *Glossário de tipos documentais em arquivos pessoais* (versão preliminar).
- Oliveira, L. M. V.; Oliveira, I. C. B. (2018). Representação arquivística e a conceituação dos tipos documentais nos arquivos pessoais. In Borges, M. M.; Sanz-Casado, E.; González-Valiente, C. L. A Ciência Aberta: o contributo da Ciência da Informação. VIII Encontro Ibérico EDICIC, 2017. *Bibliotecas. Anales de Investigación*, v.14, n.2, pp.1573-1577.
- Oliveira, L. M. V.; Sobral, C. C. (2018). Arquivos pessoais e seus tipos documentais: a perspectiva da representação. In Da Graça Simões, M.; Borges, M. M. *Tendências atuais e perspectivas futuras em organização do conhecimento. Atas do III Congresso ISKO Espanha-Portugal XIII Congresso ISKO Espanha*, Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20, pp. 1011-1017.
- Sobral, C. C. (2016). *Relatório parcial de atividades*. Fundação Casa de Rui Barbosa.
- Yeo, G. (2007). Concepts of record (1): evidence, information, and persistent representations. *The American Archivist*, v.70, n.2, pp. 315-343,.
- Thomassen, T. (2006, jan.-jun.). Uma primeira introdução à Arquivologia. *Arquivo & Administração*. Rio de Janeiro, v.5, n.1, pp.5-16.

Colecciones digitales patrimoniales especializadas. Estudio de la Red ARACNE

María-Luisa Alvite-Díez¹ y Nieves Pena Sueiro²

¹ ORCID [0000-0003-1490-8936](https://orcid.org/0000-0003-1490-8936). Área de Biblioteconomía y Documentación,
Universidad de León, León, España.
luisa.alvite@unileon.es

² ORCID [0000-0003-2028-2765](https://orcid.org/0000-0003-2028-2765). Departamento de Letras. Universidade da Coruña, A Coruña,
España.
nieves.pena.sueiro@udc.es

Resumen. Este trabajo analiza una muestra de colecciones digitales especializadas, nacidas en el ámbito de la investigación en humanidades, con el fin de perfilar la fase actual en la que se hallan estas colecciones, valorar fortalezas y debilidades de estos sistemas y proponer posibles actuaciones futuras. A tal fin, se seleccionan como objeto de estudio las bibliotecas digitales albergadas en la *Red ARACNE -Red de humanidades digitales y letras hispánicas*. Se han establecido cuatro bloques de análisis que atienden a los contenidos de la colección; a los estándares y modelos de datos; a la interfaz de usuario y a los valores añadidos desarrollados. Los resultados muestran que las colecciones digitales analizadas representan una aportación singular, tanto como herramientas para investigadores, particularmente, hispanistas, como en lo que atañe a su notable contribución a la difusión del patrimonio documental. La concepción de ARACNE como nodo que agrupa proyectos de humanidades bajo estándares de interoperabilidad incrementa la visibilidad de la investigación y subraya los beneficios de las acciones coordinadas para interrelacionar información.

Palabras clave: colecciones digitales; humanidades; interfaces; organización del conocimiento

Abstract. A sample of specialised digital collections, born in the field of humanities research, is analysed in order to outline the current phase of these collections, assess the strengths and weaknesses of these systems and propose possible future actions. For this aim, the digital libraries housed in the *ARACNE -Network of Digital Humanities and Hispanic Letters* are selected as the object of study. Four blocks of analysis have been established that deal with the contents of the collection, the standards and data models, the user interface and the added values developed. The results show that the digital collections analysed represent a singular contribution, both as tools for researchers, particularly hispanists, and as regards their notable contribution to the dissemination of documentary heritage. The conception of ARACNE as a node

that merges humanities projects under interoperability standards increases the visibility of research and underlines the benefits of coordinated actions to interrelate information.

Keywords: digital collections; humanities; interfaces; Knowledge organization.

1 Introducción

En España, desde mediados de la década de los noventa del siglo pasado, comenzaron su andadura proyectos de investigación en el ámbito de las humanidades que han conllevado la construcción de bases de datos, bibliotecas y archivos digitales especializados, tomando como punto de partida recursos tradicionales o digitales de bibliotecas, archivos y museos.

Particularmente en el ámbito filológico, estas colecciones digitales han ido conformando plataformas personalizadas que integran objetos digitales patrimoniales, descritos, analizados, contextualizados y enriquecidos con fines académicos. Estos corpus y repositorios se han diseñado con el doble objetivo de construir herramientas especializadas para la investigación y de conformar instrumentos generadores de nuevo conocimiento.

Estos desarrollos tecnológicos guardan relación con los resultados del trabajo de Given y Willson (2018) en el que se observan dos nuevas prácticas de investigación entre los académicos de las disciplinas humanísticas: el diseño de herramientas digitales adecuadas a sus necesidades específicas y la preparación de datos en formatos que sean almacenables, accesibles, analizables y usables en el proceso de escritura. Asimismo, Zeng (2019), llama la atención sobre el concepto acogido con éxito por la investigación en humanidades de “datos inteligentes” (smart data), para referirse a datos transformados y analizados para descubrir su potencial oculto. Es la misma línea, subraya esta misma autora, las estrategias de enriquecimiento semántico representan un paso crucial para la interconexión de datos relacionados, la contextualización, el descubrimiento o la reutilización, entre otros.

Las colecciones digitales diseñadas a partir de líneas de investigación humanísticas suelen ser fruto del trabajo colaborativo y presentan, generalmente, un interés multidisciplinar y multilingüe. Persiguen la creación de recursos rigurosos que contribuyan al desarrollo científico en disciplinas diversas: filología, literatura, filosofía, historia, arte, bibliografía, etc. En muchos de estos proyectos se han analizado fuentes originales, se han digitalizado textos y otras manifestaciones culturales, se han establecido enlaces a las colecciones fuente y se han firmado convenios con bibliotecas y archivos que custodian los originales para la difusión del patrimonio.

En este marco, consideramos de interés explorar una muestra de colecciones digitales especializadas, nacidas en el ámbito de la investigación, con el fin de conocer los modelos de datos implementados, las herramientas de organización del conocimiento, las funcionalidades de las interfaces desarrolladas, las posibles sinergias con archivos

y bibliotecas depositarios del patrimonio documental fuente y, en definitiva, el valor de estos recursos para la investigación y la difusión del acervo patrimonial.

2 Objetivos y metodología

Este trabajo analiza una selección de colecciones digitales creadas en el desarrollo de proyectos de investigación en humanidades con el objetivo de conocer el estadio actual de estos repositorios especializados.

A tal fin, se seleccionan como objeto de estudio las bibliotecas digitales albergadas en *ARACNE - Red de humanidades digitales y letras hispánicas*, red nacida en 2011 y que estuvo financiada por el Ministerio de Economía y Competitividad del Gobierno de España (FFI2011-15606-E) en la que participan seis equipos de investigación de larga y reconocida trayectoria. Uno de los aspectos más relevantes de ARACNE se fija en la búsqueda de la interoperabilidad mediante la implementación de un metabuscador que, no solo permite obtener resultados de todas las bases de datos y recursos de los proyectos que la integran, sino que, al aplicar el protocolo OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting), facilita la difusión y visibilidad del contenido del conjunto de las colecciones, 16 en el momento actual, y la red se convierte en agregadora de recolectores como HISPANA o EUROPEANA. En el momento actual, en el Directorio y Recolector Hispana aparecen 31546 objetos digitales procedentes de la Red ARACNE.

Se trata de una muestra que consideramos representativa para perfilar la fase actual en la que se hallan las colecciones digitales especializadas en el ámbito de las humanidades en España con la intención de valorar fortalezas y debilidades de estos sistemas y de proponer posibles actuaciones futuras.

La metodología atiende como punto de partida a trabajos que presentan modelos de análisis de bibliotecas digitales (Fuhr et al, 2007; Alvite Díez, 2009; Estivill Rius, Gascón García, Sulé Duesa, 2010; Sulé Duesa, Estivill Rius y Gascón García, 2011; Rodríguez Yunta, 2014; Sheikhshoei, et al. 2018). Se revisa paralelamente la literatura científica publicada sobre las colecciones digitales objeto de estudio para conocer su origen, las líneas de investigación perseguidas, objetivos y alcance.

Se han establecido cuatro bloques de análisis:

- Colección: descripción del proyecto, contenidos y objetivos.
- Arquitectura de la información: accesibilidad, interoperabilidad, tecnologías, estándares, modelos de datos y herramientas de organización del conocimiento.
- Interfaz de usuario: prestaciones de búsqueda y navegación, presentación y visualización de resultados, objetos digitales y sistemas de ayuda.
- Valores añadidos: interrelaciones, contextualización, enriquecimiento semántico, integración de herramientas tecnológicas derivadas, sinergias con colecciones digitales patrimoniales.

Por lo que se refiere a la exploración de los recursos seleccionados, esta se ha llevado a cabo a lo largo de los meses de marzo y abril de 2019.

3 Análisis

En este epígrafe se presenta el análisis de ARACNE y de las colecciones que integra estructurado en los bloques señalados en la metodología.

3.1 Colección: descripción del proyecto, contenidos y objetivos

ARACNE surgió como respuesta a la necesidad de coordinación de criterios comunes o consensos en la aplicación de tecnologías y en la evaluación de resultados que algunos equipos de investigación de proyectos de humanidades echaban en falta.

La iniciativa se originó en el *Seminario BIDESLITE* (Casas del Álamo, Redondo Pérez y Sánchez Bellido, 2013) y se concretó con la solicitud de una acción complementaria para la creación de la Red (solicitada por el Profesor Pedro Ruiz, IP del Grupo PHEBO), en la que participaron seis proyectos de investigación españoles (BIDISO, BIESES, BSF, CLARISEL, DIALOGYCA y PHEBO). La Red ARACNE se hizo pública el 25 de marzo de 2015 (Pena Sueiro, 2017). Los objetivos de la red fueron los siguientes:

- La elaboración de un portal que reuniese y ordenase la información sobre proyectos e instrumentos de las letras hispánicas, con el fin de servir de lugar de convergencia y coordinación entre iniciativas (las ya existentes y las emergentes).
- El desarrollo de un metabuscador, que permitiese unificar la búsqueda de datos en los proyectos participantes sin perder su especificidad, de modo que los resultados se visibilicen y aprovechen más allá de su campo de estudios específico. El desarrollo de este instrumento tecnológico debía de servir de base para la futura incorporación de otros proyectos, así como de piedra de toque para recomendaciones dirigidas a nuevos proyectos en aspectos que puedan facilitar su interoperabilidad.

Las colecciones que integra ARACNE han sido desarrolladas por los grupos de investigación mencionados anteriormente y, en algunos casos, cuentan con más de veinte años de andadura. BIDISO (Universidade da Coruña). Dio sus primeros pasos en 1993 de la mano de Sagrario López Poza. Desde 2011 la dirección fue asumida por Nieves Pena Sueiro. Ha desarrollado seis plataformas. BIESES. Bibliografía de Escritoras Españolas, comenzó en el año 2003 como proyecto formal financiado por los planes nacionales I+D, nace bajo la dirección de Nieves Baranda Leturio (UNED). BSF. Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico. Proyecto surgido en el año 2002 en la Universidad de Murcia, bajo la dirección de José Luis Villacañas Berlanga. CLARISEL (Universidad de Zaragoza), se gestó a finales de la

década de los noventa bajo la dirección de Juan Manuel Cacho Bleuca y María Jesús Lacarra. Cuenta con cinco plataformas agregadas a ARACNE. DIALOGYCA BDDH. Biblioteca Digital de Diálogo Hispánico (Universidad Complutense de Madrid). Sus inicios se remontan al año 2000, con proyectos encabezados por Consolación Baranda Leturio y Ana Vian Herrero. PHEBO. Poesía Hispánica en el Bajo Barroco (Universidad de Córdoba), surge en el año 2010 de la mano de Pedro Ruiz Pérez.

3.2 Arquitectura de la información

ARACNE establece una serie de requisitos relacionados con la arquitectura de la información de las colecciones digitales que integra. Los recursos alojados en este portal digital han de ser publicados bajo alguna modalidad de licencia que permita su acceso libre (Creative Commons o similar). Tecnológicamente, el requisito clave establecido es que estas colecciones deben compartir los metadatos de sus recursos digitales a través del protocolo OAI-PMH y seguir las especificaciones del modelo de metadatos de Europea (EDM).

El portal web de ARACNE fue desarrollado por el Laboratorio de Bases de Datos (LBD) de la Universidade da Coruña (<http://lbd.udc.es/>). Este mismo centro se ocupó del desarrollo de ARACNE como metabuscador y recolector OAI-PMH. Se ha concebido ARACNE como una herramienta de código abierto, operativa para su implementación en trabajos futuros.

En cuanto a las características individuales de los recursos que conforman la red, se observan planteamientos y desarrollos tecnológicos dispares que responden a las exigencias y objetivos de los proyectos de investigación de origen. Los problemas técnicos, metodológicos, de modelado de datos y las soluciones logradas se presentaron con distinto detalle en el *Seminario BIDESLITE* (Casas del Álamo, Redondo Pérez, y Sánchez Bellido, 2013). En general, predominan desarrollos propios asentados en modelos de bases de datos relacionales. En BIDISO las aplicaciones han sido desarrolladas con *software* libre, lenguaje de programación JAVA y tecnologías como Struts, Hibernate, Spring, etc. Los sistemas gestores de bases de datos que se han utilizado han sido PostgreSQL y MySQL. Las colecciones de CLARISEL, *COMEDIC* y *DINAM* hacen uso igualmente de MySQL. Las plataformas de BIESES y DIALOGYCA han empleado el gestor documental Knosys de Micronet (<https://www.micronet.es>). PHEBO se ha desarrollado con Drupal y el software de indexación de Apache.

Se observa en buena parte de las colecciones una especial preocupación por el control de datos de autoridades y su enriquecimiento: variantes del nombre, seudónimos, actividad profesional, dedicatarios, fechas, estado, etc. Se ha desarrollado una labor de conceptualización y categorización importante. Así, el equipo BIDISO ha elaborado para las colecciones de *Emblemas Traducidos* y *Emblemática hispánica*, un tesoro (palabras clave, onomástico y autoridades). *Relaciones de sucesos* dispone de una clasificación por tipología y subgéneros y varios tesauros (responsables de la edición -impresores, editores, autores, traductores,

etc.; lugares, bibliotecas...). *Symbola* ha implementado la clasificación Iconclass en la base de datos y emplea una taxonomía particular para los motivos pictóricos. EN BSF cuentan con "identificadores onomásticos" e "identificadores topográficos".

DIALOGYCA articula la información a partir del concepto de "obra", en lugar de "documento", esta solución permite agrupar testimonios y documentos de diversa tipología, conservados en bibliotecas muy dispersas y recoger de ese modo junto a la información literaria la tradición textual de cada diálogo.

Por su parte, el etiquetado en TEI (Textual Encoding Initiative) se ha empleado en BIESES y en dos de las colecciones de BIDISO: la *Biblioteca Digital de Libros de Emblemas Traducidos* y en la edición de textos de *Relaciones de sucesos*.

3.3 Interfaz de usuario

El metabuscador ARACNE, lanza la consulta contra el conjunto de metadatos Dublin Core de las 16 colecciones agregadas. El motor dispone de búsqueda simple y avanzada. La primera consta de una caja única y en la interfaz de búsqueda avanzada es posible combinar campos de búsqueda bibliográficos tradicionales: título, idioma, fecha, autor, editor-impresor, fuente y materia. Resulta destacable, como se observa en la Figura 1, la personalización de las etiquetas de algunos campos, complementados con una ayuda contextual que aclara la cobertura de los mismos, es el caso de "autoría" o "fuente".

Figura 1. Interfaz de ARACNE

El sistema dispone de búsqueda por frase exacta y operadores de expansión. El motor devuelve los resultados ordenados por relevancia, permitiendo *a posteriori* reordenar por autoría, título, fecha o biblioteca digital, esto es, los campos disponibles en la presentación abreviada de resultados. Los registros completos muestran los campos descriptivos completos, la colección de procedencia, el enlace al registro en la misma, así como la licencia del recurso. Es posible navegar por autores y colecciones.

Por su parte, la exportación de resultados en ARACNE está disponible en CSV, hasta un máximo de 1000 registros. En cuanto a la ayuda, aparece de forma destacada en la interfaz y consiste en una ventana emergente con instrucciones genéricas de búsqueda. No se han observado opciones de alertas para usuarios registrados, o similares.

En lo que se refiere a las interfaces de las colecciones agregadas en la red ARACNE, las diversas soluciones persiguen, en general, articular modelos que sumen a los metadatos bibliográficos, la interrogación o filtro por los distintos elementos de análisis empleados en el estudio de los objetos digitales correspondientes. Se observan desarrollos desiguales y una clara evolución hacia interfaces más usables, particularmente en aquellos grupos de investigación que cuentan con proyectos más cercanos en el tiempo o cuyos modelos conceptuales e interfaces se han remodelado a lo largo de estos años. Podemos señalar que, en general, se ha primado en el diseño la construcción de interfaces de búsqueda avanzada que permiten definir las búsquedas de los usuarios “naturales” de los sistemas, investigadores, con la máxima precisión y optimizando el ingente trabajo descriptivo y de análisis realizado. No se han observado herramientas que articulen automáticamente facetas o categorías de contenidos una vez que los sistemas muestran los resultados. Tampoco se contempla desde las interfaces la posibilidad de búsqueda en el texto completo/digitalizado de las colecciones. En este marco, el agregador ARACNE potencia la capacidad de recuperación e interrelación de contenidos y la visibilidad del conjunto de las colecciones desde una caja de búsqueda única, si bien, por el momento, no incluye funcionalidades de descubrimiento al mostrar los resultados que podrían servir para desvelar vínculos o correlaciones entre las distintas colecciones.

3.4 Valores añadidos

Una de las fortalezas indiscutible de las colecciones aquí analizadas es, sin duda, la concepción de estas bibliotecas como proyectos colaborativos e interdisciplinares de investigación. El estudio riguroso de los textos y objetos culturales tratados, la labor de digitalización emprendida en muchos casos y la tarea ingente de enlazado con las fuentes patrimoniales de origen, fundamentalmente con bibliotecas, es encomiable.

El valor añadido crucial de PHEBO se sitúa en el rigor y la calidad académica de los textos editados y disponibles en abierto en la plataforma. DYALOGICA, por su parte, ha prestado especial atención a la descripción codicológica y tipobibliográfica de los testimonios digitalizados.

CLARISEL, un grupo de investigación esencialmente filológico, incorpora entradas interdisciplinarias de historia, derecho, arte, folclore, antropología, entre otras, que puedan ayudar a interpretar y contextualizar la materia objeto de estudio. Sirve de ejemplo *COMEDIC*, colección que presenta una gran riqueza descriptiva y enlaces a testimonios manuscritos, testimonios impresos, testimonios de lectura, etc. Cada registro, distingue, no solo el nombre de su autor y sus variantes, sino también el glosador, prologuista, mecenas, etc. Se indica el título de la obra original, si es traducida, y las variantes del título que facilitan los distintos testimonios, las fechas aproximadas de composición, o de traducción, y la materia, o materias, a las que puede adscribirse por su contenido. Se facilitan los enlaces a aquellos ejemplares digitalizados. Se presta atención a las ilustraciones e imágenes, a su recepción o a los elementos paratextuales que acompañan a las ediciones. Se evidencia, en suma, el valor multidisciplinar de la colección, útil para estudios sobre diversos ámbitos de la difusión del libro y de la cultura, entre otros.

Colecciones como *BIESES* pueden servir de ejemplo en el enlazado prolijo de fuentes. Los vínculos con instituciones documentales se subrayan igualmente en muchas de las plataformas, así, BSE, Biblioteca Saavedra Fajardo, mantiene una gran riqueza de relaciones en sus registros: “Otras digitalizaciones” (en otras instituciones, en Google Libros...); “Depósito de ejemplares” en otras bibliotecas; “Documentos relacionados”, etc. Por su parte, el *Catálogo y Biblioteca Digital de Relaciones de Sucesos* (BIDISO) señala hasta 179 bibliotecas utilizadas en la colección, 30 de las cuales se han utilizado profusamente para el análisis y digitalización de fondos, estableciendo convenios para digitalizar las *Relaciones* que poseen en muchos casos. En la *Biblioteca Digital de Libros de Emblemas Traducidos* (BIDISO) a las obras digitalizadas específicamente en el proyecto han añadido las que se han localizado digitalizadas en otras bibliotecas, incorporando la ficha descriptiva, un enlace a la obra y la biblioteca que la ofrece. También aportan ayudas para explotar mejor su contenido. Por ejemplo, en las enciclopedias que están en latín, se ofrece un índice traducido al español con enlace directo a las páginas donde se da información sobre el tema.

Un apartado especial merece el etiquetado semántico en TEI abordado en algunas colecciones, como se ha adelantado. Así, por ejemplo, dentro del corpus *BIESES* es posible realizar búsquedas textuales (“Edición de paratextos” y “Buscador de paratextos de autoras”, en fase beta); se ponen de manifiesto todos los agentes que intervienen en el campo cultural de la edición de la obra de cada autora, sus posibles cruces, interferencias o similitudes, no solo de forma individual, sino también como representantes de funciones sociales que operen habitualmente sobre estos textos (el censor/es; el historiador, el editor, el círculo de autorización....) y sus modulaciones discursivas.

En cuanto a la explotación de datos y visualización en otros entornos no textuales, resulta destacable la implementación de tecnología de geolocalización en el *Catálogo y Biblioteca Digital de Relaciones de Sucesos* (BIDISO) donde se presenta un mapa asociado. En *DINAM* (CLARISEL) se han elaborado árboles genealógicos. En el caso de *BIESES* se incluyen desarrollos de valor añadido como: “Redes de sociabilidad de

las autoras” cada autora es un nodo que se relaciona con el resto de personas mencionadas en las ediciones de sus libros.

4 Conclusiones

Consideramos que las colecciones digitales analizadas representan una aportación singular, tanto como herramientas para investigadores, particularmente, hispanistas, como en lo que atañe a su notable contribución a la difusión del patrimonio documental. Debe ser reconocido el esfuerzo que durante décadas han llevado a cabo los grupos de investigación responsables de estas colecciones para construir, alimentar y mantener los proyectos.

Resulta especialmente destacable la apuesta por la interoperabilidad y los valores añadidos que aportan las colecciones en cuanto al nivel de análisis de los objetos digitales, la contextualización rigurosa de los mismos y la riqueza de interrelaciones proporcionadas, entre otros. Los metadatos que acompañan a cada objeto digital son el resultado de investigaciones rigurosas y, a su vez, contribuyen a generar nuevo conocimiento.

Aunque se observan diferencias significativas y soluciones desiguales en el comportamiento de las distintas interfaces de las colecciones, parece aconsejable, en conjunto, afrontar mejoras en la actualización del diseño de las mismas y en aspectos de usabilidad vinculados a la visualización, presentación y exportación de resultados.

Sería recomendable abordar la implementación de tecnologías semánticas (publicación en RDF y datos abiertos enlazados), la incorporación de herramientas de enriquecimiento semántico o la integración de aplicaciones para la explotación de datos (geolocalizadores, analizadores sintácticos, semánticos, etc.)

Se propone, asimismo, potenciar las sinergias con bibliotecas y archivos con el fin de mejorar la visibilidad y difusión de estas colecciones de investigación especializadas y la calidad de los datos por parte de las instituciones documentales patrimoniales. Sin duda, los convenios de Dialogyca BDDH con la Biblioteca Nacional de España y con la Biblioteca Histórica Marqués de Valdecilla señalan el camino a seguir.

Por último, se constata que la constitución de ARACNE aporta una serie de beneficios, no solo para la comunidad científica, sino para la ciudadanía en general. Evidencia la consideración de las humanidades digitales como un nuevo espacio interdisciplinar y supone la creación de un nodo que agrupa proyectos de ciencias humanas que trabajan con tecnologías y producen herramientas, recursos, etc., resultado del análisis experto. Entre los puntos destacables que aporta ARACNE podemos subrayar: el incremento y mejora de la visibilidad de la investigación en humanidades; el establecimiento de la necesidad de coordinación y creación de protocolos de trabajo y de orientaciones de investigación para interrelacionar la información y, finalmente, la creación de un interfaz de búsqueda conjunta asentada en estándares que hace posible la convergencia e integración de los 16 repositorios

implicados y ha permitido que los contenidos de estas colecciones agregadas dejen de ser opacos a los buscadores generalistas.

La apuesta por la Red ARACNE, su continuidad y mejora, ha llevado a plantear una nueva solicitud de financiación al Ministerio que persigue la actualización tecnológica (de software, diseño adaptativo, etc.) y legal (con la incorporación de exigencias de la nueva ley de protección de datos), y la ampliación de la red con la incorporación de tres nuevos grupos cuyos proyectos de investigación son referentes en las letras hispánicas, y un grupo de innovación tecnológica, que asistirá en los procesos de desarrollo técnico.

Referencias

- Alvite Díez, M.L. (2009). Interfaces y funcionalidades de Bibliotecas digitales. *Anales de Documentación*, 12, 7-23. Recuperado de <http://revistas.um.es/analesdoc/article/view/70221>
- ARACNE: Red de humanidades digitales y letras hispánicas. <http://www.red-Aracne.es/>
- Arrigoni, E., & Rodríguez López, E. (2014). La red de investigación de “Humanidades Digitales y Letras Hispánicas”: avance de Red-ARACNE. En Á. Baraibar (Ed.), *Visibilidad y divulgación de la investigación desde las Humanidades digitales: experiencias y proyectos: BIADIG: Biblioteca áurea digital* 22 (pp. 243–251). Pamplona: Universidad de Navarra. Recuperado de <http://dadun.unav.edu/handle/10171/35723>
- Baranda Leturio, C., & Rodríguez López, E. (2014). Red ARACNE: retos y objetivos de un proyecto de coordinación en letras hispánicas digitales. En S. López Poza, y N. Pena Sueiro (Eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro* (pp. 101–109). A Coruña: Universidade da Coruña, SIELAE. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/5181020.pdf>
- Casas del Álamo, M., Redondo Pérez, G., & Sánchez Bellido, S. (Eds.). (2013) *Actas del Seminario Internacional sobre Bibliotecas digitales y bases de datos especializadas para la investigación en Literaturas Hispánicas (BIDESLITE)*. Madrid, 4-5 de julio de 2011. Madrid: Grupo de estudios de prosa hispánica bajomedieval y renacentista del Instituto Universitario Menéndez Pidal, Universidad Complutense de Madrid. Recuperado de <https://eprints.ucm.es/21207/>
- Estivill Rius, A., Gascón García, J., & Sulé Duesa, A. (2010). Las colecciones digitales patrimoniales españolas: políticas de colección y presentación de la colección. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, (25), 32. doi:<http://doi.org/10.1344/105.000001673>
- Fernández Travieso, C. (2013). Estudio de codificación XML/TEI para Relaciones de Sucesos españolas, A Coruña, SIELAE. Recuperado de <https://www.bidiso.es/sielae/upload/estaticas/file/FTXMLTEIISBN2pr.pdf>

- Fuhr, N., Tsakonas, G., & Aalberg, T. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1), 21–38. doi:<http://doi.org/10.1007/s00799-007-0011-z>
- Given, L.M., & Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6), 807-819. doi:<https://doi.org/10.1002/asi.24008>
- López Poza, S. (2017). La base de datos «Symbola» de divisas o empresas históricas. Planteamiento y diseño conceptual. *Studia Aurea*, 11, 93-109. doi:<https://doi.org/10.5565/rev/studiaaurea.262>
- Pena Sueiro, N. (2017). El portal BIDISO: pasado, presente y futuro inmediato. Un ejemplo de evolución en aplicaciones de las HD. *Studia Aurea*, 11, 73-92. doi:<https://doi.org/10.5565/rev/studiaaurea.264>
- Rodríguez Yunta, L. (2014). Usabilidad y recuperación en Europeana: comparación con otras bibliotecas digitales y portales. En L. F. Ramos Simón, y R. Arquero Avilés (Coords.). *Europeana. La plataforma del patrimonio cultural europeo* (pp. 177-200). Gijón: Trea.
- Sheikhshoei, F., Naghshineh, N., Alidousti, S., & Nakhoda, M. (2018). Design of a digital library maturity model (DLMM). *The Electronic Library*. 36(4), 607-619, doi:<http://doi.org/10.1108/EL-05-2017-0114>
- Sulé, A., Centelles, M., Franganillo, J., & Gascón, J. (2016). Aplicación del modelo de datos RDF en las colecciones digitales de bibliotecas, archivos y museos de España. *Revista Española de Documentación Científica*, 39(1): e121. doi: <http://dx.doi.org/10.3989/redc.2016.1.1268>
- Sulé Duesa, A., Estivill Rius, A., & Gascón García, J. (2011). Interfaces de consulta en las colecciones digitales patrimoniales españolas. *Anales de Documentación*, 14(2). Recuperado de <http://revistas.um.es/analesdoc/article/view/113931>
- Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El Profesional de la Información*, 28(1), 1-35. doi:<http://doi.org/https://doi.org/10.3145/epi.2019.ene.03>

A dimensão social do catálogo Arca do Gosto no Brasil enquanto serviço de informação e sua relação com a organização do conhecimento

Gabrieli Aparecida da Fonseca¹ y Sonia Troitino²

¹ ORCID [0000-0002-1785-9896](https://orcid.org/0000-0002-1785-9896). Universidade Estadual Paulista-UNESP, Brasil.
gabrieli.arq@gmail.com.

² ORCID [0000-0002-7204-3283](https://orcid.org/0000-0002-7204-3283). Universidade Estadual Paulista-UNESP, Brasil.
sonia.troitino@unesp.br.

Resumo. Este artigo caracteriza-se como um recorte do projeto de doutorado intitulado “O papel da informação no resgate da cultura alimentar brasileira: uma análise do catálogo Arca do Gosto”, e diz respeito à dimensão social do catálogo Arca do Gosto, atentando para as especificidades que envolvem a sistematização da informação a partir do compartilhamento de conhecimento, principalmente a partir do encontro entre conhecimento científico e o saber popular, bem como atenta para os impactos sociais desse tipo de serviço de informação, que vem a favorecer o empoderamento de populações vulneráveis as imposições do sistema dominante. Assim, este estudo tem a proposta de realizar uma análise transdisciplinar de seu objeto, o catálogo Arca do Gosto - o qual se trata de um dos principais projetos do movimento Slow Food, e tem como proposta catalogar alimentos tradicionais em risco de extinção. Dessa forma, essa pesquisa se volta para diversos temas transversais.

Palavras-chave: Organização do conhecimento; Catálogo Arca do Gosto; Serviço de Informação; Soberania alimentar.

Abstract. This article is characterized as an excerpt from the doctoral project entitled “The role of information in the rescue of Brazilian food culture: an analysis of the Ark of Taste catalog”, and concerns the social dimension of the Ark of Taste catalog, paying attention to the specificities that involve the systematization of information based on knowledge sharing, mainly based on the encounter between scientific knowledge and popular knowledge, as well as paying attention to the social impacts of this type of information service, which favors the empowerment of vulnerable populations. impositions of the dominant system. Thus, this study proposes to carry out a transdisciplinary analysis of its object, the Ark of Taste catalog - which is one of the main projects of the Slow Food movement, and aims to catalog traditional foods at risk of extinction. Thus, this research turns to several transversal themes.

Keywords: Knowledge Organization; Ark of Taste Catalog; Information service; Food sovereignty.

1 Introdução

O presente artigo caracteriza-se como um recorte do projeto de doutorado intitulado “O papel da informação no resgate da cultura alimentar brasileira: uma análise do catálogo Arca do Gosto”, cuja temática diz respeito à dimensão social do catálogo Arca do Gosto, atentando para as especificidades que envolvem a sistematização da informação a partir do compartilhamento de conhecimento, principalmente quando se trata do encontro entre conhecimento científico e o saber popular, bem como atenta para os impactos sociais desse tipo de serviço de informação. Este estudo tem a proposta de realizar uma análise transdisciplinar de seu objeto – o catálogo Arca do Gosto – dessa forma, se volta para diversos temas transversais. Optou-se pela abordagem transdisciplinar por compreender, assim como Hjørland (1995, p.409), que esta possibilita uma visão mais orgânica e contextual, onde é possível ter dimensão do todo.

Esse artigo é resultado de um recorte temático do referido projeto e apresenta resultados parciais, voltando-se aos aspectos sociais do catálogo Arca do Gosto e temas correlatos. Para o desenvolvimento dessa pesquisa, foi escolhida uma combinação entre dois métodos: metateoria e estudo de caso, pois se acredita que a união entre dois métodos proporciona maior solidez a metodologia. O objetivo geral da pesquisa em questão é entender o papel da informação registrada no projeto Arca do Gosto e Fortalezas, do Slow Food e sua importância para a preservação da cultura alimentar brasileira. Assim, tem como objetivos específicos analisar a relação entre os critérios e itens selecionados pela Arca do Gosto no Brasil com as condições sociambientais da população, bem como a Organização do Conhecimento pode contribuir para o avanço da sociedade a partir da sistematização da informação.

Contudo, enquanto recorte de pesquisa, este artigo trata a respeito da importância da organização do conhecimento, especialmente de serviços de informação como a Arca do Gosto, para domínios específicos, no caso, para o resgate da cultura alimentar por meio da preservação de tradições e da biodiversidade, contribuindo com populações carentes e mais vulneráveis.

2 O catálogo Arca do Gosto enquanto sistema de informação e organização do conhecimento

O Slow Food é um movimento social criado em 1989 por um grupo de amigos italianos, com a proposta de defender a alimentação de qualidade, valorizar a agricultura familiar, o meio ambiente e as tradições culinárias (SLOW FOOD, 2015). O movimento também traz como lema a valorização do alimento considerado “bom,

limpo e justo” (MAKUTA, 2018, p.89), se contrapondo à disseminação do fast-food e à padronização da cultura alimentar, que implicam no desaparecimento de espécies, variedades, raças, técnicas, conhecimentos e saberes ligados à comida (FIDA, 2018).

No mundo todo, projetos liderados pelo Slow Food, como Arca do Gosto e Fortalezas, ajudam a preservar ingredientes, métodos de preparo e a salvar a biodiversidade local, da qual diversas comunidades são dependentes (SLOW FOOD, 2015).

Como um dos principais projetos do movimento, o catálogo Arca do Gosto, que é um catálogo que reúne informações a respeito de alimentos tradicionais em risco de extinção, busca identificar e cadastrar o conteúdo a respeito desses alimentos em uma base de dados de construção colaborativa, ou seja, conta com a indicação e apoio da população. A seleção dos tipos de alimentos a serem descritos no catálogo leva em consideração diversos critérios pré-estabelecidos pelo Slow Food, incluindo suas características de identidade regional, de modo que cada país possui produtos específicos descritos na Arca, conforme cultura e tradição.

A informação nesse caso, tanto é parte importante do processo que dá origem ao catálogo Arca do Gosto, quanto é o próprio produto final desse processo: o catálogo. Pois levando em conta suas características, o mesmo pode ser compreendido como um serviço de informação, que seleciona a informação relevante dentro de sua proposta e a sistematiza. A respeito da sistematização de informação Francisco Javier García Marco afirma que “es una tarea compleja y costosa que requiere, o bien un equipo de personas trabajando de forma coordinada, o la dedicación exclusiva de una persona durante mucho tempo”. (GARCÍA MARCO, s/d, p.186).

Embora o catálogo Arca do Gosto apresente características de um serviço de informação, trata-se de um trabalho complexo, especialmente pelo fato de o mesmo contar com a colaboração da sociedade para seu desenvolvimento, o que exige análise minuciosa das informações apontadas no preenchimento dos formulários de indicação de produtos. Assim, o formulário, que se encontra no site da Fundação Slow Food para a Biodiversidade, possui campos como nome do produto, área geográfica, categoria, usos gastronômicos, dentre outros. Antes de serem aprovadas pelo Comitê Técnico Internacional da Arca do Gosto, as indicações passam por avaliações da Fundação Slow Food para a Biodiversidade e da Comissão Nacional ou Regional, que verificarão se a indicação está de acordo com os critérios do projeto (MILANO; et.al, s/d, p.21).

A construção coletiva do catálogo Arca do Gosto é um aspecto importante que elucida a troca entre os saberes não apenas no que diz respeito a romper os limites impostos pela divisão desses saberes em disciplinas, em direção a transdisciplinaridade, mas também pelo fato de proporcionar a troca entre os saberes tradicionais e a ciência. Carlo Petrini salienta que o dualismo entre a ciência oficial e os saberes tradicionais fora reforçado a partir da chegada da revolução industrial, com a comercialização dos saberes através de patentes, o que não contribui para o bem comum. Para ele, assim como para o Slow Food, um mundo mais sustentável apenas pode ocorrer a partir da troca entre essas duas realidades, da dialética, e do diálogo em

nível de igualdade, evidenciando as competências e especificidades de cada um (PETRINI et. al., s/d).

É preciso ter cautela, pois qualquer pessoa pode contribuir com informações sobre produtos que tem potencial para estar na Arca do Gosto, e se não houver monitoramento atento nessa troca entre os saberes tradicionais e o científico, tal desequilíbrio pode acarretar prejuízos a Organização do Conhecimento processado pelo mesmo, correndo o risco de que informações errôneas ou inconsistências façam parte dessa sistematização. Contudo, a Organização do Conhecimento pode contribuir com os subsídios teóricos em relação ao tratamento da informação, conforme sugere Barité (2001). Além disso, a utilização de suas ferramentas comuns a Organização do Conhecimento se faz essencial para a conquista de bons resultados nessa troca, que faz parte da proposta da Arca do Gosto, pois as novas tecnologias não estão em conflito com o diálogo entre os saberes, elas podem inclusive fortalecer a catalogação e a difusão dos saberes tradicionais (PETRINI et. al., s/d).

Nesse sentido um aspecto importante da Organização do Conhecimento para o catálogo Arca do Gosto, assim como para outros catálogos semelhantes, se refere aos assuntos mais fundamentais da Organização do Conhecimento, que Hjørland se refere como sendo: Conceitos; Critérios para inclusão; Significado; Indexação; Relações semânticas; Assuntos; e Assuntos de pontos de acesso (HJORLAND, 2003).

O item critérios para inclusão, elencado por Hjørland (2003), é um ponto extremamente fundamental para a construção da Arca do Gosto, especialmente no que diz respeito a sua necessidade de se desenvolver de forma equilibrada e colaborativa, pois ao mesmo passo que “[...] pluralismo e diversidade dos sistemas de conhecimentos devem ser incentivados, pois são garantias de inovação e de uma abordagem holística capaz de dar dignidade aos saberes das comunidades” (PETRINI; et. al, s/d), se os critérios de inclusão não forem bem definidos e amparados cientificamente pode-se ter problemas como a inclusão de produtos em contexto errado. Da mesma forma, se os saberes tradicionais – o conhecimento popular, transmitido por meio da oralidade entre gerações – não forem considerados o mesmo pode ocorrer, pois “não pode existir democracia participativa sem o reconhecimento e a divulgação dos conhecimentos alimentares das comunidades, para o bem-estar das futuras gerações e do mundo natural. O direito ao alimento sem a socialização dos saberes é mera utopia” (PETRINI; et. al., s/d).

3 Arca do Gosto, preservação da biodiversidade alimentar e impacto social da informação

Marie Caroline Saglio Yatzimsky (2006) esclarece que não há o desaparecimento total das tradições culinárias, pois a comida aparece como um marcador de identidade e de distinção regional. De fato, a alimentação é um dos traços culturais mais fortes da identidade de um povo, servindo como indicador de resiliência de uma cultura em relação às ameaças, porém a imposição alimentar é uma realidade capaz de

desaparecer por completo com culturas alimentares tradicionais, sobretudo em regiões muito pobres e exploradas. Assim, “o Slow Food é uma voz importante na preservação da biodiversidade, do conhecimento tradicional, une o prazer gastronômico à alimentação consciente e responsável” (SLOW FOOD, 2015, p.55).

Além disso, o Slow Food explora as questões que envolvem a cultura alimentar de forma diferenciada, dando centralidade ao alimento. Essa dimensão fica evidente no catálogo Arca do Gosto, onde o alimento tem papel principal, uma vez que é o elemento básico para suprir a fome e garantir a biodiversidade alimentar. O movimento Slow Food possui como missão primordial salvaguardar a sociobiodiversidade, aproximar os agricultores dos coprodutores e difundir a educação alimentar e do gosto. Essa mobilização e conexão global têm apoiado diversos processos de articulação local, contribuindo com a valorização da agricultura familiar, das técnicas e conhecimentos tradicionais ligados à cultura alimentar e à identidade de diversos territórios. (MAKUTA, 2018).

Acredita-se que os hábitos alimentares merecem muita atenção, pois ao tratar cada alimento tradicional em risco de extinção como único, a Arca do Gosto exalta seu sabor, suas propriedades e valoriza seu local de origem. De acordo com o Fundo Internacional de Desenvolvimento Agrícola “valorizando as potencialidades locais, podemos desenvolver uma cadeia regional de valorização do local” (FIDA, 2018, p. 10). Como exemplo do impacto social de centralizar o alimento, no Brasil, temos o caso do umbu, que é um fruto de agradável sabor agridoce, extremamente nutritivo, sendo rico em vitamina C. Doces, sorvetes, suco concentrado, refresco, vinagre de umbu e a famosa umbuzada, são receitas populares que agregam valor de uso ao fruto e geram renda a população. Proveniente do umbuzeiro, – árvore do nordeste brasileiro, que segundo Brasil (2015) fora chamada de “árvore sagrada do Sertão” pelo famoso escritor Euclides da Cunha - possui grande reservatório de água em suas raízes, o que a faz muito resistente a períodos de estiagem. Além dos frutos, as raízes do umbuzeiro também são ricas em vitamina C e sais minerais, servindo de socorro ao nordestino nos períodos de seca, proporciona uma bebida saudável e um delicioso doce (BRASIL, 2015).

O caso do umbu demonstra o quão grande é o impacto do alimento na sociedade, tendo relação não apenas com a cultura, mas também com a autonomia dos povos locais, na luta contra a fome e na preservação do meio ambiente, como tem sido amplamente discutido ultimamente pelos defensores de propostas de produção voltadas para a agroecologia.

Contudo, “a informação e a educação são essenciais para que a sociedade entenda o potencial de transformação e o impacto gerado a partir de suas escolhas alimentares” (FIDA, 2018, p.20). Nesse sentido, o catálogo Arca do Gosto é uma ferramenta muito rica, segue o processo de seleção subjacente ao período de desenvolvimento dos padrões tradicionais de alimentação (BRASIL, 2014), conforme o Ministério da Saúde recomenda que seja a prática dos guias alimentares. O que destaca a importância da Organização do Conhecimento para a promoção da sociobiodiversidade e sustentabilidade locais, pois além da troca entre saberes, a Arca

do Gosto proporciona acesso a informações confiáveis a respeito da alimentação tradicional, empoderando a população em suas escolhas alimentares tanto no que diz respeito à produção quanto ao consumo. Para Brasil (2014) a autonomia das escolhas alimentares garante o direito humano à alimentação adequada e saudável.

4 Considerações finais

Enquanto serviço de informação, o catálogo Arca do Gosto tem muito a contribuir para a preservação da biodiversidade alimentar brasileira, pois permite a troca entre saberes e o acesso a informações importantes a respeito da cultura alimentar tradicional, uma vez que é construído coletivamente, de forma complementar.

Por conta de sua dinâmica de construção, os valores da organização do Conhecimento se fazem fortes aliados à construção do catálogo Arca do Gosto, pois oferece aporte metodológico conveniente para seu desenvolvimento e consequente alcance social.

O alcance social almejado pelo catálogo Arca do Gosto no Brasil, assim como em outros países, diz respeito primeiramente, ao combate à fome. Pois reconhecer e propagar o conhecimento a respeito de determinados hábito alimentar revela as potencialidades do mesmo as populações de regiões onde os alimentos nativos representam a força para sobrevivência, mas estão sendo esquecidos. Ou seja, recupera os saberes tradicionais da cultura alimentar por meio do auxílio da sistematização da informação e da ciência. Por outro lado, da mesma forma, também torna possível que a ciência identifique e se atente aos saberes tradicionais, usando-os a seu favor.

Referências

- Barité, M. (2001). Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: Carrara, K. (org.). *Educação, Universidade e Pesquisa*. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, p.35-60.
- Brasil, M. S. (2015). *Alimentos regionais brasileiros*. Brasília: Ministério da Saúde. Disponível em: http://189.28.128.100/dab/docs/portaldab/publicacoes/livro_alimentos_regionais_brasileiros.pdf
- Brasil, M. S. (2014). *Guia alimentar para a população brasileira*. Brasília: Ministério da Saúde. Disponível em: www.diabetes.org.br/publico/images/pdf/guia-alimentar-para-a-pop-brasiliera.pdf
- Fundo Internacional de Desenvolvimento Agrícola. (2018). *Ecogastronomia para jovens rurais do Semiárido: Compartilhando saberes e sabores locais adotando a filosofia Slow Food nos projetos FIDA*. Sergipe: Fundo Internacional de Desenvolvimento Agrícola (FIDA):

- IICA - Instituto Interamericano de Cooperação para a Agricultura. Disponível em: http://www.slowfoodbrasil.com/documentos/ecogastronomia_jovens_rurais_semiarido.pdf
- García Marco, F. J.(s/d). Los sistemas de información histórica: entre la archivística y la historia. In.: Rodríguez, A. Á. R. *Manual de archivística*. Madri: Editorial Síntesis.
- Hjørland, B.; H. Albrechtsen. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*, v.46, n.6, p.400-425.
- Hjørland, B. (2003). Fundamentals of Knowledge Organization. *Knowledge Organization*, v.30, n.2, p.87-111.
- Makuta, G. (2018). *Biodiversidade, Arca do Gosto e Fortalezas Slow Food: um guia para entender o que são, como se relacionam com o que comemos e como podemos apoiá-las*. São Paulo: Associação Slow Food do Brasil, 2018. Disponível em: http://slowfoodbrasil.com/documentos/slowfood_publicacao.pdf
- Milano, S.; et. al. (s/d). *A Arca do Gosto Como criar o maior catálogo de sabores do mundo: um patrimônio para descobrir e salvar*. Roreto di Cherasco: Stampatello. Disponível em: <http://slowfoodbrasil.com/documentos/slowfood-livreto-arcadogosto.pdf>
- Petrini, C.; et. al. (s/d). *A centralidade do alimento: documento do congresso 2012-2016*. Slow Food Internacional. Disponível em: www.slowfoodbrasil.com/documentos/a-centralidade-do-alimento-carlo-petrini.pdf
- Slow Food Internacional; Fundo Internacional de Desenvolvimento Agrícola. (2015). *Comida com gosto de licuri: receitas*. Bahia: COOPES Cooperativa de Produção da Região do Piemonte da Diamantina. Disponível em: <http://slowfoodbrasil.com/documentos/licuri-livro-de-receitas.pdf>
- Yatzimirsky, M. C. S. (2006). A comida dos favelados. *Estudos avançados*, v.20, n.58, p.123-132. Disponível em: <http://www.scielo.br/pdf/ea/v20n58/12.pdf>

Entre a “fala que narra” e a “fala que demonstra”: Iniciativas em Arquivos de História Oral e o seu valor na construção das memórias e das identidades sociais

Maria Cristina Vieira de Freitas

ORCID [0000-0002-8849-8792](https://orcid.org/0000-0002-8849-8792). Faculdade de Letras, Universidade de Coimbra, Arquivo da
Universidade de Coimbra, Centro de Estudos Interdisciplinares do Século XX - CEIS20,
Portugal.

cristina.freitas@fl.uc.pt

Resumo. Sociedades historicamente datadas experieñciam as suas memórias coletivas ora guiadas pelos mitos ora conduzidas pela razão. O texto fixa e retém a palavra num suporte e impõe padrões de comunicação distintos dos praticados no universo oral. No processo de construção das diversas identidades sociais relevam as relações de poder, produtos das divergências entre “incluídos” e “excluídos”. Este estudo assenta-se nesses pressupostos teóricos. A abordagem é qualitativa, com caraterísticas exploratórias e descritivas. Recorre-se a fontes bibliográficas e documentais. Caraterizam-se cinco iniciativas que tipificam casos de constituição de Arquivos de História Oral, tendo como foco a organização, a representação e o acesso à informação. Os principais resultados apontam para: i) o uso de estratégias semelhantes de organização e de representação da informação, embora em níveis de conformidade diferenciados; ii) o respeito pelas regras impostas pelo direito de acesso aos conteúdos. A construção e a preservação da memória coletiva e a valorização do uso da História Oral servem como justificação para a existência dessas iniciativas. Pelas suas caraterísticas, este estudo limita-se a um conjunto restrito de casos, sugerindo-se o seu alargamento, bem como o aprofundamento das análises.

Palavras-chave: História oral; materiais especiais; organização da informação; representação da informação; memória coletiva.

Abstract. Societies that are historically dated experience their collective memories either guided by myths or led by reason. The text fixes and retains the words on a medium and imposes communication patterns that are different from the ones in the oral universe. In the process of constructing the different social identities, power relations are the product of the differences between the "included" ones and the "excluded" ones. This study is based on these theoretical assumptions. The approach is qualitative, with exploratory and descriptive characteristics. It uses bibliographical and documentary sources. It characterizes five initiatives that typify research cases of constitution of Oral

History Archives, focusing on organization, representation, and access to information. The main results point to: i) the use of similar organization and information representation strategies, although at different levels of conformity; ii) the respect for the rules imposed by the right of access to the contents. The construction and preservation of the collective memory and the valorization of the use of Oral History serve as a justification for the existence of these initiatives. Due to its characteristics, this study is limited to a restricted set of cases, suggesting a future extension and the deepening of the analyzes.

Keywords: Oral history; special collections; information organization; information representation; collective memory.

1 Introdução e enquadramento

Há mais de um conceito para o termo “memória”. Na sua sobejamente conhecida obra, traduzida, publicada e citada por diversos autores/as, em igualmente diversos idiomas, Le Goff (1984) diferencia uma memória individual, que é própria da capacidade humana de conservar a informação e que possui uma índole biológica, de uma outra, a memória social, que se sujeita aos grupos e às influências do meio. Ao procurar exemplos de sociedades historicamente datadas, que experienciam ou experienciaram o acesso à memória social, esse autor investiga, por um lado, as sociedades ágrafas, guiadas pelos “mitos”, e, por outro, as sociedades da escrita, conduzidas pela “razão”. Previamente aos estudos de Le Goff (1984), Halbwachs (1990) já havia fixado e analisado em profundidade o conceito de memória coletiva, admitindo também a existência do seu oposto complementar: a memória individual.

Traduzido como a “fala que narra”, o termo *mythos* designa a forma como a consciência do eu e do outro se expressam no domínio da oralidade. Com efeito, a narrativa do mito é a que decorre de um tempo imemoriável, que fundamenta e fixa modelos exemplares de conduta social. É, pois, a “fala” primitiva e simbólica, acionada por mecanismos mnemônicos. Não se trata de uma explicação racional, destinada a satisfazer uma curiosidade intelectual, mas de uma narrativa que faz reviver uma realidade “primeira”. Esse tipo de consciência coletiva tem um poder esmagador sobre as individualidades, na medida em que suspende os interesses pessoais e submete a pessoa ao grupo, sendo esta a sua característica ontológica mais evidente (Eliade, 1986; Gusdorf, 1979; Cunha, 1992; Malinowski, 1926). E a função desta “fala que narra” é oferecer explicações para os fenômenos e responder a perguntas que de outro modo ficariam sem resposta, projetando uma visão de mundo imersa numa espécie de percepção de grupo que caracteriza o imaginário coletivo tecido no meio social.

Nestas sociedades, governadas pela consciência mítica, segundo Le Goff (1984), existem os guardiões e as guardiãs da memória, ou seja, pessoas autorizadas a preservar e a transmitir os mitos que contribuem para manter a coesão social. E essa transmissão não é feita *ipsis verbis* (Goody como citado em Le Goff, 1984, p. 430),

mas pode atualizar-se e reatualizar-se a cada narrativa, conferindo liberdade e criatividade ao narrador/a. O valor desta transmissão não está, pois, na precisão da narrativa, mas na autenticidade e na fidedignidade, que se sustentam na crença e nos valores de quem narra. Le Goff (1984) se posiciona como um defensor da autenticidade dessa memória que se manifesta na oralidade e enumera as suas principais características, entre as quais, o seu dinamismo, a sua naturalidade, o seu aspeto coletivo e o facto de ser socialmente construída e de ter um carácter dialético.

Igualmente, datado no tempo e no espaço, o vocábulo “*logos*” significa a “fala que demonstra” e que, como tal, descreve, governando-se por uma visão analítica e supostamente destituída de processos emocionais. Neste caso, a textualidade proporcionada pela escrita fixa e retém a palavra num suporte e impõe padrões de comunicação e de expressão que não se coadunam perfeitamente com aqueles que são próprios do contexto da oralidade. Para os seus críticos, essa rigidez e fixidez da palavra atinge o ápice na “Modernidade”, atribuindo-se aos movimentos iluministas e positivistas a crença no poder explicativo da “razão” e a elevação do culto ao documento escrito como fonte *ipsis litteris* da verdade (Cunha, 1992; Thompson, 1992; Le Goff, 1984).

Esses aspetos são enfatizados por Le Goff (1984), para quem o aparecimento da escrita propicia o desenvolvimento de uma “memória da inscrição”, vinculada às datas comemorativas e ao seu registo nos monumentos, ao lado de uma memória que é tangível no documento escrito e que passa a cumprir a dupla função de reter e de atualizar a informação, por meio do reexame e da retificação. Essa transformação tem um significado profundo e consubstancia-se na monumentalização da memória e do documento que, de meio para o ensino (*docere*=ensinar), transforma-se em recurso para a recordação (*monere*=recordar).

Essas transformações têm repercussões na constituição dos “monumentos de lembrança” - museus, bibliotecas e arquivos -, que cumprem a função de servir aos interesses da mentalidade de uma época e testemunham um projeto de civilização. No limite, a memória construída nas, e pelas, “sociedades da escrita” é “sequestrada” do espaço social e mantida como refém nesses lugares destinados à sua preservação, de uma forma não ritual e pouco identitária, visitando-se os “lugares como meios de memória” (De Decca, 1992; Nora, 1993).

Essa distorção identitária tem consequências imediatas na desconexão das pessoas em relação aos bens que deveriam reconhecer como “o seu” património. As instituições, por seu turno, ao constituírem as suas próprias políticas de proteção desse património não levam em conta as identidades ou não respeitam as diferenças e os valores que enformam os grupos sociais. Estes, por seu turno, acabam por não querer ou não ter um acesso igualitário ao

patrimônio que representa “a sua” cultura. Por conseguinte, o que se exprime como “patrimônio” e se constitui como um referencial identitário vem a ser fruto de uma apropriação indébita de um conceito, por uma minoria que se sobrepõe e silencia os interesses e as expectativas de uma maioria (Magnani, 1986; Paoli, 1992).

Essa apropriação nada tem do sentido original de “herança” ou de “bem comum”. Para que o patrimônio possa imbuir-se de um significado cultural, ele deve exprimir-se por meio de valores de ordem afetiva, cognitiva, formal e pragmática (Meneses, 1992). E o ato de “autonomização do sujeito”, que permite uma tomada de consciência acerca dos valores que enformam a sua própria identidade social, é um ato de “iluminação ética” (Morin, 1996). Todo esse processo de reconhecimento é, em si, um processo identitário.

Castells (1999, p. 22), define a identidade social como “o processo de construção de significado com base em um atributo cultural, ou ainda, um conjunto de atributos culturais inter-relacionados, o(s) qual(is) prevalece(m) sobre outras fontes de significado”. Esses atributos constituem-se no patrimônio cultural. E no processo de construção das identidades sociais relevam, e já o dissemos, as relações de poder, que são produtos das divergências entre os “incluídos” e os “excluídos” desse processo. Nesta contenda forjam-se, segundo o mesmo Castells (1999), diferentes tipos de identidade, consoante os papeis e a percepção do grau de inclusão ou de exclusão do processo social: i) identidade legitimadora, que justifica, expande e racionaliza o poder das instituições dominantes; ii) identidade de legitimação, que se constrói em torno das pessoas excluídas do processo social, como uma forma de resistir à dominação; e iii) identidade de projeto, que se edifica em torno de um ideal comum e com vista a uma mudança social.

2 Metodologia

2.1 Objetivos, fontes de dados e procedimentos

No presente estudo exploramos e descrevemos iniciativas de constituição de Arquivos de História Oral [AHO] e discutimos a sua relevância na construção da(s) memória(s) e da(s) identidade(s) social(is).

A abordagem é qualitativa, quer pela forma como são selecionadas as amostras, intencionais e não probabilísticas (iniciativas nacionais e internacionais que se constituam em arquivos de história oral com presença na *Web*), quer pela perspectiva analítica, recaindo o interesse nas características, na variabilidade e nos significados implicados.

Neste estudo, recorremos a fontes bibliográficas e documentais. Primeiramente, selecionamos textos de índole científica, úteis à realização de um enquadramento do tema. Na escolha desses recursos, privilegiamos alguns textos considerados clássicos sobre a matéria. Seguidamente, selecionamos iniciativas que se constituem em AHO,

visando a sua caracterização, em aspetos relacionados com a constituição, a organização e a representação da informação. Como desdobramento lógico, intentamos responder à seguinte questão de investigação: se, e de que modo, as estratégias de organização e de representação da informação contribuem para a construção e a preservação da memória e do património nos chamados AHO.

3 Arquivos de História Oral: enquadramento e iniciativas

De acordo com Alberti (1999), o sucesso da História Oral [HO], como método, é um fenómeno bastante conhecido no meio académico, sendo, no entanto, relativamente escassos os estudos que aliam a investigação à preservação das fontes orais.

As origens sistemáticas da HO, segundo recorda Thompson (1992), situam-se nos Estados Unidos, no final da Segunda Grande Guerra, quando o seu uso terá sido impulsionado com o intuito de perceber, através de relatos feitos na primeira pessoa, os problemas e as angústias enfrentadas pelos soldados envolvidos nas frentes de batalha.

Conforme expressam Freitas e Silva (2016), na literatura científica internacional e de âmbito arquivístico, há uma tendência para o crescimento dos estudos dedicados à valorização da memória coletiva e do papel das comunidades no desenvolvimento, juntamente com os arquivos, de coleções que revelem as suas identidades, sendo esse um aspeto definidor do seu “*empowerment*”, ou seja, da sua autonomia e da sua emergência enquanto comunidade (Morin, 1996). Todo esse pluralismo compreende-se melhor à luz do respeito pelo direito à diversidade e, consequentemente, da valorização das distintas identidades que compõem o tecido social (Castells, 1999; Morin, 1996).

Na atualidade, segundo as nossas pesquisas, os exemplos de AHO são variados, pelo que abordaremos apenas algumas dessas iniciativas, procurando realçar as suas particularidades e captar alguma diversidade eventualmente existente.

A primeira iniciativa a reportar é a que desenvolve o *National Archives and Records Administration* [NARA]¹. Trata-se de uma página alojada no sítio *Web* da instituição, para a qual são vertidas as transcrições das várias entrevistas realizadas com personalidades do meio arquivístico ligadas à instituição. Estas entrevistas são o resultado de um Projeto intitulado *National Archives Oral History* [NAOHP], desenvolvido com o intuito de recolher os depoimentos dos seus atuais e antigos colaboradores. Os conteúdos desses depoimentos versam sobre experiências, ideias e perspetivas das pessoas entrevistadas, no âmbito do seu trabalho em arquivos e no contexto em que o mesmo se desenvolveu.

1 Disponível em: <https://www.archives.gov/about/history/oral-history-at-the-national-archives>.

Com esse projeto, e segundo na própria página se indica, o NARA (2019) pretende compreender a sua própria cultura organizacional, por meio da identificação de práticas, tomadas de decisão, ações e eventos históricos e preservar uma parte da memória institucional. Trata-se de um projeto incremental, interativo e colaborativo, em que a comunidade pode sugerir novas pessoas a entrevistar e novos/as entrevistadores/as. Quanto à organização, a informação está classificada primeiramente pelos tipos de entrevistas e/ou entrevistadores e, seguidamente, ordenada pelos nomes das pessoas entrevistadas, dispondo-se estes na ordem direta. Os conteúdos das entrevistas encontram-se resumidos e, em alguns casos, são acompanhados pela fotografia da pessoa entrevistada. As gravações das entrevistas não se encontram, atualmente, disponíveis nessa página. As transcrições incluem os correspondentes termos de autorização para a comunicação do conteúdo. Presentemente, na página oficial do projeto contam-se 40 entrevistas realizadas entre os anos de 1972 e 2017, com personalidades relevantes para a história da Arquivística norte-americana e internacional (e.g., Lester Cappon, Oliver Holmes, Ernst Posner e Philip Bauer) e com arquivistas menos conhecidos/as do público específico. Os conteúdos são reveladores de aspetos singulares da sua carreira profissional e partilhados em livre acesso, podendo ser utilizados em investigações. Não se observam campos descritivos normalizados nas descrições apresentadas (figura 1).



Figura 1. Instantâneo da página do projeto de HO do NARA.

Fonte. Disponível em: <https://www.archives.gov/about/history/oral-history-at-the-national-archives>.

A segunda iniciativa a descrever é o projeto desenvolvido pelo *The Scottish Oral History Centre* [SOHC] que, segundo as informações facultadas pela instituição (SOHC, 2019), estabeleceu-se no *Department of History of University of Strathclyde* (Escócia), no ano de 1995, com o objetivo de fomentar atividades de ensino, pesquisa e extensão, no âmbito do uso da HO. Ainda, segundo se informa (SOHC, 2019), os recursos que compõem o fundo situam-se no período de 1981 até ao presente e em termos de descrição física ocupam 220 GB de espaço de armazenamento. Trata-se, pois, de um arquivo composto por recursos digitais. Pelos seus dados descritivos normalizados, verificamos que o *The Scottish Oral History Centre Archive* [SOHCA],

instituição encarregada da organização, da representação e da comunicação dessa iniciativa, dispõe de uma ampla coleção de depoimentos cujos conteúdos, segundo as informações descritivas disponibilizadas, encontram-se associados à história do trabalho, da saúde ocupacional e dos impactos sociais provocados pelo fenómeno da desindustrialização na Escócia² (SOHCA, 2019). Os recursos encontram-se organizados num primeiro nível em função dos projetos desenvolvidos pelos grupos de investigação, ordenados sequencialmente. Dentro de cada projeto, os dossiês ou processos são ordenados obedecendo a uma codificação numérica e sequencial, combinando-se esta com a referenciação dos nomes das pessoas entrevistadas, atribuídos em ordem direta, seguidos, em alguns casos, da sua data de nascimento ou da sua profissão. Diversos dos conteúdos são considerados sensíveis, pelo que o acesso em alguns casos pode não ser possível, estando esta restrição devidamente sinalizada. O acesso aos diferentes projetos, às descrições e, em alguns casos, às respetivas transcrições das entrevistas, é dado na zona destinada à indicação da existência de documentação relacionada, no formulário de descrição do fundo, disponibilizando-se os respetivos enlaces em forma de lista. A descrição realiza-se em campos que seguem as normas internacionais aconselhadas pelo Conselho Internacional de Arquivos [CIA], verificando-se o uso do *software* livre *Access to Memory* [AtoM] para o efeito (figura 2).

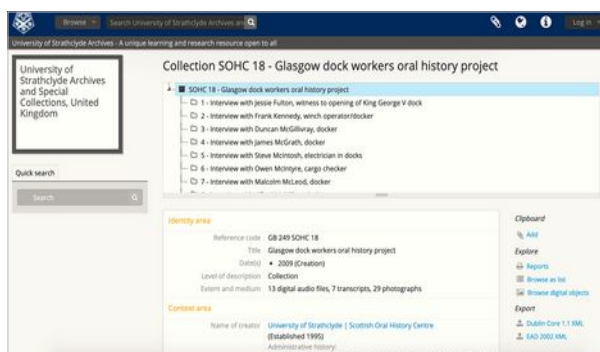


Figura 2. Instantâneo da descrição de um dos projetos de HO desenvolvidos pelo SOHCA.

Fonte. Disponível em: https://atom.lib.strath.ac.uk/glasgow-dock-workers-oral-history-project;isad?sf_culture=en.

A terceira iniciativa a indicar é a da *University of Yale* [UY], composta por 254 caixas contendo gravações e transcrições produzidas, segundo descreve a própria instituição (UY, 2019)³, no âmbito do projeto designado *Yale-United Nations Oral History Project* [YUNOHP], que documenta eventos considerados significativos para a história da UNESCO. As entrevistas são

² Disponível em: https://atom.lib.strath.ac.uk/sohc-archive;isad?sf_culture=pt.

³ Disponível em: <https://archives.yale.edu/repositories/12/resources/3915#>.

realizadas com personalidades proeminentes de vários países (políticos/as, embaixadores/as, diplomatas, etc.), para além de outros/as participantes. As datas-limite das entrevistas coligidas, segundo informa a UY (2019), são 1983 e 2005, não se prevendo ingressos adicionais. O acesso físico às cópias de uso desses recursos é livre, desde que sem fins comerciais. Pelo que se verifica, a coleção se encontra descrita, mas os recursos correspondentes não se encontram disponíveis para o livre acesso *online*. Quanto à organização, verifica-se, primeiramente, a classificação da informação pelo gênero documental (*Transcripts; Tapes*) e, seguidamente, a ordenação alfabética dos recursos pelo nome do/a entrevistado/a, na forma inversa. O *software* usado para a divulgação da coleção é o *ArchiveSpace*⁴, de uso livre (figura 3).

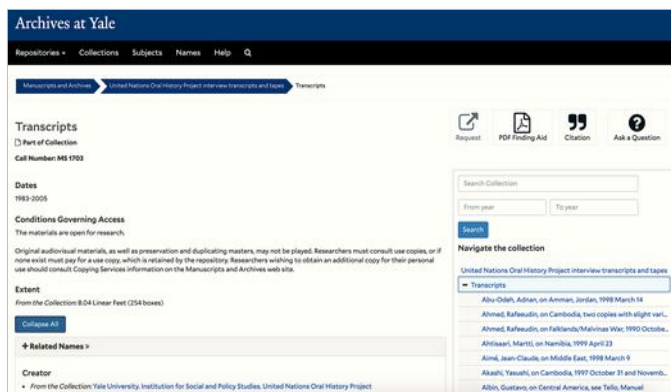


Figura 3. Instantâneo da descrição do projeto YUNOHP.

Fonte. Disponível em: https://archives.yale.edu/repositories/12/archival_objects/1256540.

A quarta iniciativa a realçar é a da *Oral History and Folklore Collection* [OHFC], desenvolvida pela *National Library of Australia* [NLA]. Na sua página de divulgação⁵, a NLA (2019) refere que o principal objetivo a cumprir com o desenvolvimento desta coleção é recolher uma série de registos de distintas manifestações artísticas e culturais do país (estórias, músicas, danças, etc.), para além dos depoimentos recolhidos de personalidades reconhecidas (cientistas, escritores, políticos, artistas, etc.) e de pessoas que vivenciam diversas situações e/ou representam diversas tendências sociais (imigrantes, refugiados, desempregados, etc.). A coleção, cujas gravações mais antigas datam de 1950, reflete a vida social, cultural e intelectual do país e tem uma extensão, segundo se informa, de 45.000 horas gravadas, encontrando-se aberta à inclusão de novos itens (NLA, 2019). Verificamos que os recursos estão descritos, no catálogo, item por item, encontrando-se disponíveis os respetivos registos bibliográficos. Também verificamos, seguindo o

4 Segundo a informação disponível no *Website* do *ArchiveSpace*. A plataforma do projeto não dá os devidos créditos no rodapé. Disponível em: <https://archivesspace.org/community/whos-using-archivesspace>.

5 Disponível em: <https://www.nla.gov.au/what-we-collect/oral-history-and-folklore>.

enlace que dá acesso direto aos itens da coleção, que os conteúdos se encontram reunidos num único ponto de acesso. As entrevistas, segundo os dados informados (NLA, 2019), são descritas como recurso integrante dos projetos desenvolvidos ou como itens individuais da coleção, se o caso não é o anterior. Algumas entrevistas vêm acompanhadas por um resumo não estruturado e de tamanho variável, incluído também no catálogo. As gravações e as transcrições também podem ser acedidas pelo mesmo catálogo, juntamente com as descrições às quais se encontram associadas. O acesso é aberto para a pesquisa e o uso pessoal e não comercial, na maioria dos itens da coleção. Naturalmente, alguns desses itens são de acesso reservado (figura 4).



Figura 4. Instantâneo da descrição das entrevistas na coleção de HO da NLA.

Fonte. Disponível em: https://catalogue.nla.gov.au/Search/Home?lookfor=&filter%5b%5d=access_type%3A%22All%20online%22&type=Array&narrow=1&filter%5b%5d=format%22Audio%22&sort=sort_date_asc&sort=sort_date_desc&page=1&sort=&page=1&sort=sort_date_desc&page=1&sort=sort_author_asc&page=1.

A quinta iniciativa a mencionar é a *Oral History Collections* [OHC] constituída no âmbito do *Center for Oral History* [COH], pertencente ao *Science History Institute* [SHI]. Segundo informa o próprio SHI (2019)⁶, trata-se de uma coleção que contém centenas de entrevistas consideradas relevantes para o estudo de diversos temas associados à história da ciência, da medicina e da tecnologia. As entrevistas começaram a ser produzidas no ano de 1979. Entre as personalidades ilustres, que figuram no rol de entrevistados/as, encontram-se cientistas laureados/as com o Prêmio Nobel (*e.g.*, Alan MacDiarmid e David Baltimore) e outros/as que foram medalhados/as pela *National Medal of Science*. As gravações e as transcrições podem ser de acesso livre ou restrito, dependendo da sensibilidade dos conteúdos. Para obter as transcrições das entrevistas que se encontram em livre acesso, deve-se preencher um rápido formulário. O *download* é gratuito e imediato. Nesta

6 Disponível em: <https://www.sciencehistory.org/oral-history-collections>.

coleção, segundo se observa, as entrevistas estão organizadas individualmente ou no âmbito de projetos de ciência. No primeiro caso, estão ordenadas alfabeticamente pelo nome da pessoa entrevistada, disposto em ordem direta. No segundo caso, a ordenação também é a alfabética, mas pelo nome do projeto (figura 5). O *software* usado para a divulgação, segundo apuramos, é também o *ArchiveSpace*⁷.



Figura 5. Instantâneo da descrição de um dos projetos desenvolvidos pelo SHI.

Fonte. Disponível em: <https://oh.sciencehistory.org/oral-histories/projects>.

4 Discussão e conclusões

Este estudo explorou, em linhas gerais, o tríplice e historicamente datado conceito de memória social, que é construído, segundo Le Goff (1984), no contexto da oralidade (dinâmica, social e inclusiva) e no contexto da textualidade (estática, individual e exclusiva). Trata-se de manifestações cíclicas experienciadas em diferentes momentos e realidades sociais. A esse conceito tridimensional associam-se as tipificações fornecidas por Castels (1999), para quem as identidades sociais manifestam-se também em diferentes aceções (legitimadora, de legitimação e de projeto), dependendo das relações de poder que se estabelecem e que resultam, na sua constituição, nos processos de inclusão ou, contrariamente, de exclusão dos diversos grupos sociais.

Este estudo também identificou diversas iniciativas de constituição de coleções de HO, em desenvolvimento em diferentes países, desde a década de 1950. Dessas iniciativas, foram selecionadas apenas cinco, usadas para tipificar casos, apoiando-se no critério da variabilidade, com a expectativa de atingir alguma riqueza na descrição e na análise dos resultados (quadro 1).

7 Segundo a informação disponibilizada no *Website* deste *software*, uma vez que a plataforma não dá os devidos créditos no seu rodapé. Disponível em: <https://archivesspace.org/community/whos-using-archivesspace>.

Quadro 1. Comparativo das informações descritivas facultadas pelos projetos de AHO selecionados

Identificação, localização e datas-limite	Principais objetivos	Organização/ representação da informação	Acesso e divulgação da informação
National Archives and Records Administration. EUA. 1972 - .	Documentar as experiências em torno do trabalho em arquivos. Compreender a cultura organizacional e preservar a memória institucional.	Pelos tipos de entrevistas/entrevistadores/as e pelo nome do/a entrevistado/a, em ordem direta; oferece resumos sinaléticos.	Acesso livre e <i>online</i> . <i>Webpage</i> alojada em <i>Website</i> oficial da instituição.
The Scottish Oral History Centre Archive, University of Strathclyde. Reino Unido (Escócia). 1981- .	Fomentar atividades de ensino, pesquisa e extensão, no âmbito da HO. Documentar a história do trabalho, da saúde ocupacional e os impactos sociais do fenómeno da desindustrialização.	Pelo nome das pessoas entrevistadas, em ordem direta; oferece resumos sinaléticos.	Acesso livre e <i>online</i> , com restrição para os conteúdos sensíveis. <i>Software</i> AtoM.
Yale-United Nations Oral History Project, University of Yale. EUA. 1983 - 2005.	Documentar eventos significativos para a história da UNESCO.	Pelo gênero documental e pelo nome das pessoas entrevistadas, em ordem inversa.	Acesso livre e local às cópias de uso. Acesso <i>online</i> às descrições arquivísticas. <i>Software</i> <i>ArchiveSpace</i> .
National Archives Oral History, National Library of Australia. 1950 -	Registrar distintas manifestações artísticas e culturais do país, pela voz de personalidades reconhecidas e de pessoas que vivenciam situações e/ou representam tendências sociais.	Pelo nome das pessoas entrevistadas, em ordem direta, ou pelo projeto; alguns resumos sinaléticos são oferecidos.	Acesso livre e <i>online</i> , com restrição para os conteúdos sensíveis. <i>Webpage</i> de divulgação da coleção mantém ligação ao catálogo bibliográfico.
Center for Oral History, Science Institute. EUA. 1979 - .	Registrar os relatos de personalidades relevantes para o estudo da história da ciência, da medicina e da tecnologia.	Pelo nome das pessoas entrevistadas, em ordem direta, ou pelo projeto; alguns resumos sinaléticos são oferecidos.	Acesso livre e <i>online</i> , mediante preenchimento de formulário para permissão de download imediato. Restrição de acesso aos conteúdos sensíveis. <i>Software</i> <i>ArchiveSpace</i> .

Fonte. Dados produzidos pela investigação.

A partir do quadro 1, verificamos que algumas dessas iniciativas são encabeçadas por bibliotecas, outras por arquivos universitários ou nacionais, outras, ainda, por centros de estudos e de investigação. A maioria destes projetos está em curso (NLA; SHI; US; NARA) e possui uma longevidade compatível com o próprio aparecimento da HO enquanto método de investigação histórica. Maioritariamente, as iniciativas selecionadas encontram-se nos EUA, país que é considerado o berço do movimento pelo uso da HO como método histórico. Algumas das instituições que abrigam estas iniciativas são de âmbito público e nacional, outras são ligadas a universidades e outras, ainda, são organizações sem fins lucrativos. Essa relativa variabilidade não foi conferida aleatoriamente. Contrariamente, atendeu aos imperativos metodológicos do estudo.

Quanto aos objetivos, constatamos que há projetos dedicados a uma realidade institucional (NARA; UY), ao lado de projetos desenvolvidos com fins que ultrapassam esses limites, reportando-se ao contexto envolvente (NLA; SHI; US). Os objetivos enunciados oferecem como tronco comum a necessidade de documentar relatos, eventos, experiências e manifestações, tendo como perspectiva as pessoas envolvidas e como pano de fundo a memória coletiva, na aceção de Halbwachs (1990), e as diferentes identidades que participam no seu processo de construção, que importa documentar e preservar (quadro 1).

Nos AHO analisados, de um modo geral, notam-se mais semelhanças do que diferenças na forma de comunicar a informação. Assim, verifica-se que a maioria dos conteúdos, disponibilizados em plataformas gratuitas, são de acesso aberto e *online*, desde que não haja nenhuma restrição imposta (*i.e.*, dados sensíveis). Quanto à organização e à representação da informação, podemos subdividir essas iniciativas em dois grupos. Há, por um lado, projetos com uma conceção menos complexa (NARA; NLA) e que, entretanto, encontram-se razoavelmente estruturados, na medida em que o/a utilizador/a descobre rapidamente a informação que pretende e compreende facilmente a forma de organização e de representação dos conteúdos. Por outro lado, há projetos um pouco mais robustos, encabeçados por Universidades e por Centros de Investigação, cujos objetivos são também, e principalmente, de natureza científica e que, por isso mesmo, recorrem a um registo diferente (UY; US; SHI). Trata-se de iniciativas com uma projeção longitudinal, que lidam com um volume de documentos mais substantivos, constituídos em torno de projetos de investigação, extensão e/ou ensino, cujos resultados são disponibilizados pelos respetivos arquivos institucionais, amparando-se em tecnologias e *softwares* conformes e adotando estruturas de dados descritivos que obedecem às normas e às convenções internacionais arquivísticas para a organização e a representação da informação (quadro 1). Nota-se, ademais, na conceção desses projetos, uma forte influência das diretrizes emanadas, no passado, do estudo pertencente à série RAMP, produzido por Moss e Mazikana (1986) e dedicado aos AHO.

Pelo exposto, e respondendo à questão inicial, conclui-se que as estratégias de organização e de representação da informação adotadas por estas instituições, ainda que em diferentes níveis de maturidade, oferecem um interessante e relevante

contributo para a construção e a preservação da memória e do património, na medida em que disponibilizam, *online*, um conjunto de conteúdos credíveis, utilizáveis e reutilizáveis. Estas iniciativas também respeitam as regras impostas pelo direito de acesso, na medida em que os conteúdos considerados sensíveis, e que pelo que pudemos apurar serão a minoria dos casos observados, não se encontram disponíveis. Também, e naturalmente, são recolhidas as devidas autorizações para a divulgação do material produzido.

Como recomendação final e tendo em conta que, pelas suas características, este estudo limita-se a um conjunto restrito de casos, sugere-se no futuro, o seu alargamento, bem como o aprofundamento da análise de outros aspetos relativos à organização e à representação da informação não abordados por este estudo.

Referências

- Alberti, V. (1999). História oral e arquivos (p.31-38). In: Z. L. Silva, *Arquivos, património e memória: trajetórias e perspectivas*. Marília: UNESP.
- Castells, M. (1999). *O poder da identidade* (vol. 2). São Paulo: Paz e Terra.
- Cunha, J. A. (1992) *Filosofia: iniciação à investigação filosófica*. São Paulo: Atual.
- De Decca, E. S. (1992). Memória e cidadania. In: Secretaria Municipal de Cultura de São Paulo. Departamento de Patrimônio Histórico. *O direito à memória: patrimônio histórico e cidadania* (p. 129-138). São Paulo: DPH.
- Eliade, M. (1986). *Mito e realidade*. São Paulo: Perspectiva.
- Freitas, M. C. V. de, & Silva, C. G. da. (2016). O novo e o atual na Arquivística internacional: a desmaterialização, a interoperabilidade, a organização e o uso da informação em evidência (2011-2016). In *Actas do XII Encontro de Arquivos Municipais: Arquivos municipais: o que há de novo?* Castelo Branco: APBAD. Disponível em: <http://www.bad.pt/publicacoes/index.php/arquivosmunicipais/article/view/1560/1470>.
- Gusdorf, G. (1979). *Mito e metafísica*. São Paulo: Convívio.
- Halbwachs, M. (1990). *A memória coletiva*. São Paulo: Vértice.
- Le Goff, J. (1984). *História e memória*. Campinas: UNICAMP.
- Magnani, J. G. C. (1986). Pensar grande o patrimônio cultural. *Lua Nova*, 3(2), 62-67.
- Malinowski, B. (1926). *Myth in primitive psychology*. New York: W.W. Norton & Company.

- Meneses, U. B. de. (1992). O patrimônio cultural entre o público e o privado. In: Secretaria Municipal de Cultura do Estado de São Paulo. Departamento de Patrimônio Histórico. *O direito à memória: patrimônio histórico e cidadania* (p.189-96). São Paulo: DPH.
- Morin, E. (1996). *A noção de sujeito: novos paradigmas, cultura e subjetividade*. São Paulo: Artes Médicas.
- Moss, W., & Mazikana, P. (1986). Archives, oral history and oral tradition: a RAMP study. Paris: UNESCO.
- National Archives and Records Administration - NARA. (2019). *Oral history at National Archives*. Disponível em: <https://www.archives.gov/about/history/oral-history-at-the-national-archives>.
- National Library of Australia - NLA. (2019). *Oral History and Folklore*. Disponível em: <https://www.nla.gov.au/what-we-collect/oral-history-and-folklore>.
- Nora, P. (1993). Entre história e memória: a problemática dos lugares. *Revista Projeto História*, 10, 7-28.
- Paoli, M. C. (1992). Memória, história e cidadania: o direito ao passado. In: Secretaria Municipal de Cultura do Estado de São Paulo. Departamento de Patrimônio Histórico. *O direito à memória: patrimônio histórico e cidadania* (p. 25-28). São Paulo: DPH.
- Science History Institute – SHI. (2019). *Oral History Collections*. Disponível em: <https://www.sciencehistory.org/oral-history-collections>.
- Scottish Oral History Centre, The – SOHC. (2019). *Collection SOHC: Scottish Oral History Centre Archive*. Disponível em: https://atom.lib.strath.ac.uk/sohc-archive;isad?sf_culture=en.
- Thompson, P. (1992). *A voz do passado: história oral*. Rio de Janeiro: Paz e Terra.
- University of Yale – UY (2019). *United Nations Oral History Project: interview transcripts and tapes*. Disponível em: <https://archives.yale.edu/repositories/12/resources/3915>.

A promoção da experiência dos utilizadores como um estímulo à aprendizagem em museus: um estudo exploratório

André P. Pacheco

ORCID [0000-0002-1810-4866](https://orcid.org/0000-0002-1810-4866). Faculdade de Letras da Universidade de Coimbra, Portugal.
andrez.pacheco@gmail.com

Resumo. O desenvolvimento e expansão das redes comunicacionais desde os meados do século XX estimularam a consolidação de uma Ciência de Informação, cujo objeto é a informação social. As instituições culturais, como os museus, são vistos enquanto serviços de informação que devem gerir os seus objetos informacionais de modo a comunicar eficiente e apelativamente o significado dos seus artefactos aos visitantes, proporcionando-lhes uma experiência de aprendizagem enriquecedora e memorável. No entanto, a experiência dos utilizadores é normalmente vista sob uma perspectiva da conceção de tecnologias. Este estudo procura alargar o conceito de experiência ao contacto do visitante com o espaço e contexto do museu. Para tal, procede-se a uma revisão de literatura qualitativa dos artigos mais relevantes que alimentam um estudo exploratório. Conclui-se que os museus devem procurar afirmar-se como locais promotores de experiências pessoais de forma a promover o enriquecimento cultural dos visitantes. O sistema tecnológico de informação é apontado como o elemento mediador entre a informação dos objetos e o conhecimento dos visitantes.

Palabras clave: *user experience*, mediação, museus, ciência da informação.

Abstract. The growth and expansion of communication networks since the middle of the 20th century promoted the consolidation of an Information Science whose object is social information. Cultural institutions, such as museums, are posited as information services that must manage their information objects in a way that communicates efficiently and appealingly the meaning of their artifacts to the visitors, ensuring a memorable, pleasant and knowledgeable experience. However, user experience is normally understood of systems design. This study seeks to expand the concept of experience to the visitor's contact with the space and context of a museum. A qualitative literature review of the most relevant papers allows for an exploratory study. It is concluded that museums should be consolidated as spaces that promote strong personal experiences in order to enhance the learning of visitors. The

information technology system is argued to be the mediating element between the information in objects and visitor's knowledge.

Keywords: user experience, mediation, museums, information science.

1 Introdução

Os museus têm um papel fortemente experiencial pautado fortemente por sensações, envolvimento e um sentido estético (Holbrook, 2000; Hassenzahl & Tractinsky, 2006). Esta dimensão é essencial para a aprendizagem dos visitantes porque as pessoas necessitam de uma representação dinâmica da história para melhor compreender a influência dos eventos passado na experiência pessoal). A investigação na experiência dos utilizadores tem dado atenção a estas várias necessidades abordadas pela tecnologia, tais como o prazer (Blythe *et al.*, 2003), a estética (Lavie & Tractinsky, 2004) e o envolvimento cognitivo (Agarwal & Karahanna, 2000). O enfoque é usualmente colocado nas dimensões emocionais, hedónicas e estéticas da experiência (Hassenzahl & Roto, 2007). No entanto, a investigação na experiência dos utilizadores é predominantemente orientada por uma matriz tecnológica de desenvolvimento das aplicações informáticas, tais como a criação de websites apelativos (Pallud & Straud, 2014). Além dos raros desenvolvimentos das noções de experiência dos utilizadores na realidade dos museus, há uma notória ausência na literatura científica de trabalhos que abordem este conceito de uma forma sistémica, enquanto forma de mediação e aproximação entre os significados dos objetos e a aprendizagem dos visitantes. Neste sentido, este estudo é motivado pelas seguintes questões de investigação:

- Pode a promoção de uma experiência positiva, enquanto um possível indicador de aumento do conhecimento do público-alvo, ser encarada como um fator-chave no sucesso de um sistema de informação museológico?
- Pode o conceito de 'experiência dos utilizadores' ser expandido para incluir não só o desenvolvimento das aplicações tecnológicas, mas antes para referir a experiência como um todo?
- Analogamente, pode o conceito de 'mediação' superar o sentido estritamente comunicacional entre humanos, para ser entendido ao nível da comunicação entre a informação de um artefacto e a sua receção por um observador?

2 Metodologia

A metodologia adotada assenta num estudo qualitativo de natureza exploratória da literatura (Gil, 2008). A revisão narrativa da literatura (Bryman, 2016) tem o objetivo de contextualizar o estudo nas correntes mais dominantes da literatura de forma a enquadrar e enriquecer a discussão dos conceitos extraídos. Foram seleccionados os textos considerados mais relevantes para as respetivas áreas de foco, nomeadamente a

Ciência da Informação, a Gestão da Informação, a Teoria da Comunicação, a teoria sistêmica e a experiência dos utilizadores, sob uma perspectiva museológica. A pesquisa de obras sobre estas temáticas foi realizada nas bases de dados de referência Web of Science, Scopus, E-LIS, *Directory of Open Access Journals* (DOAJ) e ProQuest, recuperando apenas artigos científicos nas áreas da ciência da informação (e similares, como *library science*).

3 O enfoque informacional

O desenvolvimento e disseminação das redes comunicacionais, após a II Guerra Mundial, motivaram uma expansão sem precedentes do volume e diversidade dos recursos informacionais existentes. Neste contexto, a partir dos anos 50 do século XX, assiste-se ao aparecimento de vários estudos sobre uma nova Ciência da Informação (Borko, 1968; Brookes, 1980; Silva & Ribeiro, 2008), como reflexo do valor central que a informação passa a ocupar na sociedade, a ponto de alguns autores defenderem que vivemos numa sociedade da informação (Capurro & Hjørland, 2003; Castells, 2002).

No âmbito deste trabalho, a informação social, enquanto objeto de estudo científico, é entendida como um “conjunto estruturado de representações mentais codificadas (símbolos codificantes) socialmente contextualizadas e passíveis de serem registadas num qualquer suporte material (papel, filme, banda magnética, disco compacto, etc) e, portanto, comunicadas de forma assíncrona e multidirecionada” (Silva & Ribeiro, 2008, p. 37). Por extensão, posicionamo-nos num entendimento da Ciência da Informação enquanto uma ciência de eminência social e autónoma, embora com um núcleo formado pela Biblioteconomia e pela Arquivística, enriquecida por contributos interdisciplinares a respeito do estudo da informação, tais como a Psicologia, a História e a Comunicação (Le Coadic, 1996), que “investiga os problemas, temas e casos relacionados com o fenómeno info-comunicacional perceptível e cognoscível através da confirmação ou não das propriedades inerentes à génese do fluxo, organização e comportamento informacionais (origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão e utilização da informação)” (Silva, 2006, pp. 140-141).

Neste contexto, mais do que conservar artefactos, os museus devem abordar a gestão da informação sob a perspectiva do fluxo informacional, que é visto como um sistema no qual participam vários agentes, desde os curadores aos utilizadores do museu. A finalidade dos serviços de informação é gerar produtos que sejam utilizados por indivíduos, quer internos — funcionários — quer externos — a comunidade-alvo. Por isso, mais do que uma abordagem centrada na análise e desenvolvimento dos meios tecnológicos de acesso à informação, deve promover-se o “impacto da gestão da informação na

melhoria do desempenho individual e coletivo das diversas comunidades, à escala local, nacional e global” (Marques, 2017, p. 61), espelhando uma visão na qual o proveito das pessoas, enquanto destinatárias dos serviços de informação, é a referência central para o bom desempenho desses mesmos serviços.

3.1 A mediação informacional

A nível dos museus, entendemos a mediação como o processo de comunicação que se dá entre um objeto em exposição e o olhar atento do visitante que recebe essa informação e a assimila potencialmente como conhecimento. Assim, o estudo da comunicação, enquanto um processo, não pode ser fragmentado. Todos os seus elementos têm de ser estudados em conjunto, p. o emissor, a mensagem, o canal e o recetor. Todos eles interagem numa relação de interdependência, exigindo uma abordagem sistémica ao processo infocomunicacional (Silva, 2009).

4 O museu como um sistema de informação

4.1 Caracterização do sistema

Um museu é uma instituição de carácter permanente, com ou sem personalidade jurídica, sem fins lucrativos, dotada de uma estrutura organizacional que lhe permite: a) garantir um destino unitário a um conjunto de bens culturais e valorizá-los através da investigação, incorporação, inventário, documentação, conservação, interpretação, exposição e divulgação, com objetivos científicos, educativos e lúdicos; b) facultar acesso regular ao público e fomentar a democratização da cultura, a promoção da pessoa e o desenvolvimento da sociedade (Portugal, 2004).

Apesar de a lei se referir à gestão dos artefactos físicos, entende-se o museu como uma instituição cuja missão é gerir e divulgar a informação sob a sua alçada, através do cumprimento dos procedimentos acima mencionados. Por conseguinte, a museologia afirma-se como uma disciplina inserida na epistemologia da Ciência da Informação. Alguns autores têm dado importância aos aspetos partilhados entre museus, bibliotecas e arquivos, suas disciplinas próximas (Marques, 2010, p. 27), enquanto que outros (Bellotto, cit. por Silva, 2002, p. 578) têm apontado as suas diferenças.

Os museus têm um aspeto singular dentro da Ciência da Informação: a posse de coleções de objetos bi/tridimensionais e exemplares únicos, reunidos pelo conteúdo ou pela função e produzidos pela atividade humana ou pela Natureza, com fins artísticos, culturais e funcionais. A sua gestão evoca uma abordagem sistémica, o que permite entender o museu como um sistema de informação.

Quanto ao seu objeto, certos autores consideram que os objetos dos museus são documentos porque contêm informação, formulando os museus como instituições

documentais (Briet, 1951; Otlet, 1934). Discorda-se desta posição porque se entende que um documento pressupõe a informação registada num suporte com uma intenção comunicativa, pelo que um esqueleto ou um vaso, não tendo uma intenção comunicativa a presidir à sua criação, não seriam documentos. Um quadro ou uma fotografia, por sua vez, já o podem ser, na medida em que tiveram um autor humano que materializou as suas ideias num suporte que permite comunicar essa informação a outros. Assim, defende-se a posição de Mensch (1990) e adota-se o uso do termo ‘artefactos’ para referir o objeto dos museus. Embora alguns destes objetos não resultem da atividade humana, ou não tenham uma intenção comunicativa, eles podem ser entendidos enquanto portadores de informação, uma vez que os diferentes tipos de conhecimento (histórico, social, artísticos, geográfico, etc) são suscitados pelo tipo de informação do objeto (Maroevic, 2006).

4.2 Modelos de comunicação em museus

Ao colocarmos o enfoque no valor informacional, estamos a afirmar que os museus não existem por causa dos objetos que contêm, mas sim devido aos conceitos ou ideias que esses objetos ajudam a veicular (Sola, 1986, p. 15). O mais importante é a informação de que se revestem, sobretudo o contexto humano ou natural deles decorrente, que difere do presente e, por isso, importa entender. Stransky (1974, p. 32) condensa esta noção ao atribuir aos objetos museológicos a característica de *musealium*, isto é, a de serem objetos separados da sua realidade e transferidos para um museu onde encontram nova realidade, de forma a documentar a realidade da qual são originários. Schreiner desenvolve esta noção de objeto musealizado, atribuindo-lhe as seguintes características: são testemunho de algo; são descontextualizados com fins de conservação; são selecionados ou adquiridos para uma determinada coleção e são preservados e decodificados; são utilizados com finalidades sociais, intelectuais, culturais, nomeadamente através da sua exposição e outras formas de comunicação (Schreiner, 1984, pp. 24-28). A figura 1, baseada no modelo de Shannon, pretende representar este processo.

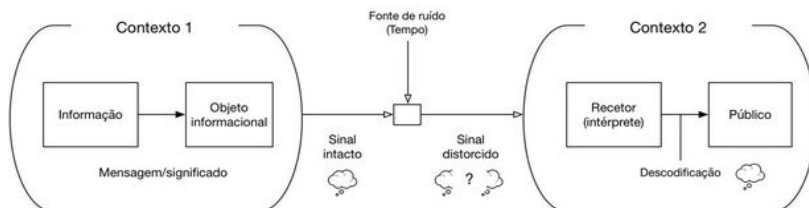


Figura 1. Modelo de codificação/descodificação da informação em museus

Uma determinada informação é materializada e registada num objeto informacional que pode ser criado por um autor (uma joia, faca, moeda) ou gerado pela Natureza (rocha, fósfil, cristal). Em ambos os casos, o artefacto possui um contexto cultural ou natural original, indispensável para a compreensão da mensagem e do significado que ele veicula. No entanto, a passagem do tempo distorce esse significado inicial, uma vez que induz a alteração do contexto. Portanto, quando o intérprete, no presente — que é um futuro em relação ao momento de criação — interage com o artefacto, ele habita num contexto diferente, dentro do qual aquele artefacto lhe é estranho. Ainda assim, ele encara o sinal distorcido do artefacto (quer fisicamente, por estar partido, incompleto ou descolorido, ou simplesmente por ser o produto de uma época distinta) e gera uma interpretação do seu significado, de modo a recriar uma hipótese do sentido inicial. Diz-se interpretação, e não conclusão, porque este nunca pode compreender na sua totalidade o contexto original, dado que nunca habitou nele. Por isso, como toda a nossa percepção do passado é uma interpretação com base nos dados disponíveis, que podem ser mais ou menos fidedignos, o intérprete nunca pode afirmar com absoluta certeza que compreende o contexto original e que, portanto, o sentido do objeto é absoluto. Formula, pois, uma hipótese interpretativa. Finalmente, fornece esse sentido reconstruído e ‘intacto’ ao público, a quem é dada a possibilidade de consumir essa informação e de a transformar em conhecimento.

O papel das ciências da informação e comunicação é precisamente o de “potenciar as transformações sociais através da mediação entre os objetos potencialmente informativos e as pessoas potencialmente usuárias da informação” (Marques, 2015, p. 57). O mais usual é o mediador ser apresentado como sendo um agente humano, como um arquivista, um bibliotecário, um museólogo ou um designer de conteúdos (Silva, 2009, p. 90). No entanto, no âmbito de um museu, faz mais sentido entender o sistema tecnológico de informação como o agente de mediação entre os objetos informacionais e os visitantes.

Embora todas as estruturas tecnológicas sejam obra humana e, por conseguinte, uma pessoa esteja sempre na conceção do processo de mediação, no âmbito de um espaço museológico (como uma exposição) o canal de comunicação é ocupado por uma tecnologia, que pode ir de uma simples etiqueta a uma apresentação audiovisual sobre o objeto. Em ambos os casos, é o STI o responsável por operar a comunicação entre o objeto e o visitante, através do qual este assimila o sentido reconstruído do artefacto e enriquece potencialmente o seu conhecimento. Para tal, Thaller identifica cinco requisitos dos STI cuja finalidade é a representação de fontes históricas: (1) *represents the artifacts as free from any interpretation as possible in the technical system*, (2) *embeds them, however, in a network of interpretations of what they imply*, (3) *provides tools which help to remove contradictions between such interpretations*, (4) *accepts, however, that such contradictions may prove to resist resolution*, (5) *as well as that all interpretations always represent tendencies, no certainties* (Thaller, 2018). Conforme visto anteriormente, salienta-se o carácter especulativo das

interpretações e a necessidade de contextualizar as informações veiculadas pelos objetos.

5 *User experience*

5.1 Definição e enquadramento

O termo *user experience* (UX), ou experiência dos utilizadores, surgiu pela primeira vez em 1995 como forma de descrever os avanços no desenvolvimento de interfaces e aplicações para utilizadores (Norman, Miller, & Henderson, 1995). A investigação e a realização de testes em UX permite analisar as relações complexas que se estabelecem entre os utilizadores, as características do sistema e o contexto de interação, de forma a desenvolver produtos e serviços que não só evitem um uso difícil, mas também que constituam uma experiência positiva (Felming-May et al., 2018, p. 1).

A UX surge assim como uma forma de ir ao encontro das expetativas dos utilizadores, e até mesmo de as superar. Todavia, é importante realçar que as expetativas são algo em permanente construção, na medida em que “refletem o desempenho anterior feito pelo cliente sobre os níveis de desempenho” (Marques, 2012, p. 349). Logo, após uma primeira visita de sucesso, a instituição tem de continuar a mostrar altos níveis de desempenho com base nas necessidades do seu público (Miller, 2000).

A perceção e a avaliação da UX têm-se afirmado como uma área de investigação crucial na CI, dando aso ao desenvolvimento de vários estudos sobre a interação entre pessoas e a tecnologia (Pallud & Monod, 2010, p. 563). No entanto, os profissionais da informação ainda possuem escassas normas para avaliar a experiência dos utilizadores (Institute of Museum and Library Services, 2006; Tost & Economou, 2007), notando-se também a falta de preparação académica sobre esta matéria (Applegate, 2016; Nitecki, Wiggins, & Turner, 2015). Ainda assim, as investigações existentes têm apontado para o carácter variado e multidimensional da UX, que exige uma abordagem multidimensional (Isomaki, 2009; McCarthy & Wright, 2004) ou, por outras palavras, sistémica. Pauta-se por uma matriz essencialmente pessoal e subjetiva, enquanto *a consequence of the user's prior experiences, attitudes, skills, habits and personality* (ISO 9241-210, p. 2000). De facto, é com base no conhecimento prévio que os utilizadores recebem a informação e a assimilam de acordo com o que, no seu entendimento, é pertinente, num processo de permanente construção e desconstrução do conhecimento (Davenport & Donald Marchland, 2004, p. 190; Marques, 2015, p. 51; Rowley, 2007, p. 173). Por este motivo, a investigação nesta área deve superar a perspetiva

puramente focada nas tecnologias que normalmente vigora no desenvolvimento dos sistemas de informação (Hassenzahl & Tractinsky, 2006, p. 92), focando-se antes numa perspetiva mais humana de compreensão das emoções e dos afetos resultantes dessa experiência de contacto.

5.2 A experiência dos visitantes nos museus

O termo *user experience* (UX), ou experiência dos utilizadores, surgiu pela primeira vez em 1995 como forma de descrever os avanços no desenvolvimento de interfaces e aplicações para utilizadores (Norman, Miller, & Henderson, 1995). A investigação e a realização de testes em UX permite analisar as relações complexas que se estabelecem entre os utilizadores, as características do sistema e o contexto de interação, de forma a desenvolver produtos e serviços que não só evitem um uso difícil, mas também que constituam uma experiência positiva (Felming-May et al., 2018). A investigação nesta área deve superar a perspetiva puramente focada nas tecnologias que normalmente vigora no desenvolvimento dos sistemas de informação (Hassenzahl & Tractinsky, 2006), focando-se antes numa perspetiva mais humana de compreensão das emoções e dos afetos resultantes dessa experiência de contacto.

Os museus são locais onde as pessoas procuram experiências e sentidos (Carr, 2001). Há um desejo inerente de interagir com o património de forma a conhecer mais sobre a história (Poria, Reichel, & Biran, 2006) e sobre nós próprios (Kimmelman, 2001). Como tal, “enhancing UX by providing cultural entertainment and education is the crucial target for museums” (ICOM, 2007). Neste sentido, os museus — e, por extensão, as tecnologias por eles usadas com este fim — devem ser capazes de garantir uma experiência apropriada às necessidades dos seus visitantes, sejam elas de natureza funcional (educativas) ou hedónicas (de entretenimento) (Pallud & Monod, 2010).

Assim, as exposições figuram como um espaço privilegiado de informação e comunicação em museus. No entanto, a informação é um conceito contextual (Mahler, 1996) e, como tal, para que seja assimilada pelos visitantes e enriqueça os seus conhecimentos, é necessário contextualizá-la através da contextualização dos artefactos. A mediação é, portanto, entendida como o processo que compreende a transmissão da informação/significado de um objeto informacional — significado esse assente na compreensão do contexto que presidiu à sua criação — e a assimilação dessa informação/significado por um visitante pertencente a um contexto diferente — e, por conseguinte, para o qual o artefacto, por si só, é um objeto estranho e incompreensível — de modo a enriquecer a sua experiência e aprendizagem.

A mediação depende em grande medida do STI, na medida em que este normalmente fornece o contexto e o sentido dos artefactos, quer seja através de etiquetas, seja através de uma componente interativa de painéis informativos adicionais. No entanto, a mediação, tal como aqui formulada, vai mais além. Umberto Eco (2005), numa conferência onde aborda o papel dos museus no terceiro milénio,

lançou uma proposta audaz: de forma a enriquecer o conhecimento do contacto entre a cultura eurocêntrica e as restantes culturas, Eco sugeriu que um museu fosse disposto na forma de um labirinto, em que cada caminho fosse consignado a uma história própria. Neste museu, por exemplo, a história da Europa seria paralela à história da América até 1492, momento a partir do qual ambos os caminhos se fundiriam num único só, e se prolongariam em conjunto. Neste exemplo, o visitante teria a possibilidade de construir a sua própria história ao mesmo tempo que a compara com outras histórias paralelas, naquilo que seria uma experiência enriquecida e exacerbada por uma componente visual e espacial que tende a despertar a sua curiosidade, interesse e aprendizagem.

Este ato interpretativo, que constitui a essência da investigação em história (Thaller, 2018), é a pedra basilar da mediação no espaço museológico. O sucesso deste ato comunicativo, dominado pela participação do STI mas a ele não reduzido, é a chave para promover uma experiência rica ao visitante, quer de natureza lúdica quer educacional, procurando para isso adaptar-se às suas necessidades, em vez de disponibilizar os seus acervos com uma autoridade unilateral e dirigista.

6 Conclusões

A valorização da experiência pessoal dos indivíduos, durante a visita, deve ser um critério determinante para fundamentar a atividade cultural dos museus e para potenciar a aprendizagem. A experiência tem de ser encarada de forma sistémica, não se resumindo unicamente à conceção e *design* de aplicações informáticas, mas compreendendo todo o processo de contacto que um visitante tem com a informação veiculada pelo objeto que, além de incluir naturalmente uma componente de mediação pelo sistema tecnológico de informação, pode também passar por uma experiência estética do espaço e da representação da obra. Justifica-se, portanto, entender os conceitos de “user experience” e de mediação num sentido mais expansivo enquanto o processo infocomunicacional entre um artefacto, enquanto um objeto informacional, e um visitante que atua como recetor dessa informação, que pode transformar em conhecimento.

Agradecimentos

Este estudo foi financiado na sua totalidade pela Fundação para a Ciência e Tecnologia através da bolsa de doutoramento SFRH/BD/131004/2017.

Referências

- Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24(4), 665–694.
- Blythe, M., Overbeeke, C., Monk, A., & Wright, P. (2003). *Funology: from usability to enjoyment*. Dordrecht: Kluwer Academic Publishers.
- Borko, H. (1968). Information Science: what is it? *American Documentation*, 19(1), 3–5. Disponível em <http://cdigital.uv.mx/bitstream/123456789/6699/2/Borko.pdf>
- Briet, S. (1951). *Qu'est-ce que la documentation?* Paris: ÉDIT.
- Brookes, B. (1980). The foundations of information science. Part I. Philosophical aspects. *Journal of Information Science*, 2, 125–133. Disponível em <http://comminfo.rutgers.edu/~kantor/601/Readings2004/Week3/r4.PDF>
- Bryman, A. (2016). *Social research methods*. Oxford University Press.
- Capurro, R., & Hjørland, B. (2003). The Concept of Information. *Annual Review of Information Science and Technology*, 37(chapter 8), 343–411. <https://doi.org/DOI: 10.1002/aris.1440370109>
- Carr, D. (2001). A museum is an open work. *International Journal of Heritage Studies*, 7(2), 173–183.
- Castells, M. (2002). *A era da informação: economia, sociedade e cultura*. Lisboa: Fundação Calouste Gulbenkian.
- Eco, U. (2005). El museo en el tercer milénio. *Revista de Occidente*, 290–291(jul./ago.), 33–53.
- Felming-May, R., Mays, R., Forrester, A., Walker, T., Tenopir, C., Bilal, D., & Allard, S. (2018). Experience assessment: designing an innovative curriculum for assessment and UX professionals. *Performance Measurement and Metrics*, 19(1), 30–39. <https://doi.org/https://doi.org/10.1108/PMM-09-2017-0036>
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. São Paulo: Atlas S.A.
- Hassenzahl, H., & Tractinsky, N. (2006). User experience - a research agenda. *Behavior & Information Technology*, 25(2), 91–97.
- Hassenzahl, M., & Roto, V. (2007). Being and doing: a perspective on user experience and its measurement. *Interfaces*, 72, 10–12.
- Holbrook, M. (2000). The millennial consumer in the texts of our times: experience and entertainment. *Journal of Macromarketing*, 20, 178–192.
- ICOM. (2007). ICOM statutes.
- Kimmelman, M. (2001). Museums in a quandary: where are the ideals? *New York Times*.

- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60(3), 269–298.
<https://doi.org/https://doi.org/10.1016/j.ijhcs.2003.09.002>
- Le Coadic, Y.-F. (1996). *A ciência da informação*. Brasília: Briquet de Lemos/Livros.
 Disponível em <http://www.restaurabr.org/siterestaurabr/CICRAD2011/M1Aulas/M1A3Aula/20619171-le-coadic-francois-a-ciencia-da-informacao.pdf>
- Mahler, G. (1996). Quantum information. In K. Kornwachs & K. Jacoby (Eds.), *Information: new questions to a multidisciplinary concept* (pp. 103–118). Berlin: Akademie Verlag.
- Maroevic, I. (2006). The museum object as historical source and document. In *Museologia e historia: un campo del conocimiento*. ICOFOM Study Series.
- Marques, M. B. (2017). Gestão da informação em sistemas de informação complexos, 60–76.
- Mensch, P. van. (1990). Methodological museology; or towards a theory of museum practice. In S. Pearce (Ed.), *Objects of knowledge* (pp. 141–157). London: The Athlone Press.
- Norman, D., Miller, J., & Henderson, A. (1995). What you see, some of what's in the future, and how we go about doing it. *Conference Companion on Human Factors in Computing Systems - CHI '95*, (January), 155.
<https://doi.org/10.1145/223355.223477>
- Otlet, P. (1934). *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelas: Editiones Mundaneum. Disponível em
<https://archive.org/details/OtletTraitDocumentationUgent>
- Pallud, J., & Monod, E. (2010). User experience of museum technologies: The phenomenological scales. *European Journal of Information Systems*, 19(5), 562–580.
- Pallud, J., & Straub, D. W. (2014). Effective website design for experience-influenced environments: The case of high culture museums. *Information and Management*, 51(3), 359–373. <http://dx.doi.org/10.1016/j.im.2014.02.010>
- Poria, Y., Reichel, A., & Biran, A. (2006). Heritage Site Perceptions and Motivations to Visit. *Journal of Travel Research*, 44(3), 318–326.
<https://doi.org/https://doi.org/10.1177/0047287505279004>
- Schreiner, K. (1984). Basic paper. In *Collecting today for tomorrow* (pp. 24–28). ICOFOM Study Series 6.
- Silva, A. M. da. (2006). *A Informação: da compreensão do fenómeno e construção do objeto científico*. Porto: Edições Afrontamento.

- Silva, A. M. da, & Ribeiro, F. (2008). *Das «ciências» documentais à ciência da informação: ensaio epistemológico para um novo modelo curricular* (2ª). Porto: Edições Afrontamento.
- Silva, A. M. da. (2009). Arquivologia e gestão da informação/conhecimento. *Informação & Sociedade: Estudos*, 19(2), 47–52. Disponível em <https://repositorio-aberto.up.pt/bitstream/10216/26548/2/MalheiroArquivologiaGestaoInf000108143.pdf>
- Sransky, Z. (1974). Metodologické otázky dokumentace soucasnoti. *Muzeologicke Sesity*, 5, 13–43.
- Thaller, M. (2018). A digital ivory tower: prolegomena for a computer science for historical studies. Retrieved June 23, 2018, from <https://ivorytower.hypotheses.org/56#more-56>.

Estratégias de reconstrução do patrimônio cultural pós-desastre, com o uso de tecnologias digitais: estudo de caso do Museu Nacional – Brasil

Fernanda Miranda de Vasconcellos Motta¹ y Ronaldo André Rodrigues da Silva²

¹ ORCID 0000-0001-8666-8906. Universidade Federal de Minas Gerais, Brasil.
fernandavasc@gmail.com

² ORCID 0000-0002-0656-8671. Pontifícia Universidade Católica de Minas Gerais, Brasil.
ronaldoandre@gmail.com

Resumo. Em situações pós-desastre, nas quais há prejuízos a construções históricas e a acervos museais, o resgate do valor patrimonial se dá por meio de processos de salvaguarda e comunicação. Com base nessa consideração, o objetivo deste estudo é pesquisar a respeito das estratégias de reconstrução do patrimônio do Museu Nacional com o uso de tecnologias digitais. Parte-se da constatação de que mais de 90% do acervo, com cerca de 20 milhões de objetos museais, foi afetado por um incêndio, ocorrido em 2018. Para a condução da pesquisa, utiliza-se metodologia qualitativa, baseada em levantamentos bibliográficos e documentais, organizados em um estudo de caso. São abordadas estratégias de digitalização e restauração digital, com reflexões sobre integridade e autenticidade do acervo. São analisadas, também, estratégias digitais colaborativas e discutidas questões referentes à autoridade e curadoria informacional. Por fim, são consideradas as estratégias de reconstrução virtual da experiência de museu. É possível inferir que o processo de reconstrução do patrimônio do Museu Nacional contempla a dimensão temporal de mudança e variabilidade, abarcando estratégias narrativas e contextuais, com o suporte das tecnologias digitais. Propõe-se, assim, avançar na compreensão dos processos pelos quais novos patrimônios e conhecimentos são incorporados ao museu.

Palavras chave: Patrimônio cultural. Reconstrução pós-desastre. Tecnologias digitais. Museu Nacional. Brasil.

Abstract. In post-disaster situations, in which there are losses to historical buildings and museum collections, the recovery of patrimonial value takes place through processes of safeguard and communication. Based on this consideration, the objective of the study is to investigate the strategies for the reconstruction of National Museum heritage that involve the use of digital technologies. It is based on the finding that more than 90% of the collection, with around 20 million museum objects, was affected by a fire in 2018. For the

conduction of the research, a qualitative methodology is used, based on bibliographic and documental studies and organized in a case study. Digitalisation and digital restoration strategies are addressed, with reflections on the integrity and authenticity of the collection. Collaborative digital strategies are analyzed too, and issues related to informational authority and curation are discussed. Finally, strategies of virtual reconstitution of the museum experience are considered. It is inferred that the process of reconstruction of the National Museum heritage contemplate the temporal dimension of change and variability, through narrative and contextual strategies, supported by digital technologies. It is proposed, thus, to understand the processes through which new heritage and knowledge are incorporated into the museum.

Keywords: Cultural heritage. Post-disaster reconstruction. Digital technologies. National Museum. Brazil.

1 Introdução

“É importante ressaltar que o Museu Nacional, apesar de ter perdido uma parte significativa do acervo, jamais perdeu a capacidade de gerar conhecimento” (Alexander Kellner, Diretor do Museu Nacional).

Estratégias de salvaguarda e comunicação, em cenários pós-desastre, constituem a base para o resgate do valor de construções históricas, acervos e coleções perdidos. A dinâmica de patrimonialização é atualizada, a partir do interesse social em relação ao bem material ou imaterial afetado e pelas possibilidades de conhecimento a respeito dele, nesse novo contexto (Davallon, 2015). Tal processo de significação social envolve a integração entre saberes, experiências e valores diversos, sendo coletivamente construído e transmitido para futuras gerações.

O museu e as políticas patrimoniais tratam artefatos, edifícios e costumes de modo a tornar inteligíveis as relações entre eles, propondo interpretações sobre o que significam para as pessoas que os evocam. Tem-se, então, espaço para construção de conhecimento. “O objetivo é a apreciação pública do patrimônio, o entendimento de sua dimensão cultural e educacional, em um mesmo universo, interconectado por um conjunto de redes multinodulares” (Portalés, Rodrigues, Rodrigues Gonçalves, Alba & Sebastián, 2018, p. 2).

Quando desastres ocorrem, o museu se torna espaço de esquecimento. Como alternativa para contornar tal situação, têm-se as tecnologias digitais, que instauram possibilidades diversas de restauração e comunicação do patrimônio, a partir de uma perspectiva sócio-cultural (Serain, 2018).

Com base nessas considerações iniciais, propõe-se pesquisar as estratégias de reconstrução do patrimônio adotadas pelo Museu Nacional, situado no Rio de Janeiro,

Brasil – com o uso de tecnologias digitais. O referido museu teve mais de 90% de seu acervo, de cerca de 20 milhões de objetos museais, afetado por um incêndio, em 02 de setembro de 2018. Após o incêndio, foram iniciadas as atividades de resgate e restauração dos artefatos, lideradas por profissionais do Museu Nacional, bem como obras de recuperação da construção histórica, sob a coordenação do Instituto do Patrimônio Histórico e Artístico Nacional [IPHAN]. Até maio de 2019, estimava-se que haviam sido recuperadas mais de 2.700 peças, entre os escombros (Sampaio, 2019). As principais fontes de financiamento do processo de reconstrução do museu são recursos do Governo Federal, do fundo emergencial da Organização das Nações Unidas para a Educação, a Ciência e a Cultura [UNESCO], além de doações internacionais e recursos arrecadados por meio de financiamento coletivo.

2 Metodologia

De forma coerente com o objetivo deste estudo, adotou-se uma abordagem metodológica qualitativa, baseada em levantamentos bibliográficos e documentais, para fundamentar o desenvolvimento de um estudo de caso sobre o processo de recuperação do patrimônio, pós-desastre, no Museu Nacional. Segundo Freitas & Jabbour (2011), “[...] o interesse do pesquisador não está focalizado em quantificar uma ocorrência ou quantas vezes uma variável aparece, mas sim na qualidade em que elas se apresentam, ou seja, como as coisas acontecem” (p.9). A estratégia de estudo de caso enfatiza, portanto, entendimentos contextuais, centrando-se na compreensão dinâmica do objeto de estudo.

3 O uso de tecnologias digitais na preservação do patrimônio

No mundo contemporâneo, tem-se uma situação híbrida, em que as esferas físicas de interação com o patrimônio cultural se fundem às esferas virtuais (Nikonova & Biryukova, 2017a). É cada vez mais usual que as pessoas aprendam sobre o passado através de aplicações tecnológicas, reconstruções virtuais e representações digitais de artefatos (Economou, 2015a). Isso se manifesta nas comunidades e grupos de discussão *online*, no compartilhamento exponencial de arquivos digitais, e no reuso destes, no contexto da colaboração aberta.

O processo de digitalização de objetos museais transforma a relação com os públicos - de estática a dinâmica – sendo que estes se tornam usuários informacionais. Se, por um lado, perde-se o “aqui e agora” do objeto museal,

alterando sua condição de singular testemunho material do mundo, por outro lado, abrem-se perspectivas de democratização de acesso e de interatividade que favorecem a fruição e o aprendizado. Nesse sentido, é útil examinar os desenvolvimentos na área do patrimônio digital, observando que abordagens têm sido adotadas, quais são os pressupostos subjacentes e as lacunas existentes (Economou, 2015b).

As noções de autenticidade e integridade do acervo entram em atualização, com o reconhecimento da dimensão imaterial deste. Henriques & Dodebei (2011a) afirmam que “talvez tenha sido necessário criar o conceito de bem imaterial para que fosse possível pensar em preservação para além da materialidade” (p.10). As referidas autoras reforçam a ideia de que é possível preservar significados, independentemente dos objetos materiais que são sua referência. O registro digital transforma o objeto museal, material ou imaterial, em informação. É preciso pensar, portanto, em como assegurar a autenticidade de uma informação digital preservada, bem como a integridade do conhecimento e das experiências criadas a partir de ambientes virtuais (Nikonova & Biryukova, 2017b). As referidas autoras ressaltam, ainda, a questão da obsolescência tecnológica, além de desafios quanto ao armazenamento virtual dos dados digitais, que se expandem continuamente. Por fim, há o desafio relacionado à transformação da massa de dados digitais em conhecimento útil para os públicos. Economou (2015c) ressalta que “o uso de ferramentas digitais pode ajudar os especialistas a organizarem essa vasta quantidade de dados, tornando-os úteis para públicos segmentados” (p.217).

Diante da diversidade e complexidade das estratégias e técnicas digitais disponíveis, não se trata apenas de tomar decisões tecnológicas, e sim de considerar o cenário mais abrangente que se apresenta para o museu, a partir de análises sociais, culturais, ambientais e econômicas.

4 Histórico do Museu Nacional

O Museu Nacional, situado no Rio de Janeiro, Brasil, foi criado por D. João VI, em 1818, com a denominação de Museu Real e com o objetivo de disseminar conhecimento sobre as ciências naturais.

Com a instauração da República, em 1892, ele passou a se chamar Museu Nacional e sua sede foi transferida para o Palácio de São Cristóvão, na Quinta da Boa Vista. O Palácio foi residência da Família Real portuguesa, entre 1808 e 1822, e da Família Imperial brasileira, até 1889. Em 1938, a construção histórica foi tombada pelo IPHAN. Em 1946, o museu foi incorporado à Universidade Federal do Rio de Janeiro [UFRJ] e se consolidou como referência em diferentes campos do saber científico, contemplando atividades de difusão cultural, educação e pesquisa.

É oportuno mencionar que o Museu Nacional é a mais antiga instituição científica do Brasil e, antes do incêndio, era o maior museu de história natural e antropológica da América Latina e o quinto maior museu do mundo, em termos de acervo. O acervo do Museu Nacional abarcava mais de 20 milhões de objetos museais, organizados em

coleções de botânica, entomologia, arqueologia, etnologia, geologia, paleontologia, zoologia e antropologia.

Nos anos 2000, o paleontólogo Sérgio de Azevedo implantou, no museu, o Laboratório de Processamento de Imagem Digital [LAPID], trabalhando em parceria com o Instituto Nacional de Tecnologia [INT], ligado ao Ministério de Ciências e Tecnologia - para tomografar e digitalizar artefatos museais.

A perda de mais de 90% do acervo, com o incêndio ocorrido em 02 de setembro de 2018 - apenas três meses após a comemoração do bicentenário do museu – levou à instauração de uma complexa operação de resgate e restauro de artefatos, aliada à busca pela ampliação e organização dos registros digitais. A partir desse acervo resgatado, vêm sendo delineadas estratégias de recuperação pós-desastre que abarcam exposições temporárias, campanhas colaborativas e o uso de recursos tecnológicos.

5 Estratégias digitais de recuperação do patrimônio do museu

Tem-se o propósito de compreender, a partir do estudo do caso sobre a recuperação pós-desastre do patrimônio do Museu Nacional, como a tecnologia favorece esse processo. “A preocupação passa a ser não apenas em salvaguardar os vestígios do passado, como também incluir nesse processo ações desencadeadas no tempo presente” (Henriques & Dodebei, 2011b, p. 10).

As possibilidades de recuperação e revisitação do patrimônio do Museu Nacional representam a continuidade da história e das funções deste. Isso envolve tanto a exibição de acervos e coleções resgatados, quanto a retomada das atividades de ensino e pesquisa, relacionadas à produção e disseminação de conhecimento.

5.1 Digitalização e restauração digital do acervo

Uma das bases para o processo de reconstrução do patrimônio foi a parceria existente entre o Museu Nacional e o INT, que viabilizou o processo de digitalização de parte do extenso acervo. Digitalizou-se, aproximadamente, 300 objetos museais. A seleção foi feita a partir da curadoria dos pesquisadores do museu, considerando-se sua importância científica, especificidade, estado de conservação e interesse social. Esses objetos museais foram capturados por *scanners* portáteis tridimensionais a *laser*, o que permite seu exame e manipulação em estudos, simulações e reconstituições, sem que ocorram danos materiais. O processo também possibilita a construção de réplicas, por meio do

escaneamento, da modelagem computacional e da impressão tridimensional, seja para estudo ou para reposição de acervo perdido.

Uma inovação no processo de restauração pós-desastre consiste na fabricação de réplicas tridimensionais dos artefatos danificados, com a incorporação de vestígios materiais coletados durante a operação de resgate. Pesquisadores do LAPID adicionam fragmentos dos artefatos resgatados às resinas específicas para impressão. As réplicas desenvolvidas passam por testes de qualidade e carregam grande carga simbólica, por incorporarem vestígios da materialidade dos objetos museais danificados ou destruídos, e por representarem registros das transformações pelas quais o patrimônio passa, ao longo do tempo.

Warthon (2005) esclarece que, no contexto contemporâneo da conservação, a autenticidade de um objeto museal não corresponde apenas ao seu estado original, mas também ao registro de cada etapa da trajetória deste objeto, o que abarca suas formas de representação, as réplicas criadas e as relações instauradas com a comunidade. Quanto à integridade, trata-se da capacidade de um artefato cultural preservar, e mesmo ampliar, seus significados, ao longo do tempo. Como desdobramento, estratégias de cooperação com a comunidade na reconstrução do patrimônio ganham ênfase, uma vez que há aspectos sensíveis, relacionados aos significados atribuídos aos bens culturais, que escapam ao conhecimento eminentemente técnico dos especialistas do museu.

5.2 Tecnologias digitais colaborativas

O uso de tecnologias, por meio de processos sociais abertos e colaborativos, para ampliar e recompor o patrimônio, representa uma estratégia contemporânea de reconstrução pós-desastre. A ampliação do acervo digital envolve a coleta, organização e indexação colaborativa de arquivos, por meio de um trabalho que envolve tanto especialistas como amadores. O propósito é a construção de uma rede íntegra de significação relacionada ao Museu Nacional e ao seu acervo.

Nesse contexto, por iniciativa de lideranças da comunidade *Wiki* foi feita uma solicitação aos usuários para que postassem arquivos e informações a respeito do museu, logo após o incêndio. Na primeira semana, cerca de 250 pessoas contribuíram com a entrada de imagens, vídeos e textos, sendo que as informações foram traduzidas em 21 idiomas. Até julho de 2019, já haviam sido incorporadas 3.928 imagens à *Wikimedia* do Museu Nacional.

Com base nesse crescimento do volume de dados, criou-se o *Wikiprojeto* do Museu Nacional. Foi desenvolvido um conjunto de páginas destinadas à gestão do tema, por parte de um grupo de editores com atuação em diferentes frentes. O projeto conta com 22 editores que são integrantes experientes da comunidade, ligados às áreas de educação, ciência, cultura e tecnologia.

Mansell (2013a) discute a respeito de tais iniciativas relacionadas às tecnologias digitais, que incorporam valores socialmente compartilhados a *hardware*, *software* e

aplicações. Segundo o referido autor, o principal desafio é transformar a massa de informações em conhecimento útil. Além da acurácia informacional, é preciso assegurar sua correta indexação no repositório digital. Muitas vezes, as informações geradas por comunidades digitais abertas se encontram fragmentadas e não podem ser utilizadas, justamente por não seguirem parâmetros formais requeridos quanto aos metadados. Mansell (2013b), nesse sentido, reforça a importância de se ter “observadores qualificados”, em aplicações colaborativas, para identificar e corrigir erros, checar conceitos, teorias e processos de indexação, favorecendo o reuso informacional e a geração de conhecimento.

5.3 Reconstituições virtuais da experiência de museu

Para além de um antagonismo, o espaço virtual complementa o físico, configurando uma solução efetiva em termos de acesso às coleções em situação de destruição.

A partir de processos de digitalização do acervo e de captura de imagens do Museu Nacional, através da ferramenta *Google Street View*, o Instituto Cultural *Google* desenvolveu, em dezembro de 2018, o projeto “Por dentro do Museu Nacional”. Ele foi hospedado na plataforma digital *Google Arts & Culture*, e exibe o acervo virtual de 164 objetos museais, organizados em oito exposições temáticas, além de oferecer o *tour* virtual 360 graus - “Descubra o Museu Nacional”, - que permite ao usuário percorrer os espaços expositivos existentes antes do incêndio. É possível fazer a visita virtual com audioguia, e há opções dos idiomas português, inglês e espanhol. Pode-se, também, utilizar um visor de realidade virtual para explorar, de modo imersivo, o museu e seu acervo.

De forma complementar a esse *tour* virtual, outra experiência imersiva desenvolvida pelo Museu Nacional, em parceria com a empresa de tecnologia LG, é a disponibilização de um *tour* sonoro, a partir de áudios tridimensionais captados no ambiente museal, antes do incêndio. Por meio deles, é possível que o usuário faça a visita virtual, acompanhada de audiodescrição, e perceba a dimensão, textura e localização de peças do acervo, vivenciando uma experiência sonora muito próxima da real. A *playlist* com os áudios está disponível no *Youtube*, e nas plataformas *Spotify* e *Deezer*, com narrações em português e inglês.

6 Conclusão

É possível inferir que, mesmo que arquivos digitais não substituam o patrimônio físico perdido, eles contribuem para que se forme uma memória digital capaz de assegurar que este não seja relegado ao esquecimento.

O patrimônio digital, além de representar um registro do passado, torna-se testemunha das transformações que o tempo produz. Ele transita para outro estado, resultante de suas adaptações, ao longo do tempo, em interação com contextos sociais, culturais e tecnológicos diversos (Holling, 2019).

Assim, a partir de objetos, informações e ambientes museais recuperados pós-desastre, surgem novos patrimônios e renovadas possibilidades de conhecimento. Torna-se de grande valor compreender o museu e suas transformações pós-desastre, contanto com a tecnologia digital para recuperar seu potencial informacional e para assegurar a continuidade das atividades de difusão cultural, ensino e pesquisa.

Nota final. A autora Fernanda Miranda de Vasconcellos Motta agradece à Fundação de Amparo à Pesquisa do Estado de Minas Gerais [FAPEMIG], pelo incentivo à realização deste estudo, por meio da concessão de bolsa de estudos de pós-graduação.

Referências

- Carvalho, E. (2019). Pesquisadores usam cinzas do Museu Nacional para reconstruir acervo destruído em incêndio. *BBC Brasil*. Rio de Janeiro. Disponível em: <http://www.bbc.com/portuguese/brasil-47781771>
- Davallon, J. (2015). Memória e patrimônio: por uma abordagem dos regimes de patrimonialização. In TARDY, C.; DOBEDEI, V. (Org.). *Memória e novos patrimônios*. Marseille: Open Edition Press. Disponível em: <http://books.openedition.org/oep/866>.
- Economou, M. (2016). Heritage in the digital age. In LOGAN, W.; CRAITH, M. N.; KOCKEL, U. (Orgs.). *A companion to heritage studies* (Chapter 15). London: John Wiley & Sons.
- Fernandes, A.C. (2019). Dentro das ruínas do Museu Nacional. *Revista FAPESP*. Edição 277. Disponível em: <http://revistapesquisa.fapesp.br/2019/03/14/dentro-das-ruinas-do-museu-nacional/>
- Freitas, W. R. S.; Jabbour, C. J.C. (2011). Utilizando estudo de caso(s) como estratégia de pesquisa qualitativa: boas práticas e sugestões. *Estudo & Debate*. Lajeado, 18 (2), 07-22.
- Google Arts & Culture (website). Por dentro do Museu Nacional. Disponível em: <http://artsandculture.google.com/project/museu-nacional-brasil>
- Henriques, R.; Dodebei, V. (2011). Os museus e os novos patrimônios. *Anais do XXVI Simpósio Nacional de História*. São Paulo: ANPUH.

- Holling, H. B. (2019). Mídia transicional: permanência, recursividade e o paradigma da conservação. In Gobira, Pablo (Org.). *A memória do digital e outras questões das artes e museologia*. Belo Horizonte: EdUEMG.
- Mansell, R. (2013). Employing digital crowdsourced information resources: managing the emerging information commons. *International Journal of the Commons*. 7 (2). August, p. 255-277.
- Museu Nacional (website). <http://www.museunacional.ufrj.br/>
- Nikonova, A. A.; Biryukova, M. V. (2017). The role of digital technologies in the preservation of cultural heritage. *Muzeológia a kultúrne dedičstvo*, 5(1), p. 169-173. Disponível em: <http://www.researchgate.net/publication/317757322>.
- Pivetta, M.. (2018). Pré-história em pedaços. *Revista FAPESP*. Edição 272. Disponível em: <http://revistapesquisa.fapesp.br/2018/10/22/pre-historia-em-pedacos/>
- Portalés, C.; Rodrigues, J. M.F.; Rodrigues Gonçalves, A.; Alba, E.; Sebastián, J. (2018). Multimodal technologies and interact. *Digital cultural heritage* (editorial). 2 (3), n. 58. Disponível em: http://www.mdpi.com/journal/mti/special_issues/Digital_Cultural_Heritage
- Sampaio, L.F.P. (2019). Resgatadas algumas peças da coleção egípcia do Museu Nacional. *Antigo Egito (online)*, publicado em 08 de maio. Disponível em: <http://antigoegito.org/anuncio-de-resgate-de-algumas-pecas-da-colecao-egipcia-do-museu-nacional/>
- Serain, C. (2018). The sensitive perception of cultural heritage's materiality through digital technologies. *Studies in Digital Heritage*. 2 (1). Disponível em: <http://scholarworks.iu.edu/journals/index.php/sdh/article/view/24606>
- Warton, G. Indigenous claims and heritage conservation: an opportunity for critical dialogue. *Public Archaeology*, n. 4, v. 2-3, p. 199-204, 2005.
- Wikipedia. *Projeto do Museu Nacional*. Disponível em: http://pt.wikipedia.org/wiki/Wikipedia/Projetos/Museu_Nacional

A utilização e tratamento de documentos fotográficos em projetos de investigação: o caso do Generation.Mobi

Joana Rodrigues¹ y Carla Teixeira Lopes²

¹ ORCID [0000-0002-1309-2122](https://orcid.org/0000-0002-1309-2122). Faculdade de Engenharia da Universidade do Porto, INESC TEC, Portugal

Joanasousarodrigues.14@gmail.com

² ORCID [0000-0002-4202-791X](https://orcid.org/0000-0002-4202-791X). Faculdade de Engenharia da Universidade do Porto, INESC TEC, Portugal
ctl@fe.up.pt

Resumo. As novas tecnologias e a imposição do digital têm assumido um papel de destaque na sociedade de informação, desencadeando novas tendências e tornando recursos altamente valorizáveis. Por sua vez, a imagem é dominante na comunicação, mas a fotografia tem ainda um tratamento diverso, no que toca à descrição, interpretação e uso sistemático. A dificuldade de utilização e tratamento de coleções de imagens é conhecida, mas tem vindo a diminuir com as potencialidades da análise automática, dos novos dispositivos de captura de imagem e da enorme capacidade de memória do espaço, que encoraja a produção e armazenamento de mais fotografias. No domínio da Ciência da Informação, nomeadamente na gestão de dados de investigação, a fotografia tende a ser um aliado dos investigadores para a captura de prova e registo de factos associados à investigação. Contudo, em diversos casos, este é, também, passível de ser incluído nos seus conjuntos de dados e suscetível de ser organizado e descrito nos sistemas de informação para que, quando necessário, seja reutilizado.

Este trabalho centra-se no uso do registo fotográfico como apoio à investigação, focando o seu estudo num projeto de investigação denominado de Generation.Mobi. O levantamento da produção fotográfica na ciência permite entender com maior clareza o comportamento informacional dos investigadores produtores de fotografia, analisando a relevância desta na investigação, pretendendo mostrar que as imagens são elementos ricos para a recuperação de informação, podendo fomentar a interoperabilidade semântica e a reutilização.

Palavras-chave: Fotografia, Investigação, Gestão de dados

Abstract. New technologies and the imposition of digital have assumed a prominent role in the information society, triggering new trends and becoming highly valorizable resources. In turn, the image is dominant in communication, but photography still has a variety of treatment, in terms of description, interpretation and systematic use. The difficulty of using and handling image collections is known, but it has been decreasing with the potential of automatic

analysis, new image capture devices and the enormous memory capacity of space, which encourages the production and storage of images, more photographs. In the field of Information Science, particularly in the management of research data, photography tends to be an ally of researchers to capture evidence and record facts associated with research. However, in a number of cases, it can also be included in its data sets and can be organized and described in information systems so that, when necessary, it is reused.

This work focuses on the use of the photographic record as a support for research, focusing its study on a research project called Generation.Mobi. The survey of the photographic production in science allows to understand with more clarity the informational behavior of the researchers producing of photography, analyzing the relevance of this in the investigation, pretending to show that the images are rich elements for the information retrieval, being able to foment the semantic interoperability and the reutilization.

Keywords: Photography, Research, Data management

1 Introdução

Capaz de salvaguardar os mais diversos momentos e manifestações, a fotografia ocupou um lugar de destaque na altura do aparecimento da televisão e de outras demonstrações artísticas, como o cinema. Esta tipologia documental, hoje, é vista como parte integrante da sociedade, já que é um mecanismo exímio no estabelecimento de comunicação. Este pode comportar diversas possibilidades interpretativas, mediante a função que lhe é atribuída. Estas funções podem passar por uma componente de negócio, de emoção, noticiosa, de estímulo, de expressão, de memória, de herança ou científica.

As imagens comportam diversas funcionalidades e a capacidade de registo de ações e de informação é uma das mais importantes. É impossível dissociar a imagem da sua competência de ser portadora de materialidade e de recursos de comunicação e expressão distintos, pois são estas características que a aproximam das restantes tipologias documentais, mesmo com as suas particularidades. Todavia, o papel da fotografia na ciência ainda não está claro, embora que o carácter científico dado ao documento fotográfico date do século XIX, altura em que começou a ser utilizada em estudos da Astronomia.

Numa perspetiva de melhor entendimento sobre o documento fotográfico, desde o fenómeno da transformação do analógico para o digital, até à sua inclusão nos projetos de investigação, o presente trabalho pretende apresentar um caso de estudo de um projeto no domínio da gestão de mobilidade dinâmica, denominado de Generation.Mobi. O trabalho começa por apresentar uma revisão de literatura, abordando as noções conceptuais do documento fotográfico e uma visão geral da evolução da produção da fotografia. Posto isto, será analisado o uso da fotografia no Generation.Mobi, começando por apresentação o projeto, até à abordagem relativa ao

fluxo de trabalho. Por fim será apresentada a análise de resultados correspondentes aos questionários realizados aos elementos do Generation.Mobi.

2 O documento fotográfico

Conceptualmente, a fotografia é o resultado de um processo, inerente à arte, em que são registadas reações químicas. Contudo, esta comporta muito mais do que uma representação e reprodução de uma realidade, ela incorpora uma série de características que a tornam complexa e essencial em diversas situações.

Tendo a intenção, a expressão e a inscrição, como componentes centrais da fotografia, Serén (2013) explica que a primeira consubstancia-se com a vontade, isto é, tem uma relação direta com os fenómenos psicológicos. Já a expressão relaciona-se com a maneira como uma ideia é exteriorizada. Por fim, a inscrição caracteriza-se como o elemento necessário para que os dois elementos anteriormente referidos sejam mantidos.

Roland Barthes (2012), um dos mentores da análise da imagem, refere a indiscutibilidade da fotografia como órgão de comunicação, já que, além do papel que desempenha nas pessoas singulares, fomentou o estudo e consolidação da linguagem visual, bem como da liberdade de expressão numa sociedade que passou de oprimida a promotora da comunicação pelos diversos canais.

A sociedade da informação e da comunicação caracteriza-se, sobretudo, pela sua transversalidade e globalidade. Lacerda (2012) vê estas características a serem também promovidas pelas competências comunicacionais da fotografia. Segundo a autora, os documentos fotográficos são autênticos portadores de materialidade e recursos de expressão capazes de mediar a comunicação entre um emissor e um recetor. É no sentido da mediação comunicacional que Boccato e Fujita (2006) apresentam aquela que consideram ser a representação esquemática da linguagem fotográfica, tendo esta, enunciados, textualidades e narrativas dentro do seu texto visual. Para os autores, o emissor é a imagem fixa, o recetor o consumidor dessa imagem e o mediador a linguagem fotográfica.

É também Serén (2013) que refere a fotografia como “uma prova de carácter científico [...] uma réplica da realidade”. Esta afirmação levou a autora a refletir mais intensamente sobre a capacidade de representação inerente à imagem fotográfica, já que é esta competência que permite a distinção entre as pinturas, os desenhos, os esboços e outras tipologias de imagem não tão capacitadas para a duplicação imaculada de uma determinada realidade. É com o início do Modernismo que a valorização da fotografia e, por consequência, da câmara fotográfica, começa a fazer-se sentir.

Serén (2013) relata a forma como a fotografia conseguiu ganhar um estatuto que a posicionou no meio cinematográfico, adquirindo autonomia nos processos de produção. Também no jornalismo a fotografia começou a ganhar destaque pelas suas capacidades de criação de emoção junto do leitor. Keene (2002) sugere que até determinado momento a fotografia era vista no jornalismo como a forma de trazer o leitor para a cena da notícia, contudo esta faceta evoluiu quando se começou a notar

que a fotografia iria muito além da capacidade de representação, ela conseguia desencadear uma série de emoções que o texto não era capaz. Vista como uma sociedade centrada no registo, a sociedade da informação e da comunicação, em que atualmente vivemos, tende a anotar todas as transações. Quem o afirma é Serén (2013) que chega mesmo a afirmar que a fotografia não deve ser despromovida do seu lugar de destaque nesta sociedade. Para a autora, esta tipologia documental constitui, efetivamente, um objeto social, com materialidade e ideias associadas.

Também Schvambach (2009) analisa a forma como a fotografia tem um destaque significativo na sociedade da informação e da comunicação. Para o autor, a dimensão da fotografia está na aptidão para a criação de novas relações sociais e de coexistência, exigência essa que provém de uma sociedade moderna e à procura de novas interpretações, novos conceitos e variadas opiniões e posições, na perspetiva constante de pensamentos inovadores que estejam além da réplica. Isto é, esta sociedade, inundada pela informação e pela comunicação, já mais não procura unicamente apresentar e expor, agora ela visa direcionar ideologias, suscitar comentários, passar significados e, sobretudo, mediar as mais diversas formas de comunicação.

Óbvio será dizer que a produção fotográfica, ao longo dos anos, tem sido significativamente intensificada. Tal facto remete-se para a valorização do visual pela sociedade, bem como para a imposição das novas tecnologias.

Num estudo de 2012 realizado *1000 memories* constatou-se este facto. Segundo o estudo, até 2012, em cada dois minutos eram capturadas mais fotografias do que todas as que foram tiradas durante todo o século XIX. Ainda mais surpreendente é o número estimado de fotografias capturadas até ao ano de 2012 que, segundo o análise realizada, rondava os 3.8 triliões de fotografias.

Outros factos interessantes que foram constatados pelo estudo prendem-se com as redes sociais. Segundo esta investigação, até 2012, o Instagram continha uma média de um bilião de fotografias. Por sua vez, o Facebook, até esse ano, recebia diariamente cerca de 300 milhões de fotografias. Interessante é, também, perceber que até ao ano do estudo, 20 por cento das fotografias que eram produzidas em todo o mundo tinham como destino o Facebook.

3 O uso da fotografia no projeto Generation.Mobi

3.1 O Generation.Mobi

Resultado de um consórcio entre a Ibéria, a Revolution Answer, o Centre of Engineering and Product Development (CEiiA), a Cardio ID e a Faculdade de Engenharia da Universidade do Porto (FEUP), o Generation.Mobi é um projeto de investigação que visa desenvolver e validar um sistema de gestão de mobilidade dinâmica, projetado para a nova geração, cujo conceito alicerça-se na noção de rede social de bicicletas interoperáveis e interativas com os ecossistemas das cidades.

Além das empresas e entidades referidas, o Generation.Mobi tem como parceiro o RENER Living Lab, uma rede Portuguesa de cidades inteligentes. Com esta aliança espera-se que seja potenciado o nível de inovação radical dos serviços e operações relacionados com a mobilidade urbana inteligente, seja a nível nacional, seja a nível internacional.

O projeto de investigação Generation.Mobi esmiuça questões desde a arquitetura de comunicações, modelos e algoritmos de previsibilidade da procura e oferta de serviços de mobilidade, interações com a rede elétrica, modularidade, flexibilidade e portabilidade de micro-hub de mobilidade urbana, arquiteturas flexíveis de bicicletas e tecnologias de biometria cardíaca aplicadas em componentes de bicicletas.

O Generation.Mobi pretende desenvolver um sistema de gestão de mobilidade que incorpore uma rede composta por devices interativos, como são exemplo as bicicletas. Este projeto foi projetado para responder a uma série de desafios sociais no domínio da Estratégia Europeia do Horizonte 2020.

3.2 Fluxo de trabalho do Generation.Mobi

O estudo do projeto Generatio.Mobi, através das interações com os seus membros, permitiu perceber qual o comportamento informacional destes, no que diz respeito ao fluxo de trabalho inerente a este projeto.

Foi perceptível que o Generation.Mobi produz e utiliza fotografias em diversos momentos da investigação. Na fase inicial do projeto, em que é realizado o levantamento de requisitos e planeamento do mesmo, o registo fotográfico é inexistente. Contudo, nas fases seguintes este está sempre presente.

A segunda fase do projeto consubstancia-se com a preparação do material. Este momento é determinante para o sucesso das experiências. Para o responsável do projeto uma má preparação do material que será utilizado nas experiência pode comprometer os resultados. Entre esta preparação está a marcação do piso, o posicionamento de bicicletas, a colocação de antenas e a medição de ângulos, por exemplo. Por vezes, os investigadores não conseguem perceber determinados erros e falhas que acontecem nas fases preliminares e com a utilização da fotografia estes conseguem, numa fase posterior, perceber esses eventuais desvios nos resultados, confirmar cenários e averiguar anomalias no material, sem que seja necessário repetir processos.

Na terceira fase do projeto, inerente aos testes, também é utilizada a fotografia, desta vez para documentar os vários momentos das experiências e dos seus intervenientes. A vantagem do uso da fotografia nesta fase do projeto é a mesma do que a apresentada para a fase anterior. Isto é, permite que, em momentos posteriores, as fotografias possam ser utilizadas para analisar os momentos em que foram produzidas.

Na quarta e última fase do projeto, a análise de resultados, a fotografia deixa de ser produzida. No entanto, o seu uso continua a existir, dessa vez como contributo para

analisar os dados recolhidos. Tal como foi explicado nos parágrafos anteriores, a capacidade de representação que a fotografia potencia, permite aos investigadores perceberem melhor os resultados, possibilitando que estes recuem até ao momento das experiências através das imagens, sem que tenham que repetir as experiências. Esta informação pode ser observada na Figura 6.

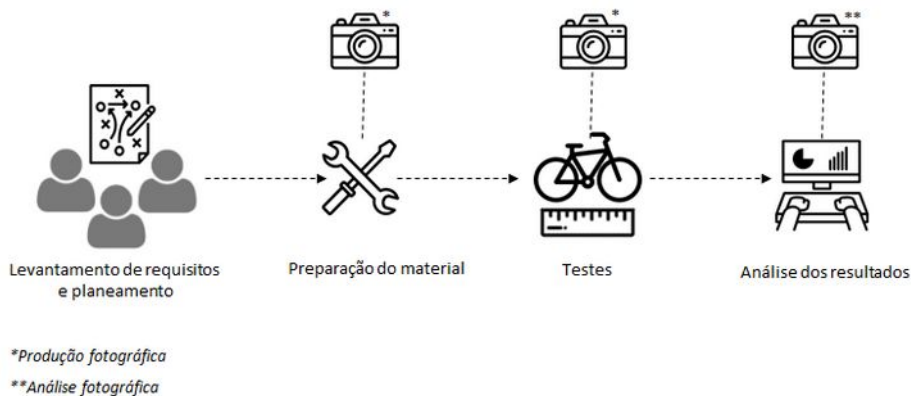


Figura 1. Fluxo de trabalho do projeto Generation.Mobi

3.3 Análise dos resultados dos questionários

Depois de percebido qual o fluxo de trabalho inerente ao Generation.Mobi, através da interação com o responsável do projeto, foi necessário realizar uma pesquisa rigorosa que possibilitasse adquirir dados significativos sobre a utilização da fotografia pelas diversos membros do projeto. Com hipóteses formuladas, foram realizadas sete questões, de forma a ser possível obter uma explicação para o que havia sido especulado.

Sendo o inquérito por questionário uma técnica de observação não participante apoiada numa sequência de perguntas escritas e dirigidas a um conjunto de indivíduos que constituem a amostra de inquiridos, é possível recolher as mais variadas informações factuais sobre a temática em análise.

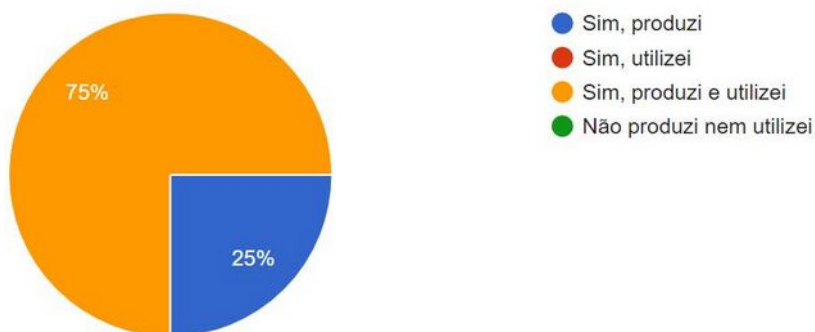
Os critérios para a elaboração do questionário em questão prenderam-se com alguns factores. Primeiro, era necessário que fosse breve, pois facilitaria a predisposição para a sua feitura. Segundo, este teria de ser intuitivo e com um léxico simples, uma vez que, a maior parte dos investigadores ainda não estão familiarizados com a terminologia inerente à gestão de dados e, por esse motivo, na sétima pergunta optou-se por colocar a explicação dos termos inseridos na pergunta, de modo a que fosse fácil de perceber o intuito da questão e das suas opções. Em terceiro, foi fundamental explicar aos inquiridos qual a finalidade do estudo, bem como o domínio de investigação inerente a este. Em quarto, foi imprescindível assegurar o anonimato

dos inquiridos, não só porque a identificação de cada um dos indivíduos não é determinante para a análise de resultados, mas sobretudo pelas questões de privacidade inerentes ao Novo Regulamento Geral de Proteção de Dados. Por estes motivos, o questionário contém um breve cabeçalho onde os inquiridos podem ver explicados e confirmados todos estes pormenores, assegurando, assim, a total transparência entre o requerente e os inquiridos.

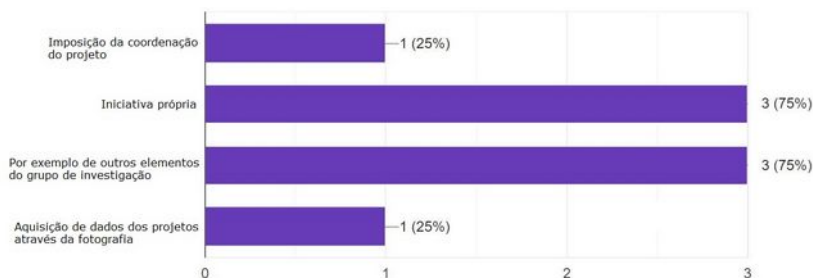
Embora o inquérito por questionário deva ser, maioritariamente, um estudo aplicado a grandes conjuntos de indivíduos, a sua utilização e pertinência não devem ser descartadas quando os grupos são constituídos por menor número. No caso do Generatio.Mobi, foram seis os elementos que integraram o projeto, pelo que esse foi o universo definido para o estudo. Contudo, apenas quatro destes elementos responderam às questões, pelo que os resultados apresentados resumem as respostas destes quatro investigadores. O questionário pode ser visto na sua totalidade no Anexo 1.

- *Análise individual das respostas*

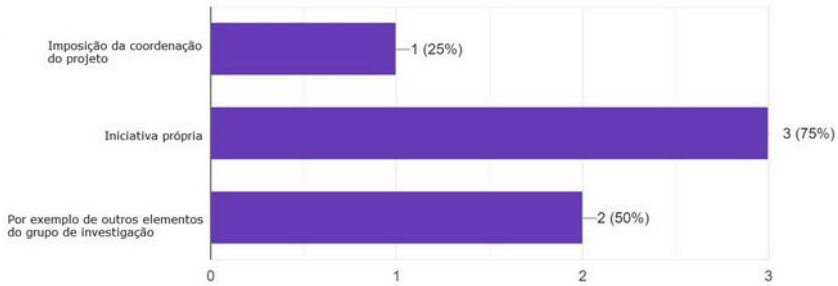
1. Foi um dos investigadores que produziu ou utilizou fotografias no processo de investigação do Generation.Mobi? Caso responda “Não”, passe para a questão 5.



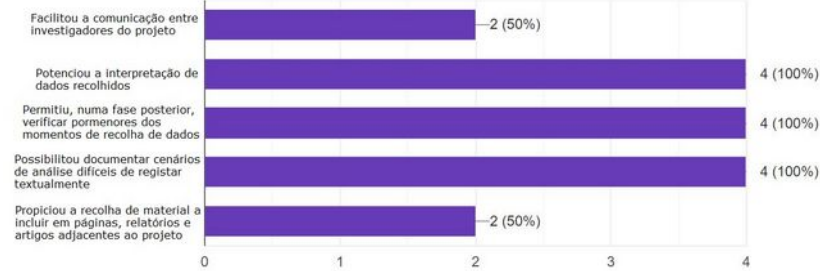
2. Quais as condições que O levaram a produzir fotografias?



3. Quais as condições que O levaram a utilizar fotografias?



4. Quais considera terem sido as vantagens da inclusão da fotografia no processo de investigação do Generation.Mobi? Selecione até 3 das seguintes opções.



5. Considere o seguinte cenário: um investigador do projeto envia-Lhe uma fotografia que relata um determinado momento de recolha de dados no qual não esteve presente. Considera que irá facilmente interpretar a informação que relata esse momento ou preferiria o mesmo conteúdo da fotografia, mas documentado textualmente?

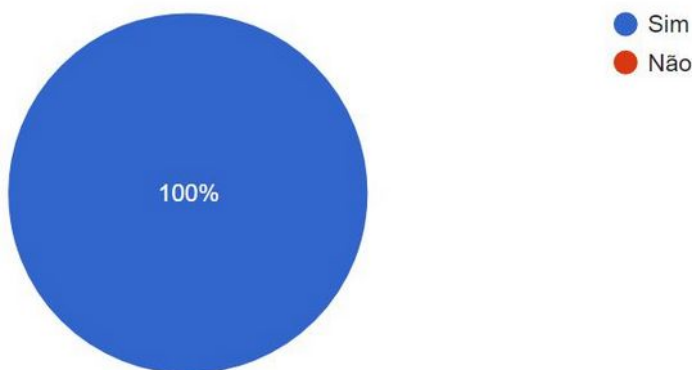
A fotografia deveria ser enviado com um texto contextual

Acho que as fotos podem mostrar como foram realizadas as experiências mas não as tecnologias usadas para a recolha de dados. Isto é, nem em todas as experiências usamos os mesmos scripts e programas para a recolha de dados e acho que isso é difícil de se enquadrar numa fotografia.

Facilmente irei interpretar o contexto com fotografia

O conteúdo da fotografia documentada textualmente seria preferível.

6. Gostaria de ver a fotografia acompanhada de elementos textuais que lhe atribuísem contexto (como descrição, palavras-chave, cobertura espacial e temporal...)?



7. Dos seguintes, quais considera serem os elementos fundamentais para descrever uma fotografia no domínio da investigação?



4 Conclusões

Este trabalho permitiu refletir sobre diversos aspetos relativos à imagem científica, em particular à fotografia inerente às práticas de investigação. Na verdade, a avassaladora dominância das novas tecnologias e as novas práticas digitais potenciam que a fotografia e a sua produção cresçam vertiginosamente, já que através dos mais diversos dispositivos, a fotografia está facilmente alcançável.

A utilização de imagens tem um efeito de recordação significativo no cérebro humano que estabelece, mais rapidamente, ligações entre os objetos e os seus contextos, potenciando um entendimento mais célere e eficiente. A fotografia não pode ser despromovida do seu carácter científico e, por conseguinte, não pode ser afastada do processo documental, já que esta contém uma linguagem e é passível de interpretação e tratamento.

Por um lado, a fotografia pode constituir um elemento valioso na salvaguarda de informação, dado que capta, com precisão, os vários componentes fotografados. Desta forma é dado aos investigadores a possibilidade de ver vários pormenores e assegurado que nada é esquecido pela memória humana que, pelas suas limitações, não garante a preservação de todas as particularidades. Assim sendo, a lembrança humana, com o auxílio da fotografia, é tornada natural. Por outro lado, a fotografia pode ser uma consequência do processo de investigação e, por isso, um dos constituintes de um conjunto de dados. Por este motivo, é passível de ser analisado e processado, de modo a obter uma série de resultados e conclusões. Assim sendo, a fotografia será suscetível de descrição, com vista a dar contexto e alcançar um nível de profundidade que de outra forma não seria possível. Com a descrição é evitada a perda de informação e a banalização das fotografias, melhorando o processo de investigação desde as fases iniciais às finais, fazendo com que o tempo dos investigadores seja rentabilizado.

A descrição é um dos alicerces base da gestão de dados, a sua feitura por parte dos investigadores acarreta inúmeras vantagens. Todavia, estes procuram que o resultado do seu trabalho seja sempre potenciado e o tempo apreendido diminuído. Do ponto de vista da imagem, particularmente da fotografia, existem ferramentas que auxiliam e facilitam a identificação da informação que nela está contida, nomeadamente através da extração de atributos e conceitos e do reconhecimento de padrões e objetos, para que a posterior interpretação seja melhorada e as imagens ganhem significado. Estas ferramentas, inerentes ao processamento automático da imagem, têm o intuito de derrubar os limites da visão humana e oferecer aos utilizadores o máximo de informação sobre determinada imagem.

É inegável que o processamento automático de imagens pode ser um contributo importante nos sistemas de informação que querem garantir uma gestão de dados eficiente. No entanto, este poderá acontecer como estágio preparatório para uma fase de descrição mais completa e pormenorizada. Os investigadores que se apoiam em exclusivo na automatização poderão não conseguir ver validadas as ações realizadas, não existirá uma verificação do erro e correção deste.

Só um sistema flexível e capaz de se adaptar às necessidades concretas de descrição de cada domínio de investigação poderá dar o contexto esperado e atribuir significado relevante ao que está a ser descrito. Só assim um investigador conseguirá aceder aos seus dados e, num curto espaço de tempo, relembrar pormenores importantes da sua pesquisa, mesmo que o seu projeto já tenha terminado. Os registos de carácter imagético, tal como os restantes, passam por várias fases, no que toca ao ciclo de vida da informação. Entre essas fases estão a captura (fase inicial) e a recuperação e reutilização (fase final). No entanto, fases intermédias como a organização e descrição garantem que o ciclo corre devidamente.

Existindo uma normalização no que diz respeito ao trabalho a desempenhar com a fotografia científica, as práticas inconsistentes serão descartadas, dando lugar a técnicas sólidas e em que se diminuirá a dificuldade em dar significado às fotografias. Para que tal facto aconteça é necessário perceber em que contornos acontece a

produção fotográfica na ciência e entender o comportamento informacional dos investigadores que utilizam a fotografia como suporte à investigação. Poder-se-á, ainda, sensibilizar para a sua importância, consciencializar para as problemáticas de gestão e averiguar se as suas necessidades informacionais são mais prontamente saciadas através da utilização de técnicas de gestão de dados de investigação. Na verdade, a educação dos investigadores em relação a esta problemática possivelmente desencadeará um comportamento mais consistente e ativo em relação à fotografia no contexto de investigação.

Referencias

- Barthes, Roland. (2012). *A Câmara Clara*. Lisboa: Edições 70
- Bocato, Vera R. C.; Fujita, Mariângela S. L. 2006. Discutindo a Análise Documental de Fotografias: Uma Síntese Bibliográfica”. Brasil: *Cadernos de Biblioteconomia, Arquivística e Documentação*, 2, 84–100
- Generation.Mobi. Disponível em: <https://generationmobi.ceiia.com/>
- Huds, Dirk. (2017). *The Impact of Photography on Societ*. Our Pastimes. Disponível em: <https://ourpastimes.com/the-impact-of-photography-on-society-12377030.html>
- Lacerda, Aline Lopes de. (2012). A Fotografia nos Arquivos: Produção e Sentido de Documentos Visuais. *História, Ciências, Saúde*, 19 (1): 283–302;
- Keene, Martin. (2002). *Fotojornalismo: Guia profissional*. Dinelivro
- Schvambach, Janaina. (n.d.). *A Fotografia como Fonte de Pesquisa e sua Ficção Documental*
- Serén, Maria do Carmo. (2013). O Documento Fotográfico: da Mediação Cultural à Mediação Técnica. *Revista do CITCEM*. Porto: 183-192
- 1000 Memories. *Number of photos ever taken*. Disponível em: <http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox>

Anexos

- Questionário

A utilização de fotografias nos processos de investigação do projeto Generation. Mobi

No âmbito do projeto Generation. Mobi em que esteve envolvido, solicito a sua colaboração no preenchimento deste breve questionário que visa averiguar o papel da fotografia no processo de investigação. O seu preenchimento não ultrapassará os cinco minutos. Todos os dados serão tratados de forma confidencial.

*Obrigatório

Foi um dos investigadores que produziu ou utilizou fotografias no processo de investigação do generation. mobi? Caso responda "Não", passe para a questão 5. *

- ☐ Sim, produzi
- ☐ Sim, utilizei
- ☐ Sim, produzi e utilizei
- ☐ Não produzi nem utilizei

Quais as condições que O levaram a produzir fotografias?

- ☐ Imposição da coordenação do projeto
- ☐ Iniciativa própria
- ☐ Por exemplo de outros elementos do grupo de investigação
- ☐ Outra: _____

Quais as condições que O levaram a utilizar fotografias?

- ☐ Imposição da coordenação do projeto
- ☐ Iniciativa própria
- ☐ Por exemplo de outros elementos do grupo de investigação

Dos seguintes, quais considera serem os elementos fundamentais para descrever uma fotografia no domínio da investigação? *

- ☐ Subjects ou Keywords (the topic of the resource)
- ☐ Description (an account of the resource)
- ☐ Author (an entity primarily responsible for making the resource)
- ☐ Title (a name given to the resource)
- ☐ Methodology (systematic, theoretical analysis of the methods applied to study)
- ☐ Instrument (type of instrument used for data collection or capture)
- ☐ Format (the file format, physical medium, or dimensions of the resource)
- ☐ Spatial Coverage (spatial characteristics of the resource (e.g. point in space, geographic limits))
- ☐ Temporal Coverage (temporal characteristics of the resource (e.g. period in time, time interval))
- ☐ Type (the nature or genre of the resource)
- ☐ Dimensions (a measurable extent of a particular kind, such as length, breadth, depth, or height)
- ☐ Outra: _____

Quais considera terem sido as vantagens da inclusão da fotografia no processo de investigação do generation. mobi? Seleccione até 3 das seguintes opções.

- ☐ Facilitou a comunicação entre investigadores do projeto
- ☐ Potenciou a interpretação de dados recolhidos
- ☐ Permitiu, numa fase posterior, verificar pormenores dos momentos de recolha de dados
- ☐ Possibilitou documentar cenários de análise difíceis de registar textualmente
- ☐ Propiciou a recolha de material a incluir em páginas, relatórios e artigos adjacentes ao projeto
- ☐ Outra: _____

Considere o seguinte cenário: um investigador do projeto envia-lhe uma fotografia que relata um determinado momento de recolha de dados no qual não esteve presente. Considera que irá facilmente interpretar a informação que relata esse momento ou preferiria o mesmo conteúdo da fotografia, mas documentado textualmente? *

A sua resposta: _____

Gostaria de ver a fotografia acompanhada de elementos textuais que lhe atribuissem contexto (como descrição, palavras-chave, cobertura espacial e temporal...)? *

- ☐ Sim
- ☐ Não

Análisis de las etiquetas de contenido de un banco de imágenes: *Agefotostock*

Críspulo Travieso-Rodríguez¹ y Blanca Rodríguez-Bravo²

¹ ORCID [0000-0002-0774-0728](https://orcid.org/0000-0002-0774-0728). Departamento de Biblioteconomía y Documentación, Universidad de Salamanca, España.
ctravieso@usal.es

² ORCID [0000-0002-9476-7602](https://orcid.org/0000-0002-9476-7602). Área de Biblioteconomía y Documentación, Universidad de León, España.
blanca.rodriguez@unileon.es

Resumen. El objetivo fundamental de este trabajo es analizar las etiquetas asignadas a los documentos fotográficos en una agencia comercial, *Agefotostock*, con el fin de conocer en qué aspectos inciden y cuáles son las características más comunes de estos descriptores libres. De entre las nueve categorías temáticas ofrecidas, se optó por la denominada “Gente” y por seleccionar, las primeras 100 fotos recuperadas en la última quincena de abril de 2019 ordenadas en función de su popularidad, lo que aseguraba contar con una muestra suficientemente amplia. El estudio de la muestra se abordó tomando los siguientes elementos de análisis: número de etiquetas asignadas por fotografía; tipo y morfología de los términos; número de términos por etiqueta; porcentaje de etiquetas de contenido frente a etiquetas de formato o técnicas; porcentaje de etiquetas denotativas y connotativas, coincidencia de etiquetas y sinonimia. Los resultados muestran una indización muy exhaustiva en la que predominan las etiquetas conformadas por un único vocablo que son mayoritariamente sustantivos, si bien la presencia de adjetivos y verbos no es desdeñable como cabía esperar para poner de manifiesto cualidades y acciones observables en las imágenes.

Palabras clave: *Agefotostock*, bancos de imágenes, etiquetas, fotografías, indización.

Abstract. The main objective of this work is to analyze the tags assigned to the photographic documents in a commercial agency, *Agefotostock*, in order to know what are the most common characteristics of these free descriptors. Among the nine thematic categories offered, the so-called “Gente” (people) was chosen and it was decided to select the first 100 photos retrieved in the last fortnight of April 2019 ordered according to their popularity, which assured to have a sufficiently large sample. The study of the sample was approached taking the following elements of analysis: number of tags assigned by

photography; type and morphology of the terms; number of terms per tag; percentage of content tags versus format tags or techniques; percentage of denotative and connotative tags, coincidence of tags and synonymy. The results show a very exhaustive indexing in which tags made up of a single word that are mostly noun predominate, although the presence of adjectives and verbs is not insignificant as might be expected to show observable qualities and actions in the images.

Keywords: *Agefotostock*, databases, images, indexing, tags.

1 Introducción

La fotografía, como apuntan Agustín y Torregrosa (2019): “Es información y es arte. Es documento y obra estética. Historia viva y presente continuo. La fotografía es luz y es memoria. Evidencia y sugerencia” Esta introducción servirá para contextualizar el trabajo que aborda las particularidades de las imágenes que actúan como condicionantes para su indización, así como las características del etiquetado libre muy utilizado de manera colaborativa en diversas plataformas.

1.1 La indización de Imágenes

Entre los elementos que condicionan la representación y la posterior recuperación de la información, uno de los más determinantes es la forma de presentación del contenido de los documentos. Y aunque tradicionalmente los presupuestos y pautas de este análisis se han aplicado a documentos escritos, las colecciones con otros componentes formales, como puedan ser las fotográficas, sonoras o audiovisuales, precisan la adaptación de esas técnicas al código y al medio de transmisión específicos de estos datos.

En función de cómo son percibidos y procesados los documentos no bibliográficos, podemos señalar dos tipologías: visuales (imagen), que son asimilados a través de su compleja organización visual y que son sincrónicos; y sonoros, que son diacrónicos, esto es, lineales y supeditados al factor tiempo, en el sentido de que no pueden ser procesados de forma libre y voluntaria. A ello se añaden los documentos audiovisuales, que combinan imagen y sonido, y que también son lineales (Rodríguez Bravo, 1996).

Como señala Moreiro González (1993), así como estos dos últimos tipos de documentos pueden ser tratados de forma más similar al análisis de los textos escritos (bien sea porque la descripción se basa en el guion de la representación audiovisual o en la transcripción escrita del discurso oral), la indización de las imágenes implica una serie de peculiaridades. En palabras de Smit (1987), la descripción de una imagen nunca es completa y no se trata de una cuestión de nivel de especificidad. Quien

trabaja con imágenes trabaja con más detalles, más información y, principalmente, con informaciones menos evidentes.

Dentro de las cualidades específicas que presenta la descripción de una imagen, además del ya mencionado carácter sincrónico, se ha de destacar también su transparencia (Smit, 1987), entendida como la correspondencia entre la realidad y su representación, entre lo mostrado y lo real. Su código icónico obliga a realizar una traducción al código verbal en el momento en que es sometida a análisis, con las dificultades que ello implica al obligar a traducir imágenes de realidades concretas a conceptos. Ello plantea ciertas limitaciones: pérdida de significado y exceso de ruido y distorsión, máxime si tenemos en cuenta que no existen imágenes de conceptos, sino de realidades, en virtud de la "transparencia" anteriormente aludida. Otras particularidades propias de este tipo de documentos son su adaptabilidad a distintos contextos (el mensaje informativo puede servir para ilustrar distintas situaciones, distantes incluso a aquella en la que fue producida) y, esta singularidad deriva de su polisemia, admitiendo distintas lecturas e interpretaciones en función del marco en que se contemple o de la persona que realice esa interpretación, como ya apuntaba Valle Gastaminza (1993). En el análisis de las imágenes interviene la memoria visual, produciéndose asociación de significados. En la interpretación visual influye el background y la personalidad del receptor (Acal Díaz, 2015) de ahí el interés en preservar todos los posibles significados de la imagen.

En el caso del análisis de las imágenes cobran relieve los recursos creativos o la retórica que componen su estructura formal, añadiendo un valor simbólico, connotativo y subjetivo a la información transmitida. El estilo define la función informativa (Vélez Cea, 2000), por ello al realizar un análisis de contenido debemos atender a la estructura plástica del documento y a los valores semánticos derivados de su iconicidad, funcionalidad, recursos connotativos y retórica, además de atender al texto si lo hay y a su relación con la imagen.

Según Valle Gastaminza (1993) el análisis de contenido de las imágenes se articula en dos niveles totalmente diferentes:

- Análisis morfológico, que estudia las características técnicas, formales y de composición de la imagen. Los aspectos morfológicos son trascendentales en la indización de las imágenes. La forma de fotografiar algo influye en su interpretación, dado que no existe la supuesta objetividad del documento visual. Por tanto, hay precisiones técnicas -el tipo de plano, el tipo de objetivo, la estructura de la representación, la luz, el eje de la toma, etc.- que son necesarias desde el punto de vista documental pues ayudan a discriminar tanto como los motivos representados (Rodríguez Bravo, 2000). Las demandas de información de los usuarios y sus criterios para aceptar los resultados de la búsqueda no se basan únicamente en los conceptos representados en el documento, el cómo se representan es tanto o más decisivo a la hora de la elección. La importancia de la forma en que se presenta el contenido es aplicable también a los documentos escritos, como han puesto de relieve Espelt y Gascón (1999). Pero las variables a considerar son más en el caso de la imagen y su trascendencia mayor.

- Indización, que implica en primer lugar su transcodificación (Smit, 1987), la traducción de ciertos elementos de esa imagen desde un código icónico a un código verbal. La representación del contenido de la imagen atenderá a dos elementos diferentes: los elementos referencialmente existentes en la imagen –denotación- y lo que esas representaciones sugieren –connotación-. El contexto en el que se encuentra la imagen también sirve de ayuda para poder ser más precisos en algunas imágenes, como es el caso de fotografías científicas o de actualidad. Ese contexto puede venir dado por el texto que acompaña a la fotografía o por el reportaje en que se inserta. Asimismo, Valle Gastaminza (2002) argumenta que de los documentos en imagen se deriva una triple información, la denotada, la connotada y el contexto que es preciso descodificar para representarlos de forma adecuada y con ello facilitar su acceso y difusión. Pinto, García y Agustín (2002) abundan en esta línea estableciendo tres niveles para el análisis de contenido de los documentos visuales: descriptivo, identificativo e interpretativo.

En la sociedad contemporánea existe un dominio de lo visual. Donde su expansión ha sido mayor es en el ámbito publicitario como señalan Marcos, Sánchez y Villegas (2005). Las imágenes requieren, así, una atención prioritaria. En consonancia con su relevancia han sido numerosos los estudios que abordan esta temática. Podemos mencionar el estudio de Angus, Stuart y Thelwall (2010) que analiza las imágenes en Flickr para conocer la frecuencia de los términos usados para etiquetar el contexto de la ubicación de los documentos en imagen. Asimismo, cabe destacar el análisis llevado a cabo por Wu (2013) que indaga en el comportamiento del usuario en la búsqueda de imágenes, o el análisis de las características compositivas y técnicas llevado a cabo por Guimaraes, Almeida y Torres (2014). También resultan de interés algunos estudios basados en el análisis de los elementos icónicos del documento publicitario y la decodificación de la información connotada con fines persuasivos como es el de Acal Díaz (2015). En publicidad el análisis de la connotación, y la representación de sensaciones, sentimientos y emociones que despiertan en el receptor cobra especial significado.

1.2 Etiquetado con Descriptores Libres

Al margen de los aspectos relacionados con la descripción de fotografías desde el enfoque conceptual de su análisis, la representación también lleva aparejada la toma de decisiones en cuanto a cómo se representan los temas que la fotografía contiene. Y no es casual que sea precisamente este tipo de documento el que de forma más evidente se haya prestado a la descripción mediante etiquetas en lenguaje libre frente a la opción de los vocabularios controlados. En buena medida, esta circunstancia puede explicarse por la proliferación de plataformas y redes que facilitan la disponibilidad de fotografías en la red de manera colaborativa. Algo habitual en el escenario surgido tras la web social, el hecho de que la propia comunidad establezca sus formas de organizar el conocimiento de modo participativo se acentúa con la

aparición de herramientas como Flickr, Pinterest, Instagram, etc. Un sistema de indización de abajo a arriba proporciona etiquetas orientadas a satisfacer las necesidades de los usuarios (Moreiro; Bolaños, 2018).

Frente al inmovilismo que se suele achacar a ciertos lenguajes documentales y su falta de actualización a la hora de adaptarse a nuevos conceptos, las folksonomías o etiquetados colaborativos han ido cobrando importancia en entornos cada vez más dinámicos, como las redes sociales y las herramientas en acceso abierto. Se han revelado, por tanto, como nuevos modelos de representación de contenidos que permiten una organización y gestión social de la información digital (Rodríguez Roche; Pérez Sanchidrián, 2014). Este tipo de indización en lenguaje natural tiene mayor incidencia en aquellos contextos en los que no se precisa ningún conocimiento teórico previo para llevar a cabo la indización (Rodríguez Yunta, 2009).

Si bien la ausencia de control terminológico y de relaciones semánticas entre los conceptos dificulta la elaboración de estrategias de recuperación y la efectividad de las búsquedas, también es cierto que en algunos entornos se prima la descripción libre mediante etiquetas. Y esta tendencia también está implantándose en unidades y sistemas de información que por la naturaleza de la documentación que gestionan prefieren el empleo del lenguaje natural frente al mantenimiento de un sistema de organización de conocimiento más estructurado. En definitiva, prevalece el interés por conseguir una mayor exhaustividad en la descripción facilitada sobre la necesidad de buscar la máxima precisión de los conceptos recogidos (Soler Monreal; Gil Leiva, 2010).

Este es el caso de la indización de las imágenes. La lectura de la imagen es más rápida que la lectura de su resumen o de sus términos de indización de ahí que la elección de la imagen apropiada para el usuario debe realizarse por este comparando directamente imágenes. Así, en la indización de las imágenes existe una regla básica: potenciar la exhaustividad en detrimento de la precisión. La indización de las imágenes proporcionará un amplio abanico de términos que derivaran de los significados denotados por la imagen y proporcionados por el contexto –términos concretos-, los significados derivados de la connotación –términos abstractos- y aquellos que representan los aspectos de representación formal del contenido. Por último, conviene considerar que la representación de las acciones y actitudes observables en las imágenes requieren la introducción de formas verbales, y las cualidades de objetos y personas conducen a la incorporación de adjetivos. Es habitual el uso de sinónimos, singulares y plurales del mismo sustantivo y/o adjetivo o de varias formas verbales. La exhaustividad trae aparejada un alto grado de redundancia en la indización.

2 Objetivo y Metodología

El objetivo fundamental de este trabajo es analizar las etiquetas asignadas a los documentos fotográficos en una agencia comercial concreta, con el fin de conocer en

qué aspectos inciden y cuáles son las características más comunes de estos descriptores libres. Se trata de analizar, en un contexto no académico, qué prácticas o patrones de descripción se revelan como más frecuentes en la indización de imágenes cuando no se emplea un vocabulario controlado.

2.1 Selección de la Fuente de Datos.

Entre los distintos recursos posibles para recopilar documentos fotográficos que contasen con una descripción de contenido se optó por emplear *Agefotostock*¹, agencia fotográfica fundada en 1973 en Barcelona que pone a disposición en su web una amplia colección de imágenes que comercializa para diversos usos. En concreto, gestiona los derechos de reproducción de imágenes fotográficas y clips de vídeo para su uso en el sector corporativo, editorial, publicitario y de diseño. Con oficinas en Barcelona y Madrid y propietaria de compañías en Nueva York y París, está especializada en hosting y distribución de fotografía de stock a través de su propia plataforma.

Esta elección obedeció a varias razones. Por un lado, se comprobó que era una fuente de información que permitía la consulta pública a partir de etiquetas de contenido. Además, el número de descriptores asignados a las fotografías era lo suficientemente amplio para aplicar diversos criterios de análisis. Como se pensaba estudiar aspectos lingüísticos concretos, también el hecho de que las palabras clave estuviesen en castellano fue un aspecto favorable en la consideración de esta fuente como objeto de análisis. Por último, se estimó como relevante que no perteneciese a un ámbito académico, sino que sirviera de base para distintas actividades (comerciales, publicitarias, artísticas).

2.2 Selección de la Muestra

De entre las nueve categorías temáticas ofrecidas, se optó por la denominada Gente, pues se infirió a priori que sería la que más enfoques podría aportar sobre una misma fotografía. Como se comprobó en las consultas iniciales, la presencia de significados connotativos era previsiblemente mayor que en otras categorías más neutras, en las que se tendía a aplicar unas etiquetas más ajustadas a lo representado realmente en la fotografía.

Finalmente se seleccionaron, dentro de la categoría mencionada, las primeras 100 fotos recuperadas en la última quincena de abril de 2019 ordenadas en función de su popularidad, lo que nos aseguraba contar con una muestra suficientemente amplia.

¹<https://www.agefotostock.com/age/es/>

2.3 Criterios de Análisis

El estudio de la muestra se abordó tomando los siguientes elementos de análisis:

- Número de etiquetas asignadas por fotografía.
- Número de términos por etiqueta.
- Morfología de los términos.
- Porcentaje de etiquetas de contenido frente a etiquetas de formato o técnicas.
- Porcentaje de etiquetas denotativas y connotativas.
- Coincidencia de etiquetas y sinonimia.

3 Resultados

Los resultados se desglosan siguiendo los indicadores establecidos en el apartado anterior.

3.1 Número de Etiquetas Asignadas por Fotografía

Se recopiló un total de 6775 etiquetas, lo que devolvía un número medio de etiquetas asignadas por fotografía más elevado de lo que se había estimado en un principio; ello da muestra de la exhaustividad de la indización de este banco de imágenes. De ese total, se eliminaron las que estaban repetidas en el conjunto, resultando una lista de 1921 etiquetas. Estas fueron verificadas, para descartar las que tenían algún error tipográfico o formal, lo que hizo que la muestra final estuviese conformada por 1913 etiquetas.

En el análisis de la composición y la morfología de las palabras clave que mostramos a continuación se han descartado las etiquetas representadas por números (un total de 73 etiquetas); todas ellas hacían referencia a rangos de edad de las personas representadas en la fotografía. La forma más habitual en que ese concepto estaba redactado era mediante números unidos por guion (por ejemplo, “25-29”), si bien se observaron variantes formales de ese mismo concepto que solo aportaban redundancia a la indización.

3.2 Número de Términos por Etiqueta

En relación con la estructura de las etiquetas, la más frecuente es la de unitérminos o etiquetas simples (78,52%) frente a las etiquetas compuestas o sintagmáticas (21,48%). En el caso de las palabras clave compuestas dominan las conformadas por dos términos (11,87%) seguidas por las estructuradas en tres partes o cuatro partes (6,83%) y (2,60%) respectivamente. La presencia de etiquetas de 5 y 6 palabras es

meramente testimonial. La tabla I muestra la composición de las etiquetas sintagmáticas.

Tabla 1.Número de términos por etiqueta

N.º de palabras	Etiquetas	Porcentaje
1	1448	78,52%
2	219	11,87%
3	126	6,83%
4	48	2,60%
5	2	0,10%
6	1	0,05%
Total	1844	100%

Se constata, por tanto, una tendencia a la sencillez en la indización siempre que es posible. No obstante, la precisión en la representación de los conceptos específicos hace necesario combinar términos en algunas ocasiones.

3.3. Morfología de los Términos

La morfología es extremadamente variada, habiéndose identificado 54 formas entre las que predominan las siguientes: sustantivo común (946 ocurrencias), adjetivo (276 ocurrencias) y la fórmula nombre + adjetivo (137 ocurrencias). En la siguiente tabla puede observarse esa heterogeneidad, pues recoge las distintas secuencias morfológicas encontradas y que al menos se repetían en dos ocasiones.

Tabla 2.Secuencias morfológicas de las etiquetas, según su frecuencia

Secuencia morfológica	N.º de etiquetas
Sustantivo común	946
Adjetivo	276
Nombre + Adj	137
Verbo	135
Nombre + Prep + Nombre	90
Nombre propio	82
Numerales	73
Adj + Nombre	29
Nombre + Prep + Art + Nombre	25
Adverbio	10
Nombre + Nombre	10
Prep + Nombre	8
Abreviatura	6
Pronombre	5
Art + Nombre	4
Nombre + Conj + Nombre	4
Adv + Adj	3
Nombre + Prep + Adj + Nombre	3

Nombre + Prep + Nombre + Adj	3
Prep + Adverbio	3
Prep + Art + Nombre	3
Adj + Prep + Art + Nombre	2
Adv + Art + Nombre	2
Adv + Prep + Nombre	2
Prep + Nombre + Prep + Nombre	2

Al margen de estas, se pudo identificar además otras 22 secuencias, representadas con solo una etiqueta, y donde la variedad llegó incluso a etiquetar conceptos con una conjunción (“and”). Esta circunstancia no es ajena a cualquier indización en lenguaje natural en la que ni antes ni después de la descripción se realiza ningún tipo de control del vocabulario.

En todo caso, si observamos la lista de términos recopilados desde la óptica de las pautas tradicionales o normativas de asignación de términos, llama la atención el abundante uso del adjetivo, bien acompañando a un nombre o bien como un término de indización en sí mismo (“alegre”, “azul”, “ensimismado”), algo, por otro lado, muy habitual en la práctica y aplicación del etiquetado social, así como imprescindible en las fotografías de una categoría que muestran gente con unas características y actitudes concretas. En cuanto a los nombres propios, todos ellos hacían referencia a aspectos geográficos, como forma de situar donde fue tomada la fotografía en cuestión – si bien en la mayoría de estos casos, se trataba de un lugar no identificable en la fotografía, por lo que a efectos de recuperación de la información no encajaría con una descripción de contenido -.

Otra de las directrices comunes en la indización es evitar el uso de verbos y optar por las formas sustantivadas. De hecho, los vocabularios controlados tradicionalmente se han limitado al uso de sustantivos y formas sustantivadas. No obstante, la indización de las imágenes requiere la presencia de verbos para representar acciones como constatan diversos estudios (García Jiménez, 2007; Marzal García-Quismondo et al., 2002). En cuanto a las formas verbales identificadas en la muestra, cabe señalar que la gran mayoría se correspondía con la forma no personal gerundio (“subiendo”, “sujetando”, “restaurando”), como una vía directa y sencilla de describir exactamente qué está sucediendo en la fotografía. Esta preeminencia era así tanto cuando se trataba de etiquetas verbales simples como cuando estaban formadas por más palabras (“mirando de reojo”, “saludando con la mano”).

En lo que respecta a los adverbios recogidos, solo hubo dos tipos: o de modo (“curiosamente” “convenientemente”) o de lugar (“encima”, “dentro”). Si bien ambos son susceptibles de ser empleados en una descripción discursiva de una imagen, su utilidad o adecuación en la recuperación de información mediante palabras clave es bastante deficiente.

En cuanto al uso de los conceptos en su forma singular o plural, y tras descartar los tipos de palabras que no admiten variación de número (nombres propios, formas verbales y adverbiales), se comprobó que para el resto (1656 etiquetas) predominaba de forma incontestable el singular (así estaba redactado el 89,85%) frente al plural

(que solo representaba un 10,15%). Este dato, esperable en todo caso, contradice también lo habitual en otros entornos donde se hace uso de vocabularios controlados. También cabe reseñar que apenas se dieron casos en los que se asignaban a los documentos ambas formas, práctica habitual cuando se prima la exhaustividad sobre la precisión.

3.4 Palabras Clave de Contenido/Técnicas

Al cotejar si las etiquetas se referían a aspectos técnicos o reflejaban el contenido mismo de la foto, la inmensa mayoría era de esta última condición (98,59%), mientras que solo 27 etiquetas hacían referencia a alguna cuestión técnica del documento (como “plano cenital”, “plano americano”, “soft focus”), no alcanzando el 2%. En este sentido, se considera importante tener en cuenta la incidencia de la elección de la fuente de los datos que se ha empleado en este estudio y el público al que va dirigido principalmente, pues no deja de ser una agencia que prima el aspecto comercial o publicitario frente a consideraciones del ámbito profesional de la fotografía y sus métodos.

Sobre la adecuación o no de la descripción a los elementos objetivamente representados, predominan las que se dedican a describir el contenido de la imagen por medio fundamentalmente de etiquetas denotativas (70,17%) frente a un 29,83% de palabras clave connotativas, esto es, destinadas a describir lo que la imagen sugiere o representa en nuestro contexto cultural. En este último grupo encontramos tanto términos concretos que no están representados de forma explícita en la fotografía y cuya inferencia implica subjetividad (como, por ejemplo, “viajes de negocios”, “fin de semana”, “soltería”) como términos abstractos que de por sí no pueden estar reproducidos mediante una imagen, pero pueden estar implícitos para el uso posterior que pretende darse a la imagen (“liderazgo”, “sentimiento positivo”, “incomunicación”).

3.5 Coincidencia de Etiquetas y Sinonimia

En cuanto a la presencia de términos coincidentes, se identificó, como se ha comentado anteriormente, un elevado número de términos redundantes en la muestra seleccionada, superando el 70%. Estos términos muestran además poca uniformidad y abundancia de sinonimia o de formas distintas de la misma familia semántica. En lo que se refiere a las etiquetas más repetidas, la tabla III muestra las que fueron asignadas al menos a 30 fotografías distintas del total analizado.

Tabla 3. Etiquetas más frecuentes

Etiqueta	N.º de veces asignada
Gente	98

Color	88
Raza Blanca	64
Día	63
Mujer	61
Adulto	59
Exteriores	55
Hombre	51
Imagen	44
Fotografía	44
Adulto Joven	43
Interior	42
Sonrisa	40
Ocio	37
Lifestyles	37
Apariencia	37
Sólo Una Mujer	36
Foto	36
Europeo	36
Sexo Femenino	35
Vista	34
Sexo Masculino	32
Una Persona	31
Horizontal	31
Feliz	31
Mirando	30

Como se observa en esa relación, las etiquetas más comunes aluden a descripciones de las figuras humanas que aparecen en las fotografías, haciendo referencia fundamentalmente a aspectos muy básicos de su presencia: género (“mujer”, “hombre”, “sexo femenino”, “sexo masculino”), o a su origen étnico o procedencia (“raza blanca”, “europeo”). Llama la atención que precisamente el título dado a la colección (“gente”) sea precisamente la etiqueta más empleada (98 de las 100 fotos la contenían), lo que merma absolutamente su capacidad como término de búsqueda y filtro de información. La segunda por orden de frecuencia fue “color”, si bien hay que señalar que no es que las 12 fotos a la que no se había asignado este concepto estuviesen en blanco y negro, con lo que la consistencia de esta asignación ha de cuestionarse.

También es destacable la aparición como etiquetas recurrentes de algunos términos con bajo poder semántico o discriminador a la hora de realizar búsquedas de información específicas en este tipo de colecciones (“apariencia”, “vista”).

Sin embargo, parece lógico que determinadas palabras descriptivas sobre el entorno de una imagen sí aparezcan entre las más empleadas (“exteriores”, “interior”, “día”). Por último, se detectaron etiquetas frecuentes que, más que encaminadas a la descripción del contenido, parecen haberse asignado indefectiblemente al tipo de

documento sin la intención de que se empleen en la búsqueda (“foto”, “imagen”, “fotografía”).

Se consignaron 93 etiquetas que estaban escritas en otro idioma (92 de ellas en inglés y una en latín (“*Triticum astivum*”). Entre las de origen anglosajón, las más comunes fueron “lifestyles” (37 ocurrencias), “millenials” (15 ocurrencias).

4 Conclusiones

Los resultados muestran una indización muy exhaustiva, hecho que es generalizado en la indización de las imágenes dado que el ruido que pudiera generarse en la recuperación es asumible habida cuenta de la capacidad humana de seleccionar de un vistazo las imágenes más convenientes.

Predominan las etiquetas conformadas por un único vocablo, es decir los unitérminos que son mayoritariamente sustantivos, si bien la presencia de adjetivos y verbos no es desdeñable como cabía esperar para poner de manifiesto cualidades y acciones observables en las imágenes. Como se ha señalado, la tendencia a asignar este tipo de términos es una práctica particular del tratamiento de las imágenes y resultan claramente necesarios para señalar el comportamiento y la actitud de las personas fotografiadas; otra cuestión sería si la abundancia de secuencias morfológicas observadas facilita o entorpece la búsqueda si esta se realiza de forma textual y no mediante la visualización consecutiva de imágenes.

Destaca el dominio del singular contrariamente al uso habitual en la indización de documentos bibliográficos donde los sustantivos se formalizan en plural si representan realidades contables y solo en singular cuando representan conceptos abstractos o incontables. Creemos que en este caso son los conceptos connotativos los que refuerzan el uso del singular utilizado, asimismo, para identificar realidades referenciales únicas en las imágenes estudiadas. En general, también serían esos términos con carga semántica implícita (no denotativa) los menos redundantes en el conjunto analizado.

La presencia de etiquetas que reflejan aspectos compositivos y técnicos es sorprendentemente reducida; ello denota que *Agefotostock* considera que la selección de las fotografías es preferible realizarla por los aspectos propiamente de contenido y que el enfoque, tipo de plano, estructura formal, formato etc., debe valorarlo el usuario directamente visionando la imagen. Si se señalan, algunos aspectos formales del contenido por medio de etiquetas que describen el entorno o el momento de la toma. Consideramos posible que en otras categorías en las que *Agefotostock* reúne las imágenes la presencia de las etiquetas compositivas cobren una mayor relevancia.

La ausencia de control del vocabulario implica una alta redundancia, por medio del uso de diversas formas para el mismo concepto: variantes en género, número, estructura, utilización de sinónimos e incluso asignación de términos con la misma raíz o de la misma familia para una sola fotografía.

En siguientes trabajos pretendemos aplicar la metodología testada a las demás categorías del banco de imágenes para así poder determinar con mayor precisión las líneas generales de actuación en la indización de *Agefotostock*. Con este fin trataremos de entrar en contacto con la agencia fotográfica para conocer sus procedimientos de indización y saber si se guían por algún instrumento de control o utilizan directrices propias. Resultaría interesante indagar sobre si han recabado datos de satisfacción de sus usuarios a partir de su experiencia en la búsqueda.

Asimismo, nos proponemos extender el análisis a otros proveedores de imágenes similares. Como línea de investigación complementaria, puede ser adecuado llevar a cabo estudios sobre cómo realizan las búsquedas los usuarios de este tipo de bases de datos fotográficas, pues se infiere que el uso final que se hará de los documentos y la propia tipología de estos determina en gran medida su descripción y descubrimiento.

Referencias

- Acal Díaz, I. (2015). Metodologías para el análisis de la imagen fija en los documentos publicitarios: revisión y aplicaciones. *Revista General de Información y Documentación*, 25-2, 425-446.
- Agustín Lacruz, C., Torregrosa Carmona, J.-F. (2019). *Formas de mirar: usos informativos y documentales de la fotografía*. Gijón: Trea.
- Angus, E., Stuart, D. & Thelwall, M. (2010). Flickr's potential as an academic image resource: an exploratory study. *Journal of Librarianship and Information Science*, 42, 4, 268-278.
- Espelt, C., Gascón, J. (1999). La rebelión de las formas: los aspectos formales del contenido como elemento de recuperación. En *4º Congreso ISKO-España*. (pp.321-330). Granada: Universidad de Granada.
- García Jiménez, A. (2007). Tesoros y ontologías para las fotografías. En *Quintas Jornadas Imagen, Cultura y tecnología* (pp.23-32). Madrid: Archiviana.
- Guimarães Pedronette, D. C., Almeida, J., & Torres, R. (2014). A scalable re-ranking method for content-based image retrieval. *Information Sciences*, 265, 91-104.
- Marcos Recio, J. C., Sánchez Vigil, J. M., & Villegas Tovar, R. (2005). La imagen en la publicidad: la fotografía al servicio de la documentación publicitaria y los derechos de autor. *Scire*, 11, .2, 119-132.
- Marzal García-Quismondo, M.; Beltrán Orenes, P., Morato Lara, J., Llorens Morillo, J., Moreiro González, J. A. & Sánchez Cuadrado, S. (2002). Propuesta para la utilización de estructuras verbales aplicadas a la recuperación y representación de la información. *Scire*, 8, 1, en.-jun., 95-102.
- Moreiro González, J. A. (1993). *Aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III/Boletín Oficial del Estado.

- Moreiro González, J. A. (2001). Significantes y significados en el análisis documental de la imagen. En *Catalogación de documentos: teoría y práctica* (pp. 395-418). Madrid: Síntesis.
- Moreiro González, J. A., Bolaños Mejías, C. (2018). Folksonomy indexing from assignment of free tags to setup subject: a search analysis into the domain of legal history. *Knowledge Organization*, 45, 7, 574-585.
- Pinto Molina, M., García Marco, F. J., & Agustín Lacruz, M. del C. (2002). *Indización y resumen de documentos digitales y multimedia. Técnicas y procedimientos*. Gijón: Trea.
- Rodríguez-Bravo, B. (1996). La indización de documentos no bibliográficos. *Accademie e Biblioteche d'Italia*, LXIV, 3, 47-54.
- Rodríguez-Bravo, B. (2000). El tratamiento documental de la imagen fotográfica: su enseñanza en la Universidad de León. En *6es Jornades Antoni Varés. Imatge i recerca*. (pp.205-2011). Gerona: Ayuntamiento de Gerona.
- Rodríguez Roche, S., & Pérez Sanchidrián, E. (2014). Aproximación al estudio de la objetivación de la realidad cubana en la blogosfera nacional a través de los términos de representación de la información. *Anales de Investigación*, 10(10), 20-36.
- Rodríguez Yunta, L. (2009). Etiquetado libre frente a lenguajes documentales. Aportaciones en el ámbito de Biblioteconomía y Documentación. En *IX Congreso ISKO España* (pp.832-845). Valencia: Universidad Politécnica de Valencia.
- Smit, J. W. (1987). A análise da imagen: um primeiro plano. En. *Análise documentaria: a análise da síntese*. 2ª ed..(pp. 99-111). Brasília:IBICT.
- Soler Monreal, C., Gil Leiva, I. (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. *Revista Interamericana de Bibliotecología*, 33(2), 361-377.
- Valle Gastaminza, F. del. (1993). El análisis documental de la fotografía. *Cuadernos de Documentación Multimedia*, 2 (junio), 43-48.
- Valle Gastaminza, F. del. (2002). Indización y representación de los documentos visuales y audiovisuales. En *Manual de Ciencias de la Documentación* (pp. 467-485). Madrid: Pirámide.
- Vélez Cea, M. (2000). *La imagen condicionada*. Granada: Galería Virtual.
- Wu, Y. (2013). Indexing historical, political cartoons for retrieval. *Knowledge Organization*, 40,5, 283-2914.

Análisis de tipos de descripción de fotografía antigua en Wikimedia Commons

Jesús Tramullas¹, Ana I. Sánchez-Casabón² y Piedad Garrido-Picazo³

¹ ORCID [0000-0002-5374-9993](https://orcid.org/0000-0002-5374-9993). Depto. Ciencias de la Documentación, Univ. de Zaragoza, España.

tramullas@unizar.es

² ORCID [0000-0002-0908-1615](https://orcid.org/0000-0002-0908-1615). Depto. Ciencias de la Documentación, Univ. de Zaragoza, España.

asanchez@unizar.es

³ ORCID [0000-0002-1750-7225](https://orcid.org/0000-0002-1750-7225). Depto. Informática e Ingeniería de Sistemas, Univ. de Zaragoza, España.

piedad@unizar.es

Resumen. Este trabajo revisa, usando una aproximación descriptiva, las características de análisis y descripción de los fondos de fotografía antigua disponibles en Wikimedia Commons. A tal fin, se han seleccionado seis fondos, proporcionados por instituciones GLAM. Sobre una selección de imágenes de estos fondos, se revisa si han sido descritos siguiendo los esquemas de metadatos Dublin Core, ISAD(G) o Europeana Data Model. Los resultados obtenidos permiten afirmar que no se aplican estándares de análisis documental ni esquemas de metadatos. El desarrollo de etiquetado semántico en Wikimedia Commons podrá ser una aproximación que ofrezca una solución a estas carencias.

Palabras clave: fotografía antigua, Wikimedia Commons, descripción, GLAM

Abstract. This paper reviews, using a descriptive approach, the characteristics of analysis and description of the ancient photography available on Wikimedia Commons. Six collections have been selected, provided by GLAM institutions. On a selection of 60 images, it is checked if they have been described following the Dublin Core, ISAD (G) or Europeana Data Model metadata schemas. The results obtained allow to affirm that description standards and metadata schemas are not applied. The development of semantic labeling in Wikimedia Commons may be an approach that offers a solution to these deficiencies.

Keywords: ancient photography, Wikimedia Commons, description, GLAM

1 Planteamiento

Wikimedia Commons se ha convertido en un importante repositorio para los fondos de fotografía, y en particular los de fotografía antigua, merced al desarrollo de proyectos de liberación de material digitalizado por parte de instituciones de la memoria (Tramullas y Ojeda, 2018). Estas instituciones participan a través del desarrollo de proyectos GLAM (*Galleries, Libraries, Archives & Museums*) los cuales se asocian a la liberación de materiales textuales, gráficos o sonoros, puestos a disposición de todos los usuarios, bajo licencias libres (Szajewski, 2013). Estas actuaciones se imbrican en el movimiento Open Access, que comenzó siendo una iniciativa de acceso abierto a las publicaciones científicas, pero que ha ido ampliando sus objetivos, hasta incluir elementos y colecciones de patrimonio cultural, que dadas las características del acceso, son resultantes de procesos de digitalización (Phillips, 2013). Además, se ha establecido una relación entre la presencia en Wikipedia y Commons de bibliotecas, museos y archivos, y de sus colecciones, y el aumento de consumo cultural, lo que resulta de gran importancia para facilitar el acceso, conocimiento y disfrute del patrimonio cultural (Navarrete y Borowiecki, 2016).

Wikimedia Commons sigue un proceso de aportación y edición colaborativa en abierto, al igual que Wikipedia. En el proceso colaborativo pueden participar tanto usuarios interesados, sin formación específica, como profesionales de la gestión de documentos. Este entorno abierto contrasta con los procesos de tratamiento de información que se llevan a cabo en las instituciones de la memoria, llevados a cabo por profesionales, y en el marco de un contexto normativo bastante estricto. En consecuencia, esta dicotomía lleva a plantear diferentes cuestiones sobre el nivel de detalle descriptivo usado en Commons, frente al aplicado en los catálogos especializados de las instituciones de la memoria. En primer lugar, hay que revisar el proceso de carga de imágenes y el esquema de descripción de contenido que se aplica, y sus características. En segundo lugar, la interacción que se lleva a cabo sobre la descripción, en especial los procesos de corrección o mejora de la misma.

Ambas cuestiones se trasladan a dos preguntas de investigación. La primera pregunta que cabe plantear es si estas prácticas se trasladan a Wikimedia Commons, o si se adoptan otros modelos descriptivos, influenciados por o adaptados al medio que los acoge. La segunda pregunta que se plantea es si las descripciones realizadas originalmente son aceptadas o son modificadas por la comunidad de editores de Commons, lo que abre un campo de investigación de interés sobre los procesos colaborativos de descripción y clasificación de documentos de imagen.

2 Metodología

La investigación sobre diferentes aspectos de Wikipedia y el resto de proyectos Wikimedia ha crecido notablemente en la última década (Tramullas, 2015). Este trabajo toma como punto de partida la propuesta clásica de atender a aspectos de

análisis formal. Dadas las características de los fondos de fotografía antigua disponibles en Commons, se ha optado por un análisis descriptivo de las características que presentan. En consecuencia, se revisará si los datos disponibles para cada fotografía responden a la información mínima requerida para una adecuada descripción del objeto. Aunque se ha señalado la dificultad que supone el abordar los procesos de análisis documental del objeto fotográfico, Sánchez-Vigil (2006: 180-181) propone la utilización de ISAD(G) para la descripción de documentos fotográficos. Iglésias-Franch (2008: 81-100) llevó a cabo una revisión de los diferentes estándares de metadatos para descripción y gestión existentes para fotografía digital, recomendando el uso de ISAD(G) etiquetado en EAD, aunque llamaba la atención sobre las limitaciones de esta aproximación.

Para el análisis de las estructuras de descripción y etiquetado de los fondos fotográficos se ha seguido la siguiente secuencia:

1. Identificación y revisión del protocolo de aportación de material fotográfico a Wikimedia Commons.
2. Identificación de los esquemas de descripción y categorización de material fotográfico.
3. Selección de fondos a revisar. A tal fin, se han identificado fondos específicos e imágenes aportados por seis instituciones de la memoria, en el marco de proyectos GLAM institucionales.
4. Revisión de los metadatos EXIF identificados.
5. Análisis de los esquemas de descripción utilizados. Comparación con las propuestas de Dublin Core (DC), ISAD(G) y Europeana Data Model (EDM).
6. Análisis de los principios y esquemas de categorización utilizados para la descripción de las imágenes.
7. Revisión del historial de cambios de las imágenes en Wikimedia Commons.

3 Análisis de fondos fotográficos

El análisis que se expone en los siguientes párrafos ha sido realizado durante el mes de mayo de 2019. En entornos cuyos servicios, funcionalidades y contenidos evolucionan a lo largo del tiempo, la revisión de los resultados obtenidos puede ofrecer variaciones dependiendo de diferentes factores, como el período transcurrido o la incorporación o modificación de funcionalidades técnicas.

El proceso de carga y la descripción inicial de material en Wikimedia Commons ha sido detallado por Saorín y Pastor (2011) y por Tramullas y Ojeda (2018). El proceso básico no se presta a realizar una descripción detallada: basta con dar un nombre para el fichero del objeto, una descripción textual no estructurada, y asignar una o varias categorías, tomadas de la estructuras preexistentes, o de nueva creación. Commons puede recomendar categorías por similitud con el texto introducido por el

editor a tal fin. Las herramientas que permiten el proceso de carga por lotes tampoco proveen la capacidad de incorporar metadatos estructurados.

Como consecuencia de las características del proceso de carga, no se han identificado esquemas de metadatos preparados previamente para la descripción de las imágenes. En aquellos casos en que las imágenes incorporan elementos específicos, éstos han sido incorporados a las mismas en un proceso de edición y modificación de contenidos posterior. En varios casos estos elementos se han organizado utilizando plantillas creadas específicamente a tal fin por las instituciones responsables del proyecto. El uso de las categorías como esquema para clasificar las imágenes muestra una tendencia clara a incorporar descriptores que intentan desempeñar funciones de identificación del fondo de origen, así como de identificación de algunos elementos básicos, como la autoría o la cronología, pero sin que esto sea consistente en todas las fotografías y fondos revisados. El sistema de categorías de Wikipedia ha sido ampliamente utilizado en diferentes ámbitos de investigación (Tramullas, Sánchez-Casabón y Garrido-Picazo, 2018).

Para el análisis de los esquemas de descripción utilizados, se han seleccionado fotografías antiguas provenientes de seis proyectos GLAM llevados a cabo por instituciones, de manera oficial, en Wikimedia Commons. Los seis proyectos se encuentran recogidos en la tabla 1.

Tabla 1. Proyectos GLAM objeto de revisión.

Institución	URL de referencia
<i>Library of Congress</i>	https://commons.wikimedia.org/wiki/Category:Collections_in_the_Library_of_Congress
<i>Deutsches Fotothek</i>	https://commons.wikimedia.org/wiki/Category:Images_from_the_Deutsche_Fotothek
<i>Metropolitan Museum of Art</i>	https://commons.wikimedia.org/wiki/Category:Images_from_Metropolitan_Museum_of_Art
<i>Georges Eastman House Collection</i>	https://commons.wikimedia.org/wiki/Category:Photographs_in_the_George_Eastman_House_Collection
<i>The J. Paul Getty Museum</i>	https://commons.wikimedia.org/wiki/Category:Google_Art_Project_works_in_The_J._Paul_Getty_Museum
<i>Fundación Joaquín Díaz</i>	https://commons.wikimedia.org/wiki/Category:Images_provided_by_Fundaci%C3%B3n_Joaqu%C3%ADn_D%C3%ADaz

Para cada uno de los seis proyectos se han buscado y seleccionado diez fotografías. En cada una de ellas se han comprobado los metadatos EXIF existentes, y la descripción, la categorización, y el uso de esquemas de metadatos para la descripción del objeto fotográfico. Los resultados obtenidos se resumen en la tabla 2.

Tabla 2. Esquemas de metadatos identificados.

Institución	EXIF	Metadatos	DC	ISAD(G)	EDM
Library of Congress	Parcial	Si	No	No	No
Deustches Fotothek	Parcial	Si	No	No	No
Metropolitan Museum of Art	Parcial	Si	No	No	No
Georges Eastman House Collection	Parcial	Si	No	No	No
The J. Paul Getty Museum	Parcial	Si	No	No	No
Fundación Joaquín Díaz	Parcial	Si	No	No	No

La aparición y formulación de los metadatos EXIF no es consistente. Dado que las imágenes revisadas provienen de diferentes políticas y procesos de digitalización, la presencia de los mismos es parcial, debido a las diferentes cabeceras generadas por las herramientas informáticas usadas en el proceso. Incluso en ocasiones se carece por completo de estos metadatos, que no han sido considerados relevantes por la institución generadora.

Artista

Desconocido

Título

The late Thomas Jones (Taliessin o Eifion)

Object type

fotografía

Género

retrato

Descripción

English: A portrait from the Welsh Portrait Collection at the National Library of Wales. Persona retratada: Thomas Jones - Welsh poet (1820 - 1876)

Fecha

hacia 1868

Técnica

fotografía

Colección

Biblioteca Nacional de Gales

Número de inventario

PA3316.

Fuente/fotógrafo

LGC NLW

LLYFRGELL GENEOLAETHOL CYMRU

THE NATIONAL LIBRARY OF WALES

This image is available from the [National Library of Wales](#)

You can view this image in its original context on the [NLW Catalogue](#)



Figura 1. Ejemplo de descripción básica de fotografía antigua.

La columna etiquetada “Metadatos” corresponde a aquellas fotografías en las cuales se han añadido metadatos básicos descriptivos, del tipo autoría, cronología, título, ubicación, técnica... (véanse figuras 1 y 2). Sin embargo, y aunque puede establecerse una correspondencia con algunos de los elementos de otros esquemas de metadatos, como puede ser, por ejemplo, Dublin Core, realmente no se ha identificado en ninguna de las fotografías revisadas el uso ni de Dublin Core, ni de ISAD(G), ni de EDM.

Rue de l'Hotel de Ville

Artista	Eugène Atget (1857-1927)
Título	Rue de l'Hotel de Ville
Descripción	English: Eugène Atget's penchant for photographing in the early morning accounts for the street's empty stillness. A sole inhabitant peers out from the hotel doorway, so unobtrusive he becomes a part of the architecture as he quietly watches the photographer work. Atget's decision to photograph when the city was deserted heightens the sense of desolate abandonment in the city's older sections, soon to be demolished to make way for grand, broad, new boulevards. The two dark, solemn buildings in the foreground thus frame the old, narrow lane with the solid weight of history. An hourglass-shaped shaft of bright daylight cuts through the frame, dividing the street into shadow and illumination.
Fecha	1921
Técnica	Albumen silver print
Dimensiones	Image: 21.6 x 18.3 cm (8 1/2 x 7 3/16 in.)
Colección	Getty Center
Número de inventario	84.XM.1034.7
Inscripciones	Title and negative number 6267 inscribed verso print in pencil.
Fuente/fotógrafo	The Getty Center, Object 62231



Esta imagen proviene del programa Open Content Program® (programa de contenido abierto) del Instituto de Getty Research, lo cual ha dicho que después de su evaluación, **no hay conocidas restricciones del derecho de autor**.
Imágenes de Open Content son reproducciones digitales de obras de arte que están en las colecciones de Getty y en el dominio público. Tenemos los derechos de autor de estas imágenes, o no somos conscientes de cualquier restricción del uso de ellas.
Mientras el Instituto de Getty Research no puede comentar con certitud al estado de derechos de una imagen en particular, ha dicho que se pueden utilizar las imágenes de Open Content por cualquier propósito sin la necesidad de permiso previo de Getty.
Para más información, véase <http://www.getty.edu/about/opencontent.html>.

Figura 2. Ejemplo de descripción básica de fotografía antigua.

Cabe destacar que en la interfaz de uso de los documentos se muestra una pestaña titulada “Datos estructurados”, y cuyo contenido deberá ser provisto a través de la introducción de descripciones estructuradas de datos a través del proyecto *Commons:Structured data*. Sin embargo, este proyecto todavía se encuentra en desarrollo, y por el momento ofrece como esquema de metadatos una versión del provisto por CIDOC para objetos museísticos. La tabla de propiedades del proyecto tienen previsto ofrecer la descripción del objeto y de su contenido a través de la propiedad P180, “representa a”, y de los calificadores correspondientes. En el conjunto estudiado no se ha encontrado ninguna imagen que implemente esta propiedad. La utilización del sistema de categorías muestra una tendencia a intentar su uso como reflejo del esquema de organización de propio fondo, estableciendo la proveniencia y las series de origen (véase fig. 3). Sin embargo, y al igual que en otros aspectos, la aproximación usada tampoco es homogénea entre instituciones, con diferencias que pueden apreciarse en el uso de categorías como sustitutas de metadatos.

Categorías (+): 1921 photographs of Paris (–) (±) (↓) (↑) | 1921 photographs (–) (±) (↓) (↑) | Photographs of Paris by Eugène Atget (–) (±) (↓) (↑) | Photographs in The J. Paul Getty Museum (–) (±) (↓) (↑) | Rue de l'Hôtel-de-Ville (Paris) (–) (±) (↓) (↑) | Images of Paris from the Getty Museum (–) (±) (↓) (↑) (↑) (+)
Categorías ocultas: Files from the Getty's Open Content Program | Artworks with known accession number | Artworks without Wikidata item | CC-PD-Mark | PD-old-80

Figura 3. Ejemplo de uso de categorías como estructura del fondo.

Finalmente, la revisión de los historiales de edición recoge el ciclo de vida de los procesos de carga y modificación de las imágenes. La primera versión de las imágenes digitalizadas es la que prevalece, habiendo detectado sólo un cambio de fichero entre los 60 analizados. El número medio de ediciones que modifican la aportación y datos originales es de 7, la mayoría correspondientes a correcciones automáticas realizadas por un bot, o a correcciones de asignación de categorías, mientras que aquellas que añaden o modifican datos metadatos se reducen a un 18% del total.

4 Conclusiones

La descripción de los elementos u objetos fotográficos disponibles en Wikimedia Commons es irregular y carece de homogeneidad, ya que no existe un estándar predefinido que haya sido establecido acorde a los requerimientos establecidos para el tratamiento documental y archivístico de este tipo de materiales. La información aportada por los metadatos EXIF es mínima, ya que en aquellos casos en los que aparece sólo informa sobre la herramienta de digitalización utilizada. En lo que concierne a los aspectos documentales y archivísticos, no se han identificado, en los materiales seleccionados, esquemas de descripción que sigan lo establecido por DC, ISAD(G) o EDM. Tampoco ha podido establecerse la aplicación de esquemas uniformes de categorización o clasificación. El análisis de los diferentes grupos y elementos individuales revisados demuestra que diferentes instituciones han utilizado diferentes aproximaciones para la descripción y categorización de los fondos fotográficos.

La revisión de los historiales de actividad sobre las imágenes revela que en muy pocos casos son modificadas las descripciones y clasificaciones llevadas a cabo en el momento de la aportación a Commons, lo que induce a pensar que la comunidad de editores da por válida la información provista por las instituciones.

El desarrollo y puesta en marcha del proyecto *Commons:Structured Data*, que pretende ofrecer un marco semántico de descripción de objetos culturales utilizando la base de Wikidata (Vrandečić y Krötzsch, 2014)) se orienta a resolver las carencias de descripción existentes, y a facilitar la reutilización y enlazado del material disponible en otros recursos de información en productos de información abiertos (de Boer et alii, 2019).

Referencias

- de Boer, V., de Bruyn, T., Brooks, J., & de Vos, J. (2019). The Benefits of Linking Metadata for Internal and External Users of an Audiovisual Archive. En *MTSR 2018: Metadata and Semantic Research* (pp. 212–223). doi: 10.1007/978-3-030-14401-2_20
- Iglésias Franch, D. (2008). *La fotografía digital en los archivos*. Gijón: Trea.

- Lally, A.M. & Dunford, C. E. (2007). Using Wikipedia to Extend Digital Collections. *D-Lib Magazine*, 13(5/6). Recuperado de: <http://www.dlib.org/dlib/may07/lally/05lally.html>
- Navarrete, T. y Borowiecki, K.J. (2016). Changes in cultural consumption: ethnographic collections in Wikipedia. *Cultural Trends*, 25(4), pp. 233-248. doi: 10.1080/09548963.2016.1241342
- Phillips, L. B. (2013). The Temple and the Bazaar: Wikipedia as a Platform for Open Authority in Museums. *Curator: The Museum Journal*, 56(2): 219-235. doi:10.1111/cura.12021
- Robledano Arillo, J. & Canela Garayoa, M. (2007). Estándares para la descripción de fotografía. *Revista d'arxius*, 6, pp. 149-188
- Sánchez Vigil, J.M. (2006). *El documento fotográfico. Historia, usos, aplicaciones*. Gijón, Trea.
- Sánchez Vigil, J. M. & Salvador Benítez, A. (2014). *Documentación fotográfica*. Barcelona, UOC.
- Saorín, T. & Pastor-Sánchez, J.A. (2011). Bancos de imágenes para proyectos enciclopédicos: el caso de Wikimedia Commons. *El Profesional de la Información*, 20(4), 424-431. doi:10.3145/epi.2011.jul.09
- Szajewski, M. (2013). Using Wikipedia to Enhance the Visibility of Digitized Archival Assets. *D-Lib Magazine*, 19(3/4). Recuperado de: <http://www.dlib.org/dlib/march13/szajewski/03szajewski.html>
- Tramullas, J. (2015). Wikipedia como objeto de investigación. *Anuario ThinkEPI*, 9, 223-226. doi:10.3145/thinkepi.2015.50
- Tramullas, J. & Ojeda, R. (2018). Fondos documentales para el estudio de la historia de la fotografía en Wikimedia Commons. En: Hernández-Latas, J.A. (ed.) *II Jornadas sobre Investigación en Historia de la Fotografía* (pp. 407-418). Zaragoza: Institución Fernando el Católico.
- Tramullas, J., Sánchez-Casabón, A. I., & Garrido-Picazo, P. (2018). Wikipedia categories in research: towards a qualitative review of uses and applications. En: *Challenges and Opportunities for Knowledge Organization in the Digital Age. Proceedings of the Fifteenth International ISKO Conference*. (pp. 490-497). Porto: Ergon Verlag.
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85. doi: 10.1145/2629489

La organización social de fotografías personales en *Flickr*: una mirada desde la gestión de información personal

Jorge Franganillo

ORCID [0000-0003-4128-6546](https://orcid.org/0000-0003-4128-6546). Departamento de Biblioteconomía, Documentación y Comunicación Audiovisual,
Universidad de Barcelona, España.
franganillo@ub.edu

Resumen. La aparición de plataformas sociales para almacenar y compartir imágenes ha traído consigo nuevas formas de organizar y difundir las colecciones de fotografías personales. En estos espacios en línea, las personas pueden compartir sus imágenes con sus contactos o con el público general, y entonces pueden usar el sitio como un escaparate digital de sus creaciones. Con lo cual, estas plataformas parecen tener un gran potencial como archivo personal duradero. *Flickr*, uno de los ejemplos más tempranos de informática social, es el depósito de fotografías más orientado a la creación de comunidad. Los usuarios de esta plataforma, en lugar de agrupar las imágenes según su significado personal, piensan ahora en un contexto más amplio y se preocupan por poner sus fotografías a disposición de toda la comunidad: las titulan, las describen y las agregan a grupos temáticos para hacerlas más fáciles de encontrar. A partir de una muestra de 200 imágenes, que corresponden a otros tantos usuarios individuales, este estudio a pequeña escala expone los esfuerzos que las personas dedican a facilitar la recuperación y el descubrimiento de las fotografías personales que publican en *Flickr*.

Palabras clave: gestión de información personal; contenido generado por el usuario; fotografía digital; informática social; etiquetado.

Abstract. The emergence of social platforms for image storing and sharing has brought new ways to organize and disseminate personal photograph collections. In these online spaces, people can share their images with their contacts or with the public, and they can use the site as an online showcase of their creations. Therefore, these platforms seem to have a great potential as a long-term personal archive. *Flickr*, one of the earliest examples of social computing, is the most community-oriented photo repository. Users of this platform don't just group their images according to their personal meaning; they now think in a broader context and are concerned about making their photographs available to the whole community: they title them, describe them and include them in pools to make them easier to find. From a sample of 200 images, corresponding to as

many individual users, this small-scale study explains the efforts made by users to facilitate the retrieval and discovery of the personal photographs they post on *Flickr*.

Keywords: personal information management; user-generated content; digital photography; social computing; tagging.

1 Introducción

Las fotografías personales, ya sean digitales o en soporte impreso, suelen constituir una de las colecciones de información personal más importantes que un individuo atesora a lo largo de su vida. Tanto si surgen de la afición personal por la fotografía como si se generan con el propósito, más o menos consciente, de documentar de forma gráfica ciertos objetos, lugares o acontecimientos, estos artefactos representan momentos y aspectos significativos de la vida de una persona. Al lograr encapsular estas pequeñas porciones de la realidad que rodea al individuo, estas piezas gráficas de información personal reflejan su propia historia y constituyen, en definitiva, una recopilación única de recuerdos y vivencias.

Es por ello que las personas dedican esfuerzos para almacenar y organizar sus colecciones fotográficas a fin de verlas de nuevo más tarde y así revivir aquellos momentos, con la consiguiente recompensa emocional; o bien para compartirlas con otras personas. A diferencia de otras piezas de información personal como, p. ej., los mensajes electrónicos o las páginas web, las fotografías personales tienden a ser de creación propia y no suelen ser estrictamente informativas ni suelen procesarse como una tarea. Forman parte del espacio personal de información porque suelen generar respuestas afectivas intensas (Frohlich *et al.*, 2002; Whittaker, 2011).

Por su singularidad, las fotografías creadas en el transcurso de la vida cotidiana son un recurso muy valorado por quien las produce, que las considera insustituibles (Petrelli, Whittaker y Brockmeier, 2008; Whittaker, Bergman y Clough, 2010). Requieren, en consecuencia, medidas oportunas de preservación. Y por su naturaleza, forman una colección que tiende a crecer con el paso del tiempo y, por lo tanto, tiende a aumentar también la dificultad para gestionarlas.

La llegada de la fotografía digital, primero, y el surgimiento de los teléfonos móviles con cámara, después, han dado lugar a una explosión en el número de fotografías que se toman y se guardan (Okabe, 2004; Whittaker, Bergman y Clough, 2010; Whittaker, 2011; Bergman y Whittaker, 2016) y han transformado tanto el contenido de las imágenes como los métodos con que estas se capturan, se organizan y se comparten (Angus y Thelwall, 2010; Stuart, 2013). En este escenario de sobreabundancia y variedad, la conveniencia de facilitar la recuperación futura plantea retos en cuanto a la adecuada organización de las fotografías personales. En este contexto es donde se desarrolla este estudio sobre la organización social de

fotografías personales que busca mejorar la posterior recuperación y el posterior descubrimiento por parte de otros usuarios.

1.1 La gestión personal de colecciones fotográficas

Las fotografías personales constituyen uno de los principales subconjuntos del espacio personal de información. Forman una colección de información personal en la cual el propietario hace esfuerzos conscientes para controlar qué piezas entran y cómo las organiza. A causa de la relación afectiva que las personas establecen con sus fotografías personales, suelen cuidarse de preservarlas durante muchos años (Bergman y Whittaker, 2016). Al ser un registro de las propias experiencias, tienen varios usos potenciales; entre otros: sustentar la visión de un hecho pasado, refrescar la memoria sobre un acontecimiento, o revivir instantes o circunstancias por simple diversión (Jones, 2007).

Sin embargo, paradójicamente, en el paradigma digital es laxo el empeño que las personas ponen en decidir qué guardan y qué no. Al percibir que el coste de conservarlo todo es casi insignificante, tienden entonces a guardar en exceso y dedican luego pocos esfuerzos a hacer limpieza. El borrado, al fin y al cabo, es un proceso difícil, según evidencia el hecho de que muchas de las fotos guardadas son semiduplicados, es decir, numerosas fotos similares de una misma escena (Kirk *et al.*, 2006).

Estudios recientes muestran, además, que en el ámbito privado las personas suelen organizar sus fotografías mediante estrategias rudimentarias y tienden a amontonar las piezas en colecciones con una estructura jerárquica bastante plana, sin apenas subcarpetas (Kirk *et al.*, 2006; Whittaker, Bergman y Clough, 2010). Igual que sucede con los documentos de papel y con el correo electrónico, las personas tienen grandes expectativas pensando que más tarde regresarán a su colección de fotos para organizarla y ponerle orden, pero está comprobado que esa racionalización rara vez se produce (Bergman y Whittaker, 2016). La falta de organización obliga a confiar excesivamente en la memoria autobiográfica y en estrategias de ensayo y error, y reduce en consecuencia el éxito en el reencuentro de piezas concretas (Kirk *et al.*, 2006).

La gestión de fotografías digitales también se ha visto influida por la convergencia de medios, cosa que actualmente implica el uso de varios dispositivos y procesos en la creación, la gestión y el almacenamiento de imágenes. La abundancia de fotografías digitales y su fragmentación en diversos dispositivos y plataformas en línea representa un obstáculo importante para el mantenimiento a largo plazo. En efecto, disuade a los creadores de ejecutar actividades de gestión y, como resultado, estos adoptan soluciones pasivas de preservación, como, p. ej., la mera acumulación (Van House, 2011; Bushey, 2015).

La necesidad de herramientas de gestión y almacenamiento conduce a las personas a depender cada vez más de aplicaciones para compartir imágenes. En ellas, sin embargo, los esfuerzos dedicados a la organización tienden a ser mayores porque a los usuarios los impulsa, sobre todo, la motivación altruista de compartir con el público general las propias creaciones, dada su utilidad potencial para otros usuarios (p. ej., como fondo de pantalla o como ilustración para un artículo). Como se comprobará más adelante, en el contexto de la informática social las tareas de organización ya no se encaminan únicamente a facilitar la propia explotación personal, sino también a aumentar la capacidad de descubrimiento en comunidades virtuales por parte de otros usuarios.

1.2 La publicación y organización social de fotografías personales: el caso de *Flickr*

La aparición de plataformas sociales para almacenar, gestionar y compartir imágenes (p. ej., *Flickr*, *SmugMug*, *500px*, *Photoshelter*, etc.) ha traído consigo nuevas formas de organizar y difundir las colecciones de fotografías personales. En estos espacios en línea, las personas pueden optar por mantener en privado las imágenes que depositan y así tratar el sitio como un archivo personal, o bien pueden compartirlas con sus contactos o con el público general y entonces pueden usar el sitio como un escaparate digital de sus creaciones (Stuart, 2019). En cualquier caso, estas plataformas parecen tener un gran potencial como archivo personal duradero (Zhao y Lindley, 2014).

Flickr, lanzado en 2004, es uno de los ejemplos más tempranos de este tipo de sistemas (Cox, Clough y Marlow, 2008) y, como tal, ha atraído más interés científico que ningún otro sitio de su especie (Stuart, 2013). Cuenta con más de 90 millones de usuarios con actividad de al menos una vez al mes (febrero de 2019) y aloja una cantidad multimillonaria de imágenes (más de 10 mil millones en octubre de 2015, último dato oficial). Permite generar álbumes, exposiciones y grupos (*pools*) sobre cualquier concepto, e incluso lanzar concursos o debates sobre cualquier tema o práctica fotográfica.

Es el depósito de fotografías más orientado a la creación de comunidad (Remick, 2010) y ello motiva a muchos usuarios a compartir sus creaciones y, especialmente, a describir con cierto detalle el contenido de cada una de ellas (Russo y Nov, 2010). El incentivo social es tan importante (Marlow, Naaman, Boyd y Davis, 2006) que los usuarios han cambiado la concepción que tenían sobre sus fotografías. En lugar de agrupar las imágenes según el significado personal, los usuarios piensan ahora en un contexto más amplio y se preocupan por poner sus fotografías a disposición de toda la comunidad: las titulan y las describen para hacerlas más fáciles de encontrar (Stuart, 2013).

En este sentido, una característica clave de la mayoría de las plataformas de alojamiento de fotografías es la posibilidad de etiquetar el contenido. El etiquetado

consiste en asignar palabras clave, que representan conceptos a los que la imagen alude de algún modo, con el objetivo de describir y organizar la colección para mejorar la recuperación y el descubrimiento de cada una de las fotos (Xu, Fu, Mao y Su, 2006). Como las etiquetas se eligen libremente y no siguen convención alguna, son de una tipología tan diversa como sus propósitos (Gupta *et al.*, 2011; Rafferty, 2018); en las plataformas de fotografía suelen representar, entre otros aspectos:

- objetos, lugares o personas que aparecen en la imagen;
- temas implícitos en la fotografía;
- el nombre, la fecha o la ubicación de un acontecimiento;
- aspectos técnicos: marca y modelo de la cámara usada, tipo de objetivo, tiempo de exposición, etc.;
- estilos y géneros fotográficos;
- elementos connotativos.

Flickr ofrece la posibilidad de que otros usuarios del sistema añadan etiquetas a fotografías públicas, aunque no es esta una práctica generalizada, ya que los usuarios la perciben como una intrusión (Marlow *et al.*, 2006; Cox, Clough y Marlow, 2008; Ding *et al.*, 2009), en parte por el carácter personal que para cada autor tienen sus propias imágenes (Stuart, 2019).

También cuenta, desde 2015, con un mecanismo de etiquetado automático, basado en la tecnología de reconocimiento de imagen, que añade, diferenciadas en otro color, etiquetas no contempladas inicialmente por el usuario (Flickr, 2018). Estas etiquetas no solo aspiran a mejorar la categorización y a facilitar la recuperación mediante palabra clave, sino también a favorecer el descubrimiento; al presentarse como enlaces hipertextuales, facilitan el acceso a piezas similares o relacionadas. Sin embargo, esta iniciativa fue objeto de controversia en el momento de su implantación a causa de llamativos fallos detectados por la comunidad usuaria (Hern, 2015). Se trataba de fallos que pueden considerarse normales en la fase embrionaria de cualquier desarrollo basado en la inteligencia artificial. Con el tiempo, el algoritmo ha aprendido de los errores y el mecanismo ha mejorado su desempeño. Por otro lado, tras la adquisición de *Flickr* por parte de SmugMug, en abril de 2018, y el consiguiente traslado del servicio a AWS, esta característica ha quedado interrumpida (en principio, de manera temporal) y, en consecuencia, el presente estudio solo puede ofrecer resultados parciales, es decir, condicionados por el inconveniente de esta interrupción.

1.3 Estudios previos

Las fotografías personales son un recurso que conviene custodiar adecuadamente durante toda la vida. Con esta convicción, Marshall (2007) reconoce en *Flickr* un sistema particularmente apto para almacenarlas, ya que, por una parte, facilita el archivo y contribuye a resolver así el problema del guardado, fundamental en la

gestión de información personal. Y, por otra parte, al ser una herramienta de almacenamiento distribuido, promueve, además, el uso compartido. Heckner, Heilemann y Wolff (2009), por su lado, muestran que los usuarios de *Flickr* utilizan la plataforma con la doble intención de gestionar y compartir sus fotografías personales.

Aquí debe tenerse presente que el acceso eficaz a las imágenes requiere medidas preparatorias, una de las cuales consiste en describir el contenido por medio de etiquetas. El etiquetado en sistemas como *Flickr* se lleva a cabo, según sugiere un amplio conjunto de investigaciones (Van House *et al.*, 2004; Van House, 2007; Nov, Naaman y Ye, 2009a y 2009b; Ames *et al.*, 2010; Angus y Thelwall, 2010), por una razón de entre cuatro posibles, o por una combinación de ellas:

- organización personal, como ayuda a la recuperación futura;
- organización social, para ayudar a otros usuarios a encontrar imágenes;
- autocomunicación, como apoyo a la propia memoria, a modo de diario personal;
- comunicación social, para expresar emociones u opiniones, o para atraer la atención de otros usuarios.

Van House *et al.* (2005), asimismo, identifican en el uso de fotografías personales un conjunto de propósitos sociales:

- creación y mantenimiento de relaciones sociales;
- memoria personal y grupal;
- expresión personal;
- presentación personal;
- funcional (usos recordatorios para uno mismo y para los demás).

En efecto, son varias las motivaciones que impulsan a un individuo a publicar sus fotografías en servicios para la compartición de ficheros. La motivación de cada usuario para utilizar el sistema determina en buena medida el uso de etiquetas (Marlow *et al.*, 2006; Kennedy *et al.*, 2007). Si alguien publica sus imágenes en un sitio como *Flickr* para que otras personas las encuentren y las vean (organización y comunicación social), es muy probable que invierta tiempo y esfuerzos en etiquetar sus imágenes. En cambio, si usa el sitio como copia de seguridad en línea (organización personal), es más probable que organice las fotografías en colecciones o álbumes, las titule y las describa para narrar de algún modo la imagen, pero seguramente no se moleste tanto en etiquetarlas (Stuart, 2013).

Pese a la popularidad de las plataformas para la publicación de imágenes, no es muy abundante la bibliografía que estudia la organización social de fotografías personales en estos espacios. Algunos estudios han analizado las motivaciones de los usuarios para publicar y etiquetar sus fotografías (Cox, Clough y Marlow, 2008; Angus y Thelwall, 2010; Russo y Nov, 2010; Stuart, 2019), ciertas actividades preparatorias que no incluyen la representación del contenido (Kirk *et al.*, 2006) y el etiquetado como principal estrategia facilitadora de la recuperación (Heckner,

Neubauer y Wolff, 2008; Heckner, Heilemann y Wolff, 2009; Stuart, 2013). Sin embargo, ninguno parece haber examinado de forma conjunta, todavía, la práctica de titular, describir, etiquetar y geolocalizar las fotografías personales que ponen a disposición pública ni se ha centrado en analizar cómo aportan los usuarios esos metadatos en el contexto multilingüe de la comunidad construida en torno a estas plataformas.

Desde el punto de vista de la gestión de información personal, el etiquetado es una estrategia unificadora del acceso (Karger, 2007). Al asignar etiquetas a cada una de las piezas de una colección se consigue agrupar aquellas entidades que comparten algún rasgo común. Basta, pues, con asignar una misma etiqueta (o varias) a las piezas que forman parte de un grupo.

Ampliar el conocimiento sobre la conducta organizadora de los usuarios, que en este contexto son también autores, puede contribuir a una mayor comprensión de los criterios con que las personas representan el contenido de sus imágenes y puede dar pistas sobre sus conductas de búsqueda; según cómo etiqueten sus propias fotografías podemos intuir cómo buscan estos contenidos de otros autores. Con esa convicción, y mediante el análisis de una muestra significativa de fotografías, este estudio quiere contribuir, siquiera sea a pequeña escala, a profundizar en ese conocimiento.

2 Objetivos

En virtud de lo expuesto en la Introducción, el presente estudio se centra en responder a una pregunta genérica, planteada en el contexto de la organización social en *Flickr*. Se trata de: ¿Cómo representan los usuarios el contenido de sus fotografías personales para facilitar la recuperación y el descubrimiento por parte de otros usuarios?

Esta cuestión se puede desglosar en cuatro preguntas de investigación concretas sobre los metadatos descriptivos de las imágenes (título, descripción, etiquetas, ubicación geográfica):

- ¿Qué metadatos se rellenan de manera más exhaustiva?
- ¿Qué aspectos representan las diversas etiquetas de cada fotografía?
- ¿En qué idiomas se representa el contenido de las fotografías?
- ¿Qué porcentaje de fotografías está georreferenciado?

Las respuestas a estas preguntas permitirán explorar, además, en qué aspectos los esfuerzos de gestión desempeñados en esta plataforma se asemejan o contrastan con los realizados en el ámbito personal.

3 Método

Para responder a las preguntas de investigación planteadas, se ha tomado una muestra de estudio constituida por las últimas 200 fotografías compartidas en el grupo *Your Best Shot 2018* <<https://flickr.com/groups/1621376@N22>>. La elección de este grupo

la motiva la diversidad temática del material y el hecho de que la participación está limitada a una única imagen por usuario, lo cual hace más representativa la muestra.

De cada pieza de este conjunto se han analizado, por una parte, el título, la descripción y las etiquetas para determinar si los usuarios han aportado estos metadatos y, en caso afirmativo, con qué tipo de información y con qué grado de especificidad. Y, por otra parte, se ha comprobado la presencia de etiquetas automáticas, metadatos técnicos (EXIF) y geográficos, y la pertenencia a álbumes y a grupos temáticos.

Se ha evaluado si el título, la descripción y las palabras clave aportadas por el usuario representan de forma fehaciente y detallada el contenido de la imagen. Se ha registrado, también, qué aspectos de la imagen representan esos elementos: denotativos, connotativos, estilísticos, etc. Las etiquetas aportadas automáticamente por el sistema se han incluido en la fase inicial del análisis para poder analizar hasta qué punto se complementan mutuamente o se diferencian entre sí.

La recogida de datos se ha llevado a cabo entre abril y junio de 2019, y coincide así cronológicamente, al menos en parte, con el tiempo durante el cual la plataforma *Flickr* ha trasladado el servicio a AWS y algunas características han quedado interrumpidas.

4 Resultados

Los resultados preliminares de este análisis muestran que, con independencia de la procedencia del usuario, el contenido se representa mayoritariamente en inglés. Esta tendencia puede entenderse, al menos en parte, como un esfuerzo colectivo por unificar la recuperación por palabra clave en el idioma establecido como la *lingua franca* de Internet. De las fotografías que incluyen una descripción, el 88,7% la tienen redactada en inglés. Y de las tituladas, una mayoría (difícil de cuantificar, dada la naturaleza de títulos como, p. ej., los que representan topónimos o antropónimos intraducibles) también se presenta en inglés. Esta proporción está alineada con los hallazgos de Koochali *et al.* (2016) sobre una muestra mayor, en la que el inglés también se presenta como el idioma predominante, usado en el 75,4% de las imágenes. No en vano, desde el punto de vista de la lingüística aplicada, la universalidad del inglés facilita la comprensión mutua y la práctica de principios cooperativos (Pikhart, 2018).

Lleva título el 96% de las imágenes de la muestra analizada. En ese subconjunto, las fotografías tienden a identificarse con títulos más bien vagos (39,1%), si bien un porcentaje significativo de usuarios asigna títulos descriptivos ricos en contenido (32,3%) o meramente simbólicos o alegóricos, con una notable carga subjetiva (20,8%). El 5,7% deja como título el nombre por defecto del fichero (p. ej., *DSC_6706*) o utiliza alguna nomenclatura de naturaleza archivística (p. ej.,

PC120125, cuyas dos iniciales indican que se trata de una fotografía en color), y el 2,1% restante otorga títulos con algún significado personal.

En el 58% de los casos, la fotografía lleva una descripción de sí misma. Se trata de descripciones de naturaleza heterogénea y en ocasiones incluyen más de un elemento. En su mayoría, son una referencia a la ubicación geográfica (28,4%) o una frase descriptiva (27,6%). Con menor frecuencia, pueden incluir especificaciones sobre los parámetros de la toma o el modelo de la cámara usada (11,2%), alguna anécdota o reflexión en torno a la toma de la fotografía (10,3%), una mención de autoría (5,2%), el mismo título, repetido o traducido (4,3%), la explicación de algún aspecto relacionado con el contenido, a modo de comentario o definición (4,3%), un enlace (3,4%), la reproducción textual de algún texto ajeno (3,4%) o una descripción amplia del contenido (2,6%). En contadas ocasiones (1,7%), la descripción no guarda una verdadera relación con la imagen.

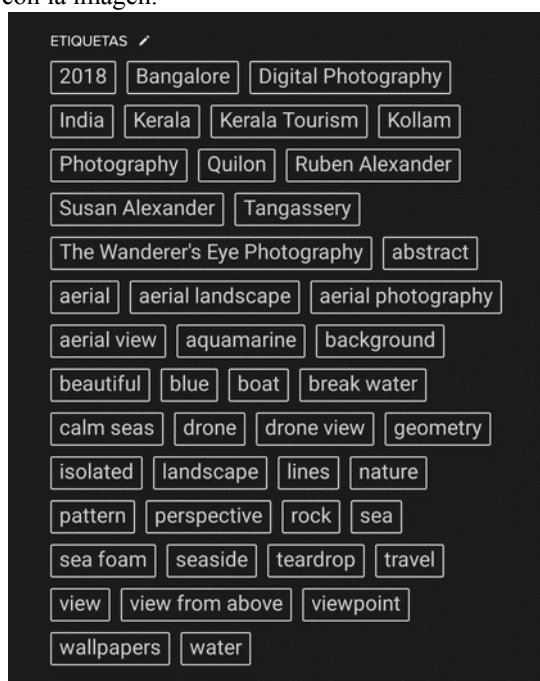


Figura 1. Algunos creadores se esmeran en representar el contenido de forma exhaustiva y desde numerosos puntos de vista: aspectos denotativos y connotativos, forma fotográfica, ubicación, fecha y referencias al propio autor, entre otros elementos

El etiquetado mediante palabras clave (fig. 1) está presente en el 71% de la muestra analizada. En ese subconjunto, entre las etiquetas suministradas manualmente predominan (43,6%) las de naturaleza denotativa (\bar{X} =10 etiquetas/imagen; σ =8,5). El resto son alusiones al equipo fotográfico (cámara, óptica, soporte, etc.: 15,9%; \bar{X} =4,1;

$\sigma=5,2$), referencias a la forma fotográfica o a la técnica (p. ej., gran angular, panorámica, larga exposición, etc.: 10,9%; $\bar{X}=2,8$; $\sigma=2,3$), aspectos connotativos (8,1%; $\bar{X}=2,1$; $\sigma=2,2$) y etiquetas heterogéneas sin un propósito descriptivo (meras referencias personales) o sin una relación explícita con la imagen (7,4%; $\bar{X}=1,9$; $\sigma=1,3$).

Durante la fase del estudio en que el etiquetado automático estaba disponible, el 45,2% de fotografías incluía palabras clave añadidas por el algoritmo de reconocimiento automático. Eran términos que solían representar elementos denotativos relativamente genéricos (p. ej., montaña, carretera, mar, edificio, nieve, cielo, etc.) y que podían resultar útiles para precisar búsquedas que abarquen algún concepto más. La tasa de acierto es del 96% y, aunque de este dato podría inferirse que el mecanismo de etiquetado automático, en las imágenes en las que se aplica, es muy fiable, cabe señalar que los usuarios siempre pueden suprimir aquellas etiquetas que consideran erróneas, por lo que resulta difícil hacer una valoración precisa.

Los metadatos técnicos EXIF están presentes en el 86% de las imágenes. Estos metadatos suelen incorporarse al fichero digital en el momento de su creación y suelen preservarse en la cadena del procesamiento fotográfico. Su ausencia puede explicarse por varios motivos: pueden haber sido eliminados u ocultos voluntariamente por el autor o pueden haber desaparecido a causa de una gestión incorrecta en la cadena de procesamiento (sobre todo, tras el revelado digital o el retoque fotográfico), p. ej., por una configuración inadecuada o por haber utilizado *software* obsoleto.

En cuanto a la geolocalización, como es bien sabido, muchas cámaras y teléfonos móviles tienen incorporado un receptor GPS capaz de identificar y almacenar la posición geográfica durante la toma de una fotografía. Cuando la cámara no dispone de esta característica, el usuario puede añadir manualmente las coordenadas GPS con la ayuda de *Adobe Lightroom* o mediante programas de geoetiquetado como *GeoSetter* o *GPicSync*, basados en la biblioteca multiplataforma *ExifTool*. La interfaz de *Flickr* no da pistas sobre el origen (nativo o manual) de este metadato, por lo que es difícil identificar en qué proporción este ha sido añadido *ex profeso*; ello es interesante, sobre todo ante la ausencia de metadatos EXIF, puesto que estos metadatos al menos darían pistas sobre la cámara usada. En cualquier caso, la geolocalización está presente en el 29% de las fotografías, que en tales casos las abre al descubrimiento por la vía de la exploración geográfica (<https://flickr.com/map>).



Figura 2. Los metadatos técnicos EXIF incluyen, entre otras informaciones, el tipo de cámara usada, los parámetros de la toma y la ubicación geográfica.

Desde el punto de vista archivístico, la concepción que *Flickr* tiene sobre el contenido generado por cada usuario es el de una colección cuyas piezas se pueden organizar libremente en álbumes o galerías, que actúan entonces como mecanismos de agregación. Esta concepción es análoga a la de otros servicios de publicación de fotografías (p. ej., *Google Fotos*, *iCloud*, *500px*) e incluso a la de aplicaciones que gestionan otro tipo de colecciones (p. ej., *iTunes*, *Kindle*), basados en una biblioteca

cuyos objetos pueden agruparse y ordenarse en listas o subcolecciones que reflejan representaciones mentales.

La inclusión en álbumes es una estrategia de agregación derivada de la práctica de clasificar en carpetas, habitual en los sistemas de archivo personales. Los álbumes son un constructo análogo a las carpetas, con la diferencia de que la organización en álbumes no permite la clasificación jerárquica en categorías y subcategorías, aunque sí la clasificación múltiple: una fotografía puede añadirse a más de un álbum, o puede incluso no formar parte de ninguno. En el contexto de la muestra analizada, cada fotografía está incluida en una media de 1,5 álbumes ($\sigma=1,7$). Como muestra del uso efectivo de la categorización múltiple, aproximadamente un tercio de las imágenes (34,5%) aparecen agregadas en más de un álbum a la vez.

Así como la inclusión de fotografías en álbumes es un método organizativo, la inclusión en grupos temáticos se suele abordar como una estrategia visibilizadora que favorece el descubrimiento por parte de otros usuarios. Los usuarios de *Flickr* pueden agregar cada una de sus fotografías a un máximo de 30 grupos en la versión gratuita del servicio, o 60 en la versión de pago. Ese límite está concebido como una función anti-*spam*, del que cabe matizar que algunos grupos han optado por excluirse y ello explica que algunas fotografías puedan estar presentes en un número superior de grupos. La comunidad usuaria, en general, parece consciente del potencial visibilizador de esta estrategia, pues cada fotografía de este subconjunto está incluida en una media de 20,2 grupos, aunque en una cantidad muy variable ($\sigma=27,5$). En efecto, el 86,5% de las fotografías de la muestra está presente en un mínimo de dos grupos, y el 19% lo está en un mínimo de tres.

5 Discusión y conclusiones

La conducta observada en este estudio revela estrategias heterogéneas de organización social que contrastan con los laxos esfuerzos de gestión desempeñados en el ámbito personal. En el contexto de *Flickr*, se aprecian, en general, esfuerzos efectivos dedicados a representar, con mayor o menor acierto, el contenido de las fotografías: la práctica deliberada de titular, describir y etiquetar las imágenes está muy generalizada y denota un claro interés por favorecer el reencuentro y el descubrimiento del material. En contraste con las prácticas rudimentarias que varios estudios observan en el ámbito personal, en el que las posibilidades de organización se centran en la estructuración mediante carpetas, las actividades de organización desarrolladas en un contexto social como *Flickr* denotan una actitud más proactiva y una motivación mayor por caracterizar la información y desarrollar estrategias preparatorias para el descubrimiento.

Se percibe, sin embargo, una notable variabilidad en el tipo de títulos asignados a las imágenes, en el estilo de las descripciones, y en el número y la naturaleza de las etiquetas. Aunque la tendencia mayoritaria es la asignación de títulos representativos,

una proporción considerable de usuarios opta por títulos de carácter simbólico o alegórico que, más que favorecer la recuperación, buscan resaltar la dimensión artística de la imagen. Algo similar ocurre con las descripciones: pese a que la mayoría presenta un texto de cierta utilidad para la recuperación, también hay una cantidad significativa de imágenes sin una descripción relevante o útil para la recuperación.

En cuanto al etiquetado, destaca la tendencia a asignar, sobre todo, etiquetas de naturaleza denotativa que representan aspectos factuales y objetivos del contenido o del contexto de las fotografías. A esta tendencia le sigue, de lejos, cierta inclinación a representar aspectos de la cámara o de la toma, una práctica que puede considerarse lógica en el contexto de una comunidad de aficionados, entusiastas y profesionales de la fotografía. En contraste con lo que se observaría en un banco de imágenes comercial, los usuarios de *Flickr* apenas asignan etiquetas de naturaleza connotativa como, p. ej., términos relativos a emociones o a atributos de algún objeto. Es razonable pensar que el carácter subjetivo de este tipo de etiquetas las hace menos obvias para el usuario y menos útiles para el descubrimiento.

La organización de las fotografías personales mediante etiquetas presenta las ventajas y los inconvenientes propios del etiquetado social, reconocidos en la bibliografía. El etiquetado, por un lado, ofrece un enfoque atractivo para la descripción y el descubrimiento de recursos, pero, por otro lado, está expuesto a una serie de inconvenientes ya conocidos como, p. ej., la tendencia a privilegiar unas cosmovisiones e ignorar o marginar otras, la irregular calidad de la indexación o las propias motivaciones por etiquetar, que abarcan razones tanto egoístas como altruistas (Rafferty, 2011 y 2018).

Uno de los principales retos está en el hecho de que la indexación basada en el usuario, para ser realmente funcional debe incluir algún elemento de control y de «autoridad representativa» (Wright, 2005). Interesa, pues, que se desarrollen herramientas orientadas a disciplinar el etiquetado, p. ej., mediante sistemas de recomendación de etiquetas que fomenten la consolidación del vocabulario al promover etiquetas adecuadas a un recurso (Ding *et al.*, 2010).

Aun así, los resultados confirman los hallazgos de Marlow *et al.* (2006), Ames y Naaman (2007) y Wash y Rader (2007): el uso de las diversas estrategias organizadoras depende en buena medida de la motivación de cada persona para usar el sistema. Los esfuerzos dedicados al etiquetado denotan interés por la organización social y por facilitar el descubrimiento, mientras que el uso del sitio como copia de seguridad en línea (organización personal) conduce a centrar los esfuerzos en organizar las fotografías en colecciones o álbumes, sin ignorar por ello que también las palabras clave tienen una función agrupadora.

El etiquetado automático, por su parte, es una ayuda efectiva a la recuperación, especialmente en aquellas piezas en las que el creador no ha incluido suficientes palabras clave representativas del contenido de la imagen. La tasa de acierto es muy

elevada, por lo que se trata de un mecanismo fiable que presenta un evidente potencial para mejorar la exhaustividad y la precisión de los resultados de la búsqueda.

Por último, dado que la gestión también implica protección, cabe alertar sobre los riesgos que presenta *Flickr*, como cualquier otro servicio de alojamiento remoto, relativos a la disponibilidad de la información. Al depositar sus materiales en un servicio externo, los usuarios están indefensos frente a los caprichos y las eventualidades de la empresa propietaria, que de forma repentina podría cambiar su política e incluso cesar su actividad, circunstancias que conllevarían la pérdida de la información allí alojada. Muestra de ello es que los términos de uso del servicio advierten de que cualquier cuenta puede cerrarse sin previo aviso. *Flickr* es tristemente notorio, además, por su escasa fiabilidad; son frecuentes las fallas y caídas del servicio que comprometen el acceso a la información. Ante todos estos peligros cabe cuestionarse, pues, la conveniencia de usar este tipo de sitios como archivo personal duradero o como copia de seguridad. *Flickr* puede ser un eslabón más en la cadena de las actividades de mantenimiento y protección de fotografías personales, pero no debería ser la única solución para la preservación a largo plazo, ya que no hay garantía de que el proveedor del servicio mantendrá de forma permanente el acceso al contenido.

Referencias

- Ames, Morgan; Eckles, Dean Naaman, Mor, Spasojevic, Mirjana; Van House, Nancy (2010). Requirements for mobile photoware. *Personal and Ubiquitous Computing*, v. 14, n. 2, p. 95–109. <https://doi.org/10.1007/s00779-009-0237-4>
- Ames, Morgan; Naaman, Mor (2007). Why we tag: motivations for annotations in mobile and online media. En: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Nueva York: ACM Press, p. 971–980.
<http://infolab.stanford.edu/~mor/research/chi2007-Ames-whyWeTag.pdf>
- Angus, Emma; Thelwall, Mike (2010). Motivations for image publishing and tagging on Flickr. En: *Proceedings of the 14th International Conference on Electronic Publishing*, p. 189–204. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.833.8235>
- Bergman, Olef; Whittaker, Steve (2016). *The science of managing our digital stuff*. Cambridge, MA: MIT Press.
- Bushey, Jessica (2015). Convergence, connectivity, ephemeral and performed: new characteristics of digital photographs. En: Pang, Natalie; Liew, Kai Khuin; Chan, Brenda (eds.). *Participatory archives in a world of ubiquitous media*. Londres; Nueva York: Routledge, p. 33–47. <https://doi.org/10.1080/01576895.2014.881262>

- Cox, Andrew M.; Clough, Paul D.; Marlow, Jennifer (2008). *Flickr: a first look at user behaviour in the context of photography as serious leisure*. *Information Research*, v. 13, n. 1, marzo. <http://informationr.net/ir/13-1/paper336.html>
- Ding, Ying; Jacob, Elin K.; Zhang, Zhixiong; Foo, Schubert; Yan, Erjia; George, Nicolas L.; Guo, Lijiang (2009). Perspectives on social tagging. *Journal of the American Society for Information Science and Technology*, v. 60, n. 12, p. 2388–2401. <http://ntu.edu.sg/home/sfoo/publications/2009/2009-JASIST-PerspectivesOnSocialTagging.pdf>
- Flickr (2018). Tag keywords in *Flickr*. En: *Flickr Help*. https://help.flickr.com/en_us/tag-keywords-in-flickr-BJUJpQoyX
- Frohlich, David; Kuchinsky, Allan; Pering, Celine; Don, Abbe; Ariss, Steven (2002). Requirements for photoware. En: *CSCW '02: Proceedings of the 2002 ACM conference on computer-supported cooperative work*, p. 166–175. <https://doi.org/10.1145/587078.587102>
- Gupta, Manish; Li, Rui; Yin, Zhijun; Han, Jiawei (2011). An overview of social tagging and applications. En: Aggarwal, Charu C. (ed.). *Social Network Data Analytics*. Nueva York, Springer, p. 447–497. https://doi.org/10.1007/978-1-4419-8462-3_16
- Heckner, Markus; Heilemann, Michael; Wolff, Christian (2009). Personal information management vs. resource sharing: towards a model of information behaviour in social tagging systems. En: *Proceedings of the Third International Conference on Weblogs and Social Media: ICWSM 2009*, p. 42–49. <https://aaai.org/ocs/index.php/ICWSM/09/paper/view/212>
- Heckner, Markus; Neubeuer, Tanja; Wolff, Christian (2008). Tree, funny, to_read, google: what are tags supposed to achieve? A comparative analysis of user keywords for different digital resource types. En: *SSM '08: Proceedings of the 2008 ACM workshop on Search in social media*, p. 3–10. <https://doi.org/10.1145/1458583.1458589>
- Hern, Alex (2015). *Flickr faces complaints over “offensive” auto-tagging for photos* *The Guardian*, 20 de mayo. <https://theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos>
- Jones, William. Personal information management. En: Cronin, Blaise (ed.), *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, 2007, v. 41, p. 453–504. <https://doi.org/10.1002/aris.2007.1440410117>
- Karger, David R. Unify everything: it's all the same to me. En: Jones, William; Teevan, Jaime (2007). *Personal information management*. Seattle y Londres: University of Washington Press, p. 127–152.
- Kennedy, Lyndon; Naaman, Mor; Ahern, Shane; Nair, Rahul; Rattenbury, Tye (2007). How *Flickr* helps us make sense of the world: context and content in community-contributed media

- collections. *Proceedings of the 15th ACM international conference on Multimedia, September 24–29, Augsburg, Germany*. Nueva York: ACM, p. 631–640.
- Kirk, David; Sellen, Abigail; Rother, Carsten; Wood, Ken (2006). Understanding photowork. En: *CHI 2006: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Montreal, 22–27 de abril.
- <https://doi.org/10.1145/1124772.1124885>
- Koochali, Alireza; Kalkowski, Sebastian; Dengel, Andreas; Borth, Damian; Schulze, Christian (2016). What languages do people speak on *Flickr*? A language and geo-location study of the YFCC100m dataset. En: *MMCommons '16: Proceedings of the 2016 ACM Workshop on Multimedia Commons*, p. 35–42. <https://doi.org/10.1145/2983554.2983560>
- Marlow, Cameron; Naaman, Mor; Boyd, Danah; Davis, Marc (2006). Position paper, tagging, taxonomy, flickr, academic article, toRead. En *Hypertext '06: Proceedings of the 17th conference on Hypertext and hypermedia (Odense, August 22–25 2006)*. Nueva York: ACM, p. 31–40. <https://danah.org/papers/Hypertext2006.pdf>
- Marshall, Catherine C. How people manage personal information over a lifetime. En: Jones, William; Teevan, Jaime (2007). *Personal information management*. Seattle y Londres: University of Washington Press, p. 57–75.
- Nov, Oded; Naaman, Mor; Ye, Chen (2009a). Analysis of participation in an online photosharing community: A multidimensional perspective. *Journal of the American Society for Information Science and Technology*, v. 61, n. 3, p. 555–566. <https://doi.org/10.1002/asi.21278>
- Nov, Oded; Naaman, Mor; Ye, Chen (2009b). Motivational, structural and tenure factors that impact online community photo sharing. En: *Proceedings of the Third International ICWSM Conference*, p. 138–145. <https://aaai.org/ocs/index.php/ICWSM/09/paper/view/206>
- Okabe, Daisuke (2004). Emergent social practices, situations and relations through everyday camera phone use. *2004 International Conference on Mobile Communication*. Seúl. http://itofisher.com/mito/archives/okabe_seoul.pdf
- Petrelli, Daniela; Whittaker, Steve; Brockmeier, Jens (2008). AutoTopography: what can physical mementos tell us about digital memories?. En: *CHI '08 ACM conference on human factors in computing systems*. Florencia: 5–10 abril, p. 53–62. <http://shura.shu.ac.uk/2919>
- Pikhart, Marcel (2018). Multilingual and intercultural competence for ICT: accessing and assessing electronic information in the global world. En: Choroś, Kazimierz; Kopel, Marek; Kukla, Elżbieta; Siemiński, Andrzej (eds.). *Multimedia and network information systems: proceedings of the 11th International Conference MISSI 2018*. Cham: Springer, p. 273–278. https://doi.org/10.1007/978-3-319-98678-4_28
- Rafferty, Pauline (2011). «informative tagging of images: the importance of modality in interpretation. *Knowledge Organization*, v. 38, n. 4, p. 283–298.

- Rafferty, Pauline (2018). Tagging. *Knowledge Organization*, v. 45, n. 6, p. 500–516.
<http://isko.org/cyclo/tagging>
- Russo, Paul; Nov, Oded (2010). Photo tagging over time: a longitudinal study of the role of attention, network density and motivations. En: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, p. 146–153.
<https://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1511>
- Stuart, Emma (2013). Organizing photographs: past and present. En: *New directions in information organization*. Bingley: Emerald, p. 137–155.
[https://doi.org/10.1108/S1876-0562\(2013\)0000007011](https://doi.org/10.1108/S1876-0562(2013)0000007011)
- Stuart, Emma (2019). Flickr: organizing and tagging images online. En: *ISKO Encyclopedia of Knowledge Organization*. <http://isko.org/cyclo/flickr>
- Van House, Nancy (2007). Flickr and public image-sharing: distant closeness and photo exhibition. En: *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, p. 2717–2722. <https://doi.org/10.1145/1240866.1241068>
- Van House, Nancy (2011). Personal photography, digital technologies and the uses of the visual. *Visual Studies*, v. 26, n. 1, June, p. 125.
<https://doi.org/10.1080/1472586X.2011.571888>
- Van House, Nancy; Davis, Marc; Ames, Morgan; Finn, Megan; Viswanathan, Vijay (2005). The uses of personal networked digital imaging: An empirical study of cameraphone photos and sharing. En: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. Nueva York: ACM, p. 1853–1856. <https://doi.org/10.1145/1056808.1057039>
- Van House, Nancy; Davis, Marc; Takhteyev, Yuri; Ames, Morgan; Finn, Megan (2004). The social uses of personal photography: methods for projecting future imaging applications. Documento de trabajo.
http://people.ischool.berkeley.edu/~vanhouse/photo_project/pubs/vanhouse_et_al_2004b.pdf
- Wash, Rick; Rader, Emilee (2007). Public bookmarks and private benefits: an analysis of incentives in social computing. En: *Proceedings of the American Society for Information Science and Technology*, Milwaukee, Wisconsin. <https://doi.org/10.1002/meet.1450440240>
- Whittaker, Steve. Personal information management: from consumption to curation. En: Blaise Cronin (ed.). *Annual Review of Information Science and Technology*. Medford, NJ: Information Today, v. 45, p. 1–62. <https://doi.org/10.1002/aris.2011.1440450108>
- Whittaker, Steve; Bergman, Ofer; Clough, Paul (2010). Easy on that trigger, dad: a study of long term family photo retrieval. *Personal Ubiquitous Computing*, v. 14, n. 1, January, p. 31–43. <https://doi.org/10.1007/s00779-009-0218-7>

- Xu, Zhichen; Fu, Yun; Mao, Jianchang; Su, Difu (2006). Towards the semantic web: collaborative tag suggestions. En: *Proceedings of Collaborative Web Tagging Workshop*. Edimburgo. <http://www.rawsugar.com/www2006/13.pdf>
- Zhao, Xuan; Lindley, Siân. Curation through use: understanding the personal value of social media. En: *32nd Annual ACM Conference on Human Factors in Computing Systems, CHI 2014*. Nueva York: ACM, p. 2431–2440. <https://doi.org/10.1145/2556288.2557291>

Contributos da indexação social para tratamento temático de documentos fotográficos provenientes de arquivos públicos

Susana Sofia Cunha¹, Maria da Graça Melo Simões² y Luciana de Souza Gracioso³

¹ ORCID [0000-0003-4444-4319](https://orcid.org/0000-0003-4444-4319). Faculdade de Letras da Universidade de Coimbra.
cunha.susana@gmail.com

² ORCID [0000-0001-5575-0261](https://orcid.org/0000-0001-5575-0261). Faculdade de Letras da Universidade de Coimbra, Portugal.
gsimoes@fl.uc.pt

³ ORCID [0000-0002-6320-4946](https://orcid.org/0000-0002-6320-4946). Universidade Federal de São Carlos, BR. (Universidade de Coimbra/CAPES)
luciana@ufscar.br

Resumo. O carácter polissémico e subjetivo da fotografia dificulta a implementação de políticas e procedimentos comuns ao nível da sua descrição e recuperação. O surgimento da *web 2.0* potencia novas formas de publicação, partilha e organização da informação *online*. A folksonomia surge como um recurso interativo e inclusivo para representar estes conteúdos através da linguagem livre do utilizador. Tem-se como hipótese que a utilização da folksonomia poderia ser aplicada e teria utilidade no tratamento fotografias. O objetivo geral é analisar que contributos traz a folksonomia aos processos de indexação de documentos fotográficos provenientes de arquivos públicos, com vista à preservação e divulgação do património. A metodologia adotada consistiu numa pesquisa exploratória sistemática do tema da folksonimia em contexto arquivístico e aplicada à documentação fotográfica, para identificar trabalhos científicos em repositórios e bases de dados *online*. Paralelamente, foi criada uma coleção de fotografias *online* no Flickr, pertencentes à Coleção David Freitas do Arquivo Fotográfico da Câmara Municipal de Évora, sobre as quais serão promovidas ações de indexação social. O estudo prático levado a cabo contribuirá para aferir a importância do uso de sistemas de divulgação com possibilidade de indexação social, por parte das instituições detentoras de coleções de fotografia.

Palavras chave: Folksonomia; Flickr; Arquivo; Fotografia.

Abstract. The polysemous and subjective nature of photography has made it difficult to implement common policies and procedures concerning its description and recovery. The emergence of web 2.0 provides new ways of publishing, sharing and organizing information online. Folksonomy thus emerges as an interactive and inclusive resource for representing these contents through the user's free language. It is theorized that the use of folksonomy could be applied and would be useful in the treatment of photographic images. The main goal is to analyze which contributions folksonomy brings to the indexation processes of photographic documents coming from public archives,

in order to preserve and share this heritage. The adopted methodology consisted in a systematic exploration of folksonomy in archival context and applied to photographic documentation, to identify scientific works in online repositories and databases. An online collection of photographs was created on Flickr, belonging to the David Freitas Collection of the Photographic Archive of Évora City Hall, on which social indexing actions will be promoted. The practical study carried out will contribute to ascertain why systems that allow social indexing features are important to institutions holding photograph collections.

Keywords: Folksonomy; Flickr; Archive; Photography

1 Contextualização

A fotografia está presente em vários arquivos e bibliotecas e deve ser encarada como um documento com necessidades de tratamento e descrição próprios. O seu carácter polissémico e subjetivo, bem como os diferentes tipos de processos, suportes e materiais que a constituem, dificultam a implementação de políticas e procedimentos comuns ao nível da sua gestão, tratamento, descrição e recuperação, o que resulta em as instituições apliquem critérios próprios em função do tipo de documentação que possuem com base nas normas, orientações e legislação nacional e internacional, e em função dos seus objetivos e dos seus utilizadores. Matusiak (2006) aponta alguns desafios aquando da indexação de imagens, desde logo a ambiguidade da linguagem e as próprias limitações humanas: falta de consenso sobre que atributos de uma imagem devem ser indexados, dificuldades em determinar o nível apropriado de indexação, subjetividade, falta de consistência nos termos e dificuldade em fazer coincidir os mapas mentais de indexador com os do utilizador. Santos, Neves, & Albuquerque (2018) referem que *“o indexador, por meio do processamento de dados sensoriais, pode destacar os significados da imagem que mais lhe interessam, no momento da indexação, provenientes de sua percepção, podendo não necessariamente satisfazer as necessidades dos usuários”* (p.979). Neste sentido, a introdução da linguagem natural poderá ser útil como recurso complementar às atividades de indexação.

Hoje em dia aumentam as exigências dos utilizadores e a apresentação de documentos com suporte em papel é substituída pelo computador e pelos processos digitais. Iglesias Franch (2016) salienta que os responsáveis por património fotográfico se devem questionar sobre as oportunidades oferecidas pela evolução tecnológica e analisar as vantagens que estas trazem aos arquivos e instituições detentoras destes acervos. Estes desafios sobre os processos de representação de imagens tornam-se mais complexos quando transpostos para o ambiente digital, virtual e aberto a construções colaborativas. O surgimento da *web 2.0* ou, ainda, a *web* pragmática, potencia novas formas de publicação, partilha e organização da informação disponível *online*, nomeadamente de documentos fotográficos. O termo *folksonomy*, proposto por Thomas Vander Wal (2005) para designar a classificação

informal de conteúdos que vinha surgindo na *web 2.0*. surge, assim, como um recurso interactivo e inclusivo para representar estes conteúdos através da linguagem livre do utilizador. O Flickr é uma plataforma *online* de gestão e partilha de imagens e foi um dos pioneiros na utilização da folksonomia. Este espaço virtual é utilizado por várias instituições detentoras de documentação fotográfica para divulgação dos seus acervos e foi já objeto de análise por parte de vários autores (Beaudoin, 2006; Marlow, Naaman, Davis, & Hall, 2006; Aquino, 2008, Leitão, 2010, Rorissa, 2010; Stvilia & Jørgensen, 2010; Freixa-Font, 2011; Santos, 2016; Kipp, Beak, & Choi, 2017). Em Portugal, a Biblioteca de Arte da Fundação Calouste Gulbenkian, no Flickr desde 2008, é um bom exemplo de como uma instituição cultural pode beneficiar da presença nas redes sociais. Leitão (2010) faz uma análise desta participação, concluindo que o projeto resultou num aumento significativo da utilização das coleções disponibilizadas e que atraiu novos públicos. A marcação de imagens como favoritas, os comentários e as *tags* atribuídas pelos utilizadores resultaram em novas informações e conhecimento para a própria biblioteca.

Neste contexto, põe-se a hipótese da folksonomia poder ser aplicada, de modo complementar, a outras formas de tratamento temático de documentos, e ter utilidade, em algumas circunstâncias, no tratamento de imagens, já que possibilitaria uma indexação mais exaustiva, maior possibilidade de análise da imagem e a representação dessa análise em linguagem natural. Santos, Neves & Albuquerque (2018) desenvolveram um trabalho sistemático de análise de conteúdo das principais publicações sobre indexação de fotografias que confirmam este posicionamento. Da análise de 33 trabalhos académicos, concluem que as pesquisas sobre indexação social crescem e enfatizam a contribuição desta prática para *“reformulação/atualização dos métodos, recomendações, políticas de indexação e instrumentos tradicionais de indexação de imagens digitais (...) para fins de representação e recuperação da informação imagética”* (p. 976).

Propõe-se como objetivo geral analisar de que modo a indexação social contribui para os processos de indexação de documentos fotográficos provenientes de arquivos públicos, com vista à preservação do património. Os objetivos específicos são: definir e caracterizar a folksonomia enquanto prática de indexação social de fotografias, sobretudo em arquivos; selecionar e disponibilizar fotografias da coleção David Freitas do Arquivo Fotográfico da Câmara Municipal de Évora (AFCME), através do Flickr; descrever e sistematizar a interação dos utilizadores da plataforma em relação à prática de indexação social das imagens disponibilizadas e verificar em que perspectivas as palavras indicadas pelos utilizadores se aproximam da indexação controlada e da indexação automática (promovida pelo Flickr), sinalizando se a indexação coletiva ampliou ou não as possibilidades de representação temática das fotografias do respectivo arquivo.

No contexto dos objetivos delimitados identificamos, no âmbito da Ciência da Informação, os contributos de Mathes (2004), Quintarelli (2005), Golder & Huberman (2005), Aquino (2007), Catarino & Baptista (2009), Strehl (2011) e Yedid (2013) para definir e contextualizar a folksonomia em arquivos, e de Neal (2010), Gracioso

(2014) e Cordeiro (2018), quanto à sua aplicação em imagens. Autores como Marlow, Naaman, Davis & Hall (2006), Rorissa (2010), Leitão (2010) e Freixa-Font (2011) abordam o Flickr como instrumento de divulgação de coleções fotográficas. Considerando que o propósito desta pesquisa é a organização e representação da documentação fotográfica, apoiámo-nos, nesta linha, nos trabalhos de Kipp, Beak & Choi (2017) e Santos (2017).

O AFCME conserva cerca de 500 000 espécies fotográficas. Tem como objetivos proteger o património fotográfico local, recolher a produção fotográfica contemporânea e histórica, divulgar e promover coleções próprias e em depósito e disponibilizar à consulta pública essas coleções. A coleção David Freitas foi adquirida pela autarquia em 2000 e é constituída por 6470 espécies fotográficas (negativos e positivos). As imagens, realizadas entre 1940 e 1970, retratam a cidade de Évora, Alentejo e outras zonas de Portugal. Encontram-se, na coleção, grandes grupos temáticos: aspectos da cidade e do concelho, reportagens, levantamentos de obras municipais e exemplos de atividades sociais, constituindo-se como uma ferramenta essencial para a compreensão da história da cidade no século XX. A escolha desta coleção como objeto deste estudo justifica-se, assim, não só pela sua pertinência histórica, mas também por ser uma das coleções mais requisitadas do Arquivo Fotográfico.

2 Técnicas e procedimentos metodológicos

A metodologia adotada neste trabalho consistiu, numa primeira fase, numa pesquisa exploratória sistemática do tema da indexação social em contexto arquivístico e aplicada à documentação fotográfica. Procedeu-se à revisão da literatura nacional e internacional, tendo por base uma pesquisa selectiva para identificar trabalhos científicos em repositórios e bases de dados *online* (BDTD (Biblioteca Digital de Teses e Dissertações brasileiras), BRAPIC (Base de dados de revistas científicas em Ciência da Informação e áreas correlatas), BENANCIB (Repositório de trabalhos publicados nos Anais dos Encontros Nacionais de Pesquisa em Ciência da Informação), *Web of Science (Information Science & Library Science)*, EBSCO (*Library & Information Science Source*), B-on (Biblioteca do Conhecimento Online), RCAAP (Repositório Científico de Acesso Aberto em Portugal) e OASIS (Portal brasileiro de publicações científicas em acesso aberto)). A pesquisa, feita por títulos e assuntos, visou artigos publicados em revistas científicas ou atas de congressos, em acesso aberto e texto integral. Os termos utilizados para delimitar a pesquisa, em inglês, foram “Folksnom*” e suas variações terminológicas, como “Tag*” e “Social index*”, e a conjugação destes com “archiv*”, “imag*” e “photograph*”. Uma vez que parte deste estudo incide sobre a plataforma Flickr, este termo foi também pesquisado. Analisaram-se as referências bibliográficas dos artigos para identificar quais os autores mais referenciados, recuperando alguns desses artigos a partir dos links disponibilizados. No total, foram considerados 26 trabalhos que serviram de

base para a revisão da literatura. Paralelamente, foi desenvolvida uma pesquisa aplicada, voltada para a criação de uma coleção de fotografias *online* (no Flickr) sobre as quais foram promovidas ações de indexação controlada e social, que serão posteriormente objeto de análise e descrição¹. As 139 imagens da coleção David Freitas foram inseridas entre os dias 16 de Outubro de 2018 e 5 de Janeiro de 2019. As categorias temáticas elencadas com base na coleção existente foram: comércio e indústria, equipamentos e serviços, evolução urbana e sociabilidade, relativas à cidade de Évora. A cada imagem foi atribuída uma legenda com o título, cota, série, data e autor, bem como processo e formato original. Num segundo momento as imagens serão indexadas seguindo as orientações da literatura e tendo por base a lista de descritores organizada pelo AFCME, com vista à promoção de uma indexação controlada, que posteriormente servirá de base para as discussões sobre os contributos da folksonomia.

3 Considerações parciais

O número de visualizações nos primeiros dias da divulgação do AFCME no Flickr apontaram a importância que o projeto significou para a comunidade: entre Outubro de 2018 e Maio de 2019 foram contabilizadas 84 818 visualizações à Coleção David Freitas. A cobertura de visualizações é de 100% (todas as imagens foram visualizadas). As *tags* colocadas pelo AFCME correspondem aos critérios onomástico (nome do autor e, quando conhecido, o nome da pessoa fotografada), geográfico (Évora e, ocasionalmente, o local específico) e tipológico (processo fotográfico original). O AFCME não dispõe de um tesouro para indexação de fotografia, pelo que se optou pela linguagem natural. Das 139 imagens publicadas, 119 foram objeto de indexação social (85,6%), contabilizando-se 275 *tags*. Para a obtenção deste total foram somadas todas as ocorrências, repetidas ou não. Foram identificados 206 termos diferentes. A maioria das *tags* é em língua portuguesa; isto explica-se pela predominância de seguidores portugueses, sendo também portugueses todos os utilizadores que atribuíram *tags* às imagens disponibilizadas. Embora utilizando conceitos diferentes, observamos que as categorias utilizadas para a descrição e indexação de imagens (“quem”, “quando”, “onde”, “o que/ como”) foram utilizadas na indexação social e são os termos relacionados com o assunto os mais observados: objetos (como “automóvel”, “brinquedo”), aspetos arquitetónicos (como “arcadas”, “calçada romana”), edifícios (como “biblioteca”, “armazém”), marcas (como “Sagres”, “Ferguson”) e locais (“Alentejo”, “Praça do Giraldo”). Em menor número temos *tags* relativas ao momento temporal (“outono”, “*mid century*”) e foram identificados, também, conceitos abstratos, como “amizade”, “imaginário”, “condição feminina”. Registam-se casos pontuais de *tags* subjetivas, isto é, *tags* que refletem a opinião ou emoções do utilizador (“humor negro”, numa imagem de duas enfermeiras com um esqueleto, “recordaréviver”, alusivo ao slogan da Kodak, ou “romeuejulieta”,

1 <https://www.flickr.com/photos/arquivo fotografico cme/albums/72157701232379211>

numa imagem de duas crianças). Encontramos algumas das desvantagens associadas à folksonomia, como a ausência de padronização de substantivos (coincidem plural e singular, como em “mulher” e “mulheres” e feminino e masculino, em “operário” e “operária”), homografia (“arcada”, nome de um café em Évora e “arcada” referindo-se aos arcos da Praça do Giraldo), erros tipográficos e ocorrência de mais do que um idioma (português e inglês). Analisando outras investigações realizadas acerca da dinâmica das *tags* no Flickr, nomeadamente a de Leitão (2010), constatamos algumas semelhanças. Desde logo, a reduzida participação na atribuição de *tags*, explicada por alguns autores pelo facto de as pesquisas poderem ser feitas por títulos e descrições, assumindo as *tags*, para o utilizador, um papel secundário. Isto pode indicar que os utilizadores não têm por hábito atribuir *tags* às imagens disponibilizadas por terceiros. O maior obstáculo reside em envolver e motivar os utilizadores para colaborarem com estes sistemas, como salienta Rorissa (2010). O reconhecimento automatizado de imagens do Flickr proporciona a criação de *tags* que são convertidas em pontos de acesso na pesquisa e recuperação da informação. Foram contabilizadas 326 *tags* automáticas, atribuídas a 77,7% das imagens (31 imagens não foram objeto de identificação automática). Pese embora um número maior de ocorrências, encontramos um número menor de termos (44 termos diferentes). A maioria destes termos descreve o conteúdo (“edifício”, “rua”, “árvore”) e um descritor confere dados acerca do processamento da imagem (“monocromático”).

Carecendo, ainda, de uma análise mais aprofundada, nomeadamente o cruzamento das *tags* atribuídas pelos utilizadores e pelo sistema com um vocabulário controlado para descrição de fotografia, podemos já aferir que a atribuição de *tags* pelos utilizadores numa plataforma *online* de um arquivo, biblioteca ou centro de documentação fotográfico, pode, à partida, refletir um interesse individual de recuperar o documento. Se, por um lado, a seleção das *tags* é subjetiva e resulta de diferentes interpretações e motivações, por outro, como salienta Santos (2016), esse conjunto de termos constitui diversidade e maior quantidade de pontos de acesso em linguagem natural. A indexação social potencia o desenvolvimento da consistência do utilizador-indexador, já que quem atribui as *tags* é, muito provavelmente, quem as vai utilizar para recuperar a informação, juntamente com aqueles que consigo partilham os mesmos interesses e vocabulários. A prática na área profissional na conservação de fotografia leva a concluir que muitas coleções fotográficas de arquivos e bibliotecas não são eficazmente divulgadas por falta de partilha *online*. Por outro lado, quando divulgadas, as plataformas não respondem às necessidades de recuperação e organização do utilizador comum. O estudo prático levado a cabo na plataforma Flickr contribuirá para aferir a importância do uso de sistemas de divulgação com possibilidade de indexação social por parte das instituições detentoras de acervos fotográficos. Desta forma, não só potenciariam o conhecimento do acervo e criariam pontos de acesso úteis para a recuperação de informação, como ajudariam a compreender se a utilização de folksonomias se assume como uma prática eficiente e eficaz na recuperação de imagens para fins de investigação, estudo e outros que não os meramente sociais.

Referências

- Aquino, M. C. (2007). Hipertexto 2.0, folksonomia e memória coletiva : Um estudo das tags na organização da web. *Revista E-Compós*, 9, 1–18.
- Aquino, M. C. (2008). A folksonomia como hipertexto potencializador de memória coletiva: Um estudo dos links e das tags no de.licio.us e no Flickr. *Liinc Em Revista*, 4(2), 303–320. <https://doi.org/10.18617/liinc.v4i2.263>
- Beaudoin, J. (2006). Flickr image tagging : Patterns made visible. *Bulletin of the American Society for Information Science & Technology*, 34(1), 26–30.
- Catarino, M. E., & Baptista, A. A. R. P. (2009). Integração das folksonomias aos metadados : um perfil de aplicação. In “*Encontro Nacional de Pesquisa em Ciência da Informação*”. João Pessoa : ANCIB. Retrieved from <http://repositorium.sdum.uminho.pt/handle/1822/9697>
- Cordeiro, R. I. N. (2018). O delineamento de uma pesquisa em imagens e audiovisuais na ciência da informação: o “tagueamento” como quarta dimensão. *Informação & Informação*, 23(1), 6–30. <https://doi.org/10.5433/1981-8920>.
- Freixa-Font, P. (2011). Patrimonio fotográfico y web 2.0: La experiencia Flickr The Commons. *El Profesional de La Información*, 20(4), 432–438. <https://doi.org/10.3145/epi.2011.jul.10>
- Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198–208. <https://doi.org/10.1177/0165551506062337>
- Golder, S., & Huberman, B. A. (2005). The structure of collaborative tagging systems. Retrieved from <http://arxiv.org/abs/cs/0508082>
- Gracioso, L. S. (2014). Organização social de fotografias na Web: Fundamentos, métodos e orientações. In L. M. A. Sousa, M. S. L. Fujita, & L. S. Gracioso (Eds.), *A imagem em Ciência da Informação: Reflexões teóricas e experiências práticas* (pp. 207–226). Marília, São Paulo: Cultura Acadêmica Editora.
- Iglesias Franch, D. (2016). Fotografía, tecnología y comunicación. Una confluencia de intereses para el archivo digital. *Patrimonio Cultural de España*, (11), 49–58.
- Kipp, M. E. I., Beak, J., & Choi, I. (2017). Motivations and intentions of Flickr users in enriching Flickr records for Library of Congress photos. *Journal of the Association for Information Science and Technology*, 68, 2364–2379. <https://doi.org/10.1002/asi>
- Leitão, P. (2010). Uma biblioteca nas redes sociais: o caso da Biblioteca de Arte da Fundação Calouste Gulbenkian no Flickr. *Actas Do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas - Políticas de Informação Na Sociedade Em Rede*, (10). Retrieved from <http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/189>
- Marlow, C., Naaman, M., Davis, M., & Hall, S. (2006). Position paper, tagging, taxonomy, Flickr, rticle, toread. *Structure*, 9. <https://doi.org/10.1145/1149941.1149949>

- Mathes, A. (2004). Folksonomies - Cooperative classification and communication through shared metadata. Retrieved May 11, 2018, from <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Matusiak, K. K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services*, 22(4). <https://doi.org/10.1108/10650750610706998>
- Neal, D. M. (2010). Emotion-based tags in photographic documents : The interplay of text, image, and social influence, (3). Retrieved from <https://eds.b.ebscohost.com/eds/detail/detail?vid=26&sid=54313bef-49f7-491d-9c8c-40dada12083c%40sessionmgr103&bdata=JkF1dGhUeXB1PWlwLGNvb2tpZSxzGhLiHVpZCZsYW5nPB0LWJyJnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=74267901&db=lxh>
- Quintarelli, E. (2005). Folksonomies: power to the people. In *Comunicação apresentada no ISKO Italy - UniMib meeting*. Milano. Retrieved from <http://www.iskoi.org/doc/folksonomies.htm>
- Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science & Technology*, 61(11), 2230–2242. <https://doi.org/DOI: 10.1002/asi.21401>
- Santos, R. F., Neves, D. A. B., & Albuquerque, M. E. B. C. (2018). Pesquisas sobre indexação colaborativa de imagens na ciência da informação: Abordagens e perspectivas de estudos. In *Comunicação apresentada no Encontro nacional de pesquisa em Ciência da Informação - XIX ENANCIB* (pp. 976–999). Londrina PR. Retrieved from <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/1303>
- Santos, T. H. N. (2016). Indexação social de imagens por meio do Flickr. *Revista Photo & Documento*, (1). Retrieved from <http://gpaf.info/photoarch/index.php?journal=phd&page=article&op=view&path=14>
- Santos, T. H. N. (2017). *A taxonomia e a folksonomia na recuperação da informação: Um estudo no acervo de imagens da Fundação Marques da Silva (FIMS)*. (Tese de Doutorado). Faculdade de Letras da Universidade do Porto, Porto, Portugal. Retrieved from <http://hdl.handle.net/10216/105840>
- Strehl, L. (2011). As folksonomias entre os conceitos e os pontos de acesso : as funções de descritores, citações e marcadores nos sistemas de recuperação da informação. *Perspectivas Em Ciência Da Informação*, 16(2), 101–114. <https://doi.org/10.1590/S1413-99362011000200007>
- Stvilia, B., & Jørgensen, C. (2010). Member activities and quality of tags in a collection of historical photographs in Flickr. *Journal of the American Society for Information Science & Technology*, 61(12), 2477–2489. <https://doi.org/DOI: 10.1002/asi.21432>

Vander Wal, T. (2005). Folksonomy definition and Wikipedia. Retrieved April 21, 2018, from <http://www.vanderwal.net/random/entrysel.php?blog=1750>

Yedid, N. (2013). Introducción a las folksonomías: Definición, características y diferencias con los modelos tradicionales de indización. *Informacion, Cultura y Sociedad*, 29(1), 13–26.

Análisis de los servicios 'a la carta' de las televisiones autonómicas españolas

Laura Anton¹ y Javier Guallar²

¹ ORCID [0000-0001-8615-5932](https://orcid.org/0000-0001-8615-5932). Betevé, Barcelona, España.
lauraansa@gmail.com

² ORCID [0000-0002-8601-3990](https://orcid.org/0000-0002-8601-3990). Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona, Centre de recerca en Informació, Comunicació i Cultura CRICC, Barcelona, España.
jguallar@gmail.com

Resumen. Se estudian los archivos audiovisuales en internet de las televisiones autonómicas españolas, en base a veintiséis indicadores agrupados en cinco apartados (aspectos generales, contenido, sistema de consulta, página de resultados y características del audiovisual). Se analizan las doce televisiones autonómicas pertenecientes a la FORTA y se comparan los resultados actuales con los recogidos en 2012 en un trabajo anterior de los autores. Como principales conclusiones, se puede señalar que los servicios 'a la carta' de las televisiones autonómicas españolas aún tienen un nivel de desarrollo medio, ya que de los 58 puntos que las televisiones pueden alcanzar en el total de indicadores, las cadenas analizadas consiguen entre 27 y 38 puntos. Entre sus puntos débiles, se puede destacar que se limitan a ofrecer una búsqueda simple, sin opciones de filtros ni de gestión de los resultados. Estas carencias se han convertido en un problema enquistado de estos servicios, ya que hace siete años las televisiones analizadas presentaban estos mismos puntos débiles y solo en un caso se ha observado una mejora.

Palabras clave: Televisión; Documentación audiovisual; Archivos audiovisuales; Archivos de televisión; Servicios

Abstract. We study the audiovisual archives on the Internet of Spanish autonomous a la carta televisions, based on twenty-six indicators grouped into five sections (general aspects, content, query system, results page and audiovisual characteristics). The twelve regional televisions belonging to the FORTA are analyzed and the current results are compared with those collected in 2012 in a previous work by the authors. As main conclusions, it can be pointed out that the 'a la carte' services of the Spanish autonomous televisions still have a medium level of development, since of the 58 points that the televisions can reach in the total of indicators, the analyzed archives get between 27 and 38 points. Among its weaknesses, it can be noted that they simply offer a simple search, without filter options or results management. These deficiencies have become a cyclical problem of 'a la carte services': seven years ago the televisions analyzed presented these same weaknesses and only in one case an improvement has been observed.

Keywords: Television; Audiovisual documentation; Audiovisual archives; Television archives; A la carte services.

1 Introducción

La presente comunicación estudia los sistemas de acceso a la información audiovisual que ofrecen las televisiones en sus sitios web, concretamente, las características de los servicios ‘a la carta’ de las televisiones autonómicas españolas. Los datos obtenidos en la actualidad se comparan con los obtenidos en un estudio anterior de los autores (Anton y Guallar, 2014).

2 Antecedentes

La presente comunicación se inscribe en la especialidad de la documentación audiovisual, un campo de estudio académico de particular y notable dinamismo en España, en correspondencia a la importancia de esta actividad profesional en las empresas televisivas (Caldera-Serrano y Arranz-Escacha, 2012; Giménez-Rayó, 2012; Hidalgo Goyanes, 2013; López-de Quintana-Sáenz, 2014). Dentro del amplio repertorio de trabajos publicados sobre documentación en televisión, esta investigación guarda más relación con aquellos que se centran en analizar diversas televisiones en su conjunto; como los dedicados a centros de documentación de televisiones (Giménez-Rayó y Guallar, 2014; Inarejos y Guallar, 2015), a sus páginas web (Díaz-Campo, 2014), y especialmente, a los dedicados al objeto de estudio de esta comunicación, los archivos de las televisiones en internet (Anton y Guallar, 2014; De Mon y Guallar, 2014).

Por otra parte, la metodología de investigación evaluativa empleada en este trabajo en base a una serie de indicadores a los que se asigna una puntuación, tiene varios precedentes, aplicados en diversos contextos y temáticas, como por ejemplo: sitios web en general (Pedraza, Codina y Guallar, 2016); portales de archivos (Perpinyà Morera y Cid-Leal, 2018); hemerotecas de prensa (Guallar, Abadal y Codina, 2012), bases de datos de prensa (Linares, Codina, Abadal y Guallar, 2016); webs de televisiones (Díaz Campo, 2014); software de archivos audiovisuales (Cascón-Katchadourian, Ruiz-Rodríguez y Alberich-Pascual, 2018); y, como señalábamos antes, archivos de televisiones en internet (Anton y Guallar, 2014; De Mon y Guallar, 2014). Todos estos referentes se han tenido en cuenta para la confección de la presente propuesta.

3 Objetivos

El objetivo principal de la presente comunicación es analizar y evaluar las características de los servicios ‘a la carta’ de las televisiones autonómicas españolas, y comparar los datos recogidos con los obtenidos en el año 2012 y publicados en 2014. Concretamente, se establecen cuatro objetivos específicos:

- Actualizar el método de evaluación propuesto en el año 2012 en función de las novedades y los cambios tecnológicos observados en los últimos años.

- Ampliar la muestra estudiada para obtener una visión más amplia del panorama.
- Detectar puntos fuertes y carencias de los archivos estudiados y compararlos con los resultados obtenidos en la evaluación de 2012.
- Obtener una visión global de la evolución de los servicios ‘a la carta’ en 7 años.

La evaluación debe permitir conocer aspectos como la tipología de contenidos que ofrecen, la cobertura temporal, el grado de actualización, las prestaciones del sistema de consulta y las características de cada documento audiovisual.

4 Metodología

Para responder a estos objetivos se realiza un estudio de investigación de orientación evaluativa sobre una muestra de las televisiones autonómicas pertenecientes a la FORTA (Federación de Organismos de Radio y Televisión Autonómicos). En nuestro estudio anterior se seleccionaron las seis televisiones de la Federación con mayor audiencia en internet según la clasificación de OJD entonces:

- Televisió de Catalunya <https://www.ccma.cat/tv3/>
- Canal Sur <http://www.canalsur.es/>
- À Punt <https://apuntmedia.es/>
- Televisión de Galicia <http://www.crtvg.es/>
- Euskal Irrati Telebista <https://www.eitb.eus/>
- Telemadrid <http://www.telemadrid.es/>

El caso de la televisión valenciana es particular. En el anterior trabajo se analizó Canal 9, por entonces la televisión de la comunidad. En noviembre de 2013 cesaron sus emisiones y la cadena cerró y en abril de 2018 nació À Punt, una televisión nueva pero a la vez heredera de Canal 9. Este aspecto se ha tenido en cuenta en la comparativa con los datos obtenidos en 2012.

Con el objetivo de aumentar la muestra y así obtener una visión más global, en este trabajo se analizan también las seis televisiones autonómicas restantes pertenecientes a la FORTA:

- Aragón Televisión <http://www.aragontelevision.es/>
- Televisión del Principado de Asturias <https://www.rtpa.es/television>
- Televisión Canaria <http://www.rtv.es/television/>
- Televisión de la Región de Murcia <http://7tvregiondemurcia.es/>
- Castilla-La Mancha Media <http://www.cmmedia.es/>
- IB3 <https://ib3.org/televsio>

El análisis de los servicios a la carta de las televisiones se ha llevado a cabo entre los meses de marzo y junio de 2019 siguiendo la propuesta metodológica de evaluación que se detalla en el siguiente apartado.

5 Sistema de análisis de los servicios de Televisión a la carta

Para evaluar las características de los servicios ‘a la carta’, se han fijado una serie de indicadores agrupados en cinco grandes epígrafes: aspectos generales, contenido, sistema de consulta, página de resultados y características del audiovisual.

Para los indicadores se ha establecido un sistema de puntuación: 0 (inexistencia), 1 (nivel bajo), 2 (nivel medio) y 3 (nivel alto), si bien este sistema no se aplica de forma estricta en todos los casos: algunos indicadores se puntúan del 0 al 2 y en otros casos son binarios (0-1), cuando el aspecto analizado no permite establecer una gradación, señalando entonces su ausencia o presencia. Además, se han incluido dos indicadores únicamente informativos que no tienen asignada una puntuación.

Los indicadores establecidos se basan mayoritariamente en nuestro anterior trabajo (Anton y Guallar, 2014), que sirve de referencia principal al presente, con algunos cambios, que se detallan a continuación:

- Se han añadido dos indicadores procedentes de un trabajo posterior (De Mon Martín y Guallar, 2014). Se trata concretamente de los indicadores de Publicidad y Calidad del visionado. En este último caso, De Mon Martín y Guallar establecían una puntuación de 0 a 2 y se ha considerado conveniente reducir este rango, así que se asignará una puntuación de 0 a 1.
- Se han modificado dos indicadores procedentes de nuestro estudio anterior, aplicando un cambio en la puntuación. Se trata de los indicadores relativos al Lenguaje de interrogación y las Opciones de visionado. Este aspecto se ha tenido en cuenta en la comparativa con los datos del estudio anterior.
- Se han añadido dos indicadores nuevos a la propuesta: Servicios personalizados y Televisión en directo, que recogen algunos de los cambios observados en los últimos años en las televisiones en internet.

A continuación se define cada uno de los 26 indicadores de análisis propuestos, especificando el sistema de puntuación aplicado en cada caso:

A. ASPECTOS GENERALES	
Indicador	Descripción
A1. Denominación	Término o conjunto de términos empleados por cada televisión para nombrar su archivo audiovisual en línea. La denominación debe ser comprensible e inequívoca para cualquier usuario que visite el sitio web de la televisión. Este indicador es únicamente informativo y, por este motivo, no tiene asignada una puntuación.
A2. Ubicación	Localización del acceso al servicio ‘a la carta’ en la página inicial del sitio web de la televisión. Puntuación:

	3. Ubicación en el menú principal. 2. Ubicación en menús secundarios. 1. Ubicación en apartados menos visibles. 0. No presencia en la página inicial del web.
A3. Información del archivo audiovisual	Información general sobre la cobertura y características generales, tecnología empleada... Puntuación: 3. Información exhaustiva sobre el contenido y el funcionamiento. 2. Información básica sobre el contenido y funcionamiento. 1. Información mínima sobre el contenido. 0. No se proporciona ningún tipo de información.
A4. Coste	Acceso gratuito o de pago al servicio 'a la carta'. Este indicador es únicamente informativo y, por este motivo, no tiene asignada una puntuación.
A5. Servicios personalizados	Posibilidad de registrarse y de crear un perfil de usuario para obtener servicios personalizados. 1. Permite crear un perfil de usuario para obtener servicios personalizados. 0. Sin posibilidad de crear un perfil de usuario.
B. CONTENIDO	
Indicador	Descripción
B1. Géneros	Géneros televisivos que cubre el servicio 'a la carta'. Para este indicador se establecerán los siguientes grandes grupos: informativos, deportes, programas divulgativos y de actualidad, programas de entretenimiento, programas musicales, programas infantiles y ficción. Puntuación: 3. Todos o gran parte de los géneros emitidos. 2. Informativos, deportes y programas de actualidad. 1. Únicamente informativos.
B2. Origen	Tipo de producción (propia, ajena o coproducción) de los contenidos ofrecidos en línea. Puntuación: 3. Tres tipos de producción. 2. Producción propia y coproducciones. 1. Producción propia.
B3. Cobertura temporal	Alcance temporal que cubre el servicio 'a la carta'. Puntuación: 3. Se ofrece en línea la totalidad del archivo. 2. Se ofrece una cobertura temporal superior a 5 años, aunque no sea completa. 1. Se ofrece una cobertura temporal inferior a 5 años.
B4. Actualización	Tiempo transcurrido entre la emisión del contenido y su disponibilidad en línea. Puntuación: 3. Contenidos disponibles en línea el mismo día de emisión. 2. Contenidos disponibles en línea el día posterior a la emisión.

	1. Contenidos disponibles en línea dos o tres días posteriores a la emisión de los contenidos.
B5. Televisión en directo	Posibilidad de acceder a la emisión en directo de los diferentes canales de la cadena de televisión. Puntuación: 1. Televisión en directo. 0. Sin televisión en directo.
B6. Publicidad	Inclusión de anuncios durante el visionado. Puntuación: 2. Sin anuncios. 1. Con anuncios que se pueden saltar. 0. Con anuncios que se han de visualizar obligatoriamente.
C. SISTEMA DE CONSULTA	
Indicador	Descripción
C1. Tipos de consulta	Opciones de búsqueda que ofrece el servicio ‘a la carta’. Puntuación: 3. Búsqueda simple, avanzada y profesional. 2. Búsqueda simple y avanzada. 1. Únicamente búsqueda simple. 0. No se ofrece búsqueda por palabra clave.
C2. Lenguaje de interrogación	Posibilidad de realizar búsquedas utilizando operadores booleanos, de proximidad, frases literales, truncamiento, enmascaramiento... Puntuación: 3. Sistemas que admiten más de tres operadores. 2. Sistemas que admiten tres operadores. 1. Sistemas que admiten uno o dos operadores. 0. Sistemas que no admiten lenguaje de interrogación.
C3. Recuperación por navegación	Posibilidad de acceder a los documentos mediante la navegación alfabética, por fecha de emisión, por categorías o por rankings (los más vistos, destacados...) Puntuación: 3. Más de tres índices y rankings. 2. Dos o tres índices y rankings. 1. Un único índice o ranking. 0. Ningún índice o ranking.
C4. Opciones de limitación o filtro	Posibilidad de restringir la búsqueda por fecha de emisión, programa, género televisivo ... Puntuación: 3. Sistemas que ofrecen tres o más opciones de limitación. 2. Sistemas que disponen de dos tipos de limitación. 1. Sistemas que únicamente permiten un tipo de limitación. 0. Sistemas sin ningún tipo de limitación.
C5. Ayuda	Existencia de textos explicativos sobre el funcionamiento del sistema y la realización de las búsquedas. Puntuación: 1. Existencia de ayuda. 0. Inexistencia de ayuda.

D. PÁGINA DE RESULTADOS	
Indicador	Descripción
D1. Gestión de los resultados	<p>Posibilidad de establecer el número de registros por página y el criterio de ordenación de los resultados (fecha de emisión, relevancia...).</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 3. Gestión completa de los registros. 2. Posibilidad de modificar únicamente el criterio de ordenación. 1. Posibilidad de modificar únicamente el número de registros por página. 0. Imposibilidad de gestionar los registros.
D2. Campos visualizados de cada registro	<p>Número de campos que se muestran de cada registro en la página de resultados.</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 3. Se muestran los campos básicos (título, programa y fecha de emisión), una breve sinopsis y un <i>frame</i> del contenido audiovisual. 2. Se muestran entre tres y cuatro campos o se ofrecen más, pero falta algún campo básico. 1. Se muestran dos o menos campos.
D3. Agrupación de los resultados por categorías	<p>Presentación de los resultados clasificados por categorías (informativos, deportes, entretenimiento, infantil...).</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 1. Agrupación de los resultados por categorías. 0. Listado de resultados sin ninguna agrupación.
D4. Identificación de los términos de búsqueda	<p>Aparición destacada de los términos buscados, de manera que el usuario pueda localizarlos rápidamente en cada resultado.</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 1. Términos de búsqueda resaltados. 0. Términos de búsqueda sin resaltar.
E. CARACTERÍSTICAS DEL AUDIOVISUAL	
Indicador	Descripción
E1. Información básica	<p>Se considera información básica del audiovisual, todo aquello que permite identificarlo: título y/o número de capítulo, programa, fecha de emisión y duración.</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 3. Se muestra la información básica al completo. 2. Se muestra el título, el programa y la fecha de emisión. 1. Se muestra el programa y la fecha de emisión.
E2. Descripción del contenido	<p>Conjunto de informaciones que permiten conocer el contenido del audiovisual: sinopsis, transcripción del audio y etiquetas.</p> <p>Puntuación:</p> <ol style="list-style-type: none"> 2. Tiene los tres elementos.

	1. Tiene uno o dos elementos. 0. No se incluye descripción del contenido.
E3. Opciones de visionado	Opciones que ofrece el sistema para visionar el contenido audiovisual (pantalla completa, cambio de reproductor, subtítulo...). Puntuación: 3. Sistemas que ofrecen todas las opciones mencionadas. 2. Sistemas que ofrecen dos opciones de visionado. 1. Sistemas que sólo permiten pantalla completa.
E4. Calidad del visionado	Opción de visionado de vídeo en varios formatos o en alta calidad (HD). Puntuación: 1. Permite elegir la calidad de visionado. 0. No tiene ninguna opción.
E5. Opciones de gestión del vídeo	Posibilidad de marcar el vídeo como favorito, comentarlo, votarlo, suscribirse (RSS o <i>Podcast</i>), enviarlo por correo electrónico, compartirlo en redes sociales, insertarlo en una página web ... Puntuación: 3. Se ofrecen todas las opciones mencionadas. 2. Se ofrece un menor número de opciones. 1. Sólo se permite compartir el contenido en redes sociales. 0. No se permite gestionar el vídeo.
E6. Vídeos y páginas relacionadas	Posibilidad de recuperar vídeos relacionados y de acceder a la página web específica del programa al que pertenece el contenido. Puntuación: 3. Sistemas que enlazan el vídeo con la web del programa y permiten recuperar contenidos relacionados temáticamente. 2. Sistemas que enlazan el vídeo con la web del programa y permiten recuperar otros vídeos del mismo programa. 1. Sistemas que únicamente enlazan el vídeo con la web del programa u ofrecen contenidos relacionados. 0. Sistemas que no establecen ningún tipo de relación entre contenidos.

6 Resultados

A continuación se presentan los resultados del análisis de las 12 televisiones que integran la muestra agrupados según los indicadores detallados en el apartado anterior.

A. Aspectos generales

A1. Denominación

El término ‘a la carta’ es el usado de manera unánime por once de las doce televisiones de la muestra. Únicamente la Televisión Canaria utiliza otra denominación: ‘Multimedia’. Cabe destacar el caso de la Televisión de Murcia, que utiliza dos términos distintos para designar sus contenidos audiovisuales en línea: ‘a la carta’ para contenidos actuales (2015-) y ‘archivo’ para emisiones anteriores (2009-2015).

A2. Ubicación

Las televisiones cada vez dan más importancia a los contenidos audiovisuales en línea y prueba de ello es que sitúan el acceso a los servicios ‘a la carta’ en el menú principal de su sitio web. La Televisión Canaria es la única que sitúa el acceso en un menú secundario bajo la pestaña relativa a la sección de televisión, ya que se trata del sitio web de la corporación. Independientemente de ofrecer un acceso en el menú principal o secundario, la mayoría de las televisiones dedican un apartado de la página inicial de la cadena al servicio ‘a la carta’ donde acostumbran a aparecer los últimos contenidos disponibles en línea.

A3. Información del archivo audiovisual

Éste es uno de los puntos más débiles de las televisiones analizadas, ya que sólo dos de ellas ofrecen alguna información sobre el servicio ‘a la carta’. En el caso de la Televisión de Murcia, esta información está situada en la cabecera de la página del ‘archivo’. La explicación es mínima pero muy útil ya que la cadena usa dos denominaciones distintas (‘a la carta’ y ‘archivo’) y esta información permite al usuario conocer qué contenidos encontrará en cada sección. TVC proporciona también una información muy básica pero no es fácil para el telespectador encontrarla, ya que no está ubicada en las páginas del servicio ‘a la carta’. Si el usuario quiere conocer el funcionamiento y las características del archivo, debe dirigirse al apartado de atención a la audiencia.

A4. Coste

En ningún caso el acceso al archivo en línea requiere pago. La gratuidad del servicio permite justificar su misión como televisiones autonómicas públicas y ampliar la difusión de los contenidos y, por tanto, la audiencia.

A5. Servicios personalizados

Tres de las doce televisiones analizadas ofrecen a los usuarios la posibilidad de crear un perfil de usuario y obtener servicios personalizados. Se trata de un producto que claramente las cadenas autonómicas aún tienen que desarrollar e implantar. À Punt permite al usuario crear una lista de favoritos, mientras que Televisión de Murcia ofrece la posibilidad de guardar vídeos para visionar más tarde, comentar, etiquetar los contenidos, etc. TVC ofrece un servicio de valor añadido distinto. En lugar de ser el usuario el que interactúa con los contenidos, es el propio sistema que interactúa con el usuario, ofreciéndole contenidos que pueden interesarle en función de los vídeos que visiona. Televisión Canaria, por su parte, permite registrarse en el sitio web pero el registro no tiene ninguna relación con el servicio ‘Multimedia’. Es por ello que no obtiene ninguna puntuación en este indicador.

B. Contenido

B1. Géneros

Todas las televisiones alcanzan la máxima puntuación ofreciendo todos o la mayoría de géneros emitidos por la cadena. Este indicador está directamente relacionado con el tipo de producción. Muchos géneros como el cine o la ficción son de producción ajena y por lo tanto están sujetos a unos derechos que limitan o impiden su difusión en línea.

B2. Origen

Como se ha comentado anteriormente, la limitación de derechos que conlleva el tipo de producción obliga a la mayoría de televisiones a incluir únicamente contenidos de producción propia y asociada. Cinco de las cadenas analizadas, TVC, À Punt, Canal Sur, TVG y ETB, cuentan también con producción ajena en sus servicios ‘a la carta’. En muchos casos estos

contenidos se ponen a disposición del usuario por un tiempo limitado, habitualmente unos quince días, y posteriormente se retiran del archivo en línea.

B3. Cobertura temporal

Canal Sur, Telemadrid y CMM son las televisiones que alcanzan una puntuación más baja en este indicador, ofreciendo una cobertura temporal inferior a 5 años. De las tres, destaca negativamente Canal Sur, ya que su servicio ‘a la carta’ sólo incluye emisiones del último año. Las demás televisiones disponen de una cobertura temporal más amplia, superior a los 5 años. Algunas, como TVC o ETB, están incluyendo también contenidos retrospectivos de la cadena. À Punt destaca entre todas las demás por ser la única que ofrece en línea contenidos audiovisuales desde su primera emisión, el 10 de junio de 2018. Este hecho es excepcional, puesto que la cadena nació hace un año. Aunque se trata de la sucesora de Canal 9, es una televisión de nueva creación.

B4. Actualización

La actualización del servicio ‘a la carta’ es uno de los indicadores donde las televisiones analizadas obtienen mejor nota. Once de ellas alcanzan la puntuación máxima incorporando los contenidos audiovisuales el mismo día de su emisión. Televisión de Murcia tiene una actualización irregular, incorpora los contenidos el mismo día o el día posterior a su emisión. Por este motivo no alcanza la máxima puntuación.

B5. Televisión en directo

Los cambios de hábito de consumo de los telespectadores hace que todas las televisiones ofrezcan televisión en directo.

B6. Publicidad

De igual modo que en la televisión convencional, muchas cadenas incluyen anuncios publicitarios o promociones propias en los contenidos que se visualizan a través del servicio ‘a la carta’. En este aspecto, hay variedad de opciones. À Punt, Telemadrid, TPA, Televisión Canaria, Televisión de Murcia, CMM e IB3 permiten visionar los contenidos sin publicidad. Canal Sur, TVG y Aragón TV incluyen anuncios antes o durante el visionado del contenido pero se puede saltar transcurrido unos segundos. TVC y ETB, en cambio, incluyen publicidad que el usuario debe visionar obligatoriamente.

C. Sistema de consulta

C1. Tipos de consulta

En general, las televisiones alcanzan una puntuación muy baja en este indicador. Diez de ellas únicamente disponen de búsqueda simple y, en algunos casos, se trata de una búsqueda de *Google* genérica para todo el sitio web. Un hecho negativo para el usuario, que tiene más dificultad para distinguir los contenidos audiovisuales de las noticias u otras páginas del sitio web de la cadena. En este aspecto, es interesante destacar cuatro casos con casuísticas muy distintas:

- Televisión de Murcia cuenta con dos repositorios diferenciados y ofrece un sistema de búsqueda diferente para cada uno de ellos. ‘A la carta’ cuenta con búsqueda simple y ‘archivo’ con avanzada, pudiendo buscar por título y/o programa.
- TVG no ofrece búsqueda por palabra clave para todo el servicio ‘a la carta’, pero sí en la página propia de cada programa, donde permite buscar por título.

- Aragón TV cuenta con búsqueda simple y avanzada, pero ésta última no funciona. Por este motivo se ha contabilizado únicamente la búsqueda simple.
- Castilla-La Mancha Media obtiene un cero en este indicador. La cadena dispone de búsqueda simple para las noticias y las páginas de los programas pero no para el servicio 'a la carta'.

C2. Lenguaje de interrogación

El lenguaje de interrogación es otro de los puntos débiles de los sistemas de consulta de los servicios 'a la carta'. Ocho de las televisiones no admiten ningún lenguaje de interrogación o aplican automáticamente la búsqueda por frase exacta. TVC, Canal Sur, TPA y Telemadrid permiten al usuario realizar búsquedas utilizando operadores booleanos o comillas. La televisión madrileña es la más completa ya que admite un número mayor de operadores.

C3. Recuperación por navegación

Las televisiones han optado por dar prioridad a la recuperación por navegación como principal sistema de consulta de los contenidos audiovisuales. La mitad de las cadenas ofrecen más de tres índices (TVC, Canal Sur, TVG, ETB, Aragón TV, IB3), cuatro de ellas tienen entre dos y tres índices (À Punt, Telemadrid, TPA, CMM) y dos ofrecen solo un índice (Televisión Canaria y Televisión de Murcia). El ranking de 'lo último' y los índices de programas y categorías/géneros son los más utilizados. En menor medida, las televisiones analizadas cuentan con rankings de 'lo más visto', 'destacados' y el índice de colecciones. Es interesante destacar el caso de TVC y ETB que muestran los contenidos de los últimos días en formato parrilla de programación.

C4. Opciones de limitación o filtro

La simplicidad de los sistemas de recuperación se observa también en este indicador. De las doce televisiones, solo TVC, TVG, Televisión Canaria, Televisión de Murcia y IB3 permiten restringir mínimamente la búsqueda. Las cinco permiten limitar los resultados por fecha de emisión. Además, Televisión Canaria permite restringir la búsqueda por categoría y TVG limitar los resultados a programas completos. Aragón Televisión ofrece limitación por fecha y programa pero no funcionan. Por este motivo se le ha asignado un cero en este indicador.

C5. Ayuda

Ninguna de las televisiones analizadas dispone de ayuda para que el usuario pueda conocer los tipos de consulta del sistema y los lenguajes de interrogación que admite.

D. Página de resultados

D1. Gestión de los resultados

En general, las televisiones alcanzan una puntuación muy baja en este indicador. Solo dos de ellas permiten gestionar los resultados: Canal Sur y Telemadrid. En ambos casos, el usuario puede ordenar los resultados por fecha o relevancia/coincidencia.

D2. Campos visualizados de cada registro

Las cadenas no obtienen una mala puntuación en este indicador, pero muchas pueden mejorar ampliando la información mostrada de cada resultado. Cuatro de ellas (TVC, À Punt, Televisión de Murcia y IB3) muestran una información bastante completa de cada resultado, es decir, el título, el programa, la fecha de emisión, una breve sinopsis y un *frame*. TVG, Telemadrid, Aragón Televisión y CMM muestran la mayoría de campos básicos, pero no ofrecen la sinopsis o aparece el nombre del programa pero no el título (y al revés). ETB,

TPA, Televisión Canaria y Canal Sur también obtienen menor puntuación (entre uno y dos puntos). Las cuatro tienen en común que no muestran la fecha de todos los videos en la página de resultados. Se trata de un aspecto negativo, ya que la fecha es uno de los datos importantes para identificar los contenidos.

D3. Agrupación por categoría

Ninguna televisión ofrece los resultados agrupados por categoría (informativos, deportes, entretenimiento, infantil...). Se trata de un punto débil que se acentúa en las cadenas que cuentan con una búsqueda genérica para todo el sitio web. Los resultados se muestran listados sin ningún tipo de agrupación y es más difícil para el usuario encontrar aquello que busca.

D4. Identificación de los términos de búsqueda

En general, en la página de resultados no se resaltan los términos buscados. Se trata de una carencia de la mayoría de las televisiones analizadas, exceptuando Canal Sur y TPA. Este punto débil afecta directamente al usuario, que tiene más dificultades para valorar si los resultados responden a su búsqueda.

E. Características del audiovisual

E1. Información básica

La mayoría de las televisiones alcanzan la máxima puntuación en este indicador. De las doce analizadas, nueve de ellas muestran los datos básicos que identifican cada audiovisual: título y/o número de capítulo, programa, fecha de emisión y duración. Televisión Canaria y Televisión de Murcia ofrecen la fecha, el programa y la duración, pero no el título del vídeo (o no en todos los casos). Televisión de Murcia, además, cuenta con diferencias en sus dos repositorios en línea: ‘a la carta’ y ‘archivo’. El primero ofrece la información anteriormente mencionada. En cambio, en el ‘archivo’ solo se ofrecen los datos básicos del audiovisual en la página de resultados. Cuando el usuario accede al vídeo en cuestión, la información del contenido es inexistente.

E2. Descripción del contenido

Aunque ninguna cadena obtiene la máxima puntuación en este indicador, existe unanimidad entre ellas. Todas incluyen una sinopsis, excepto la Televisión Canaria, que no incluye descripción del contenido. Es interesante destacar los casos de TPA, Telemadrid y TVG, que complementan la sinopsis con otros elementos. TPA, en algunos casos, muestra la transcripción del audio; Telemadrid cuenta con etiquetas; y TVG ofrece una tabla de contenidos con el *time code* de cada tema para magazines, informativos y programas donde es difícil hacer una sinopsis.

E3. Opciones de visionado

Todas las televisiones ofrecen la pantalla completa como opción de visionado. Además, TVC, Televisión Canaria y TPA incorporan la opción de subtítulo. En el caso de la cadena canaria, el contenido se visualiza a través de la plataforma Youtube y el subtítulo se genera automáticamente. La incorporación de subtítulos es un aspecto importante que las televisiones, especialmente las públicas, deben desarrollar para la accesibilidad de los contenidos audiovisuales.

E4. Calidad del visionado

Las tres televisiones con más opciones de visionado son las únicas que ofrecen los vídeos en distintas calidades. Aunque solo TVC y TPA cuentan con alta calidad entre sus opciones.

Para evaluar este indicador es importante tener en cuenta que muchas autonómicas emiten en HD desde hace relativamente pocos años y aún no han desarrollado este punto en el servicio 'a la carta'.

E5. Opciones de gestión del vídeo

En cuanto a este aspecto, todas las televisiones permiten compartir el vídeo en las redes sociales. Un hecho que demuestra la importancia de las redes en la sociedad actual. Canal Sur, Telemadrid y CMM son las únicas que solo ofrecen esta opción. Las demás televisiones también permiten enviar el contenido por correo electrónico, insertarlo en una página web, marcarlo como favorito o votarlo. Televisión de Murcia e IB3, además de algunas de las opciones ya mencionadas, permiten descargar el vídeo en formato mp4. IB3 también permite al usuario compartir únicamente un fragmento del vídeo (marcando un *in* y un *out*). Televisión Canaria contempla la descarga de los vídeos pero la opción está inactiva.

E6. Vídeos y páginas relacionadas

En general, todas relacionan los vídeos con el programa y permiten recuperar otros contenidos del mismo. Telemadrid es la única que va más allá y también ofrece contenidos relacionados temáticamente, una opción muy interesante para el usuario.

Resultados globales

El resumen de los resultados globales se muestra en tabla 1.

Tabla 1. Resultados

	TVC	Canal Sur	À Punt	TVG	ETB	Telemadrid	Aragón TV	TPA	TV Canaria	7TV Murcia	CMM	IB3	Total / puntuación máxima
A. Aspectos generales													
A1. Denominación	--	--	--	--	--	--	--	--	--	--	--	--	--
A2. Ubicación	3	3	3	3	3	3	3	3	2	3	3	3	35/36
A3. Información del archivo	1	0	0	0	0	0	0	0	0	1	0	0	2/36
A4. Coste	--	--	--	--	--	--	--	--	--	--	--	--	--
A5. Servicios personalizados	1	0	1	0	0	0	0	0	0	1	0	0	3/12
B. Contenido													
B1. Géneros	3	3	3	3	3	3	3	3	3	3	3	3	36/36
B2. Origen	3	3	3	3	3	2	2	2	2	2	2	2	29/36
B3. Cobertura temporal	2	1	3	2	2	1	2	2	2	2	1	2	22/36
B4. Actualización	3	3	3	3	3	3	3	3	3	2	3	3	35/36
B5. Televisión en directo	1	1	1	1	1	1	1	1	1	1	1	1	12/12
B6. Publicidad	0	1	2	1	0	2	1	2	2	2	2	2	17/24
C. Sistema de consulta													
C1. Tipos de consulta	1	1	1	1	1	1	1	1	1	2	0	1	12/36
C2. Lenguaje de interrogación	2	2	1	1	0	3	1	2	1	1	0	0	14/36
C3. Recuperación por navegación	3	3	2	3	3	2	3	2	1	1	2	3	28/36
C4. Opciones de limitación/filtro	1	0	0	2	0	0	0	0	2	1	0	1	7/36
C5. Ayuda	0	0	0	0	0	0	0	0	0	0	0	0	0/12
D. Página de resultados													
D1. Gestión de los resultados	0	2	0	0	0	2	0	0	0	0	0	0	4/36
D2. Campos de cada registro	3	1	3	2	2	2	2	2	1	3	2	3	26/36
D3. Agrupación por categorías	0	0	0	0	0	0	0	0	0	0	0	0	0/12
D4. Ident. términos búsqueda	0	1	0	0	0	0	0	1	1	0	0	0	3/12
E. Características del audiovisual													
E1. Información básica	3	3	3	3	3	3	3	3	2	1	3	3	33/36
E2. Descripción del contenido	1	1	1	1	1	1	1	1	0	1	1	1	11/24
E3. Opciones de visionado	2	1	1	1	1	1	1	2	2	1	1	1	15/36
E4. Calidad del visionado	1	0	0	0	0	0	0	1	1	0	0	0	3/12
E5. Opciones de gestión del vídeo	2	1	2	2	2	1	2	2	2	2	1	2	21/36
E6. Vídeos y páginas relacionadas	2	2	2	2	2	3	2	2	2	2	2	2	25/36
Total / puntuación total máxima	38/58	33/58	35/58	34/58	30/58	34/58	31/58	35/58	31/58	32/58	27/58	33/58	

7 Discusión

Los resultados recogidos en la tabla anterior, fruto del análisis y la asignación de puntuaciones, permiten establecer un ranking de televisiones. Algunas de ellas han obtenido la misma puntuación total y por este motivo aparecen en la misma posición:

1. Televisió de Catalunya (TVC)
2. À Punt y Televisión del Principado de Asturias (TPA)
3. Televisión de Galicia (TVG) y Telemadrid
4. Canal Sur y IB3
5. Televisión de la Región de Murcia
6. Aragón Televisión y Televisión Canaria
7. Euskal Telebista (ETB)
8. Castilla-La Mancha Media (CMM)

Televisió de Catalunya (TVC) encabeza la lista con buenas puntuaciones en cuanto a contenido, características del audiovisual y recuperación por navegación. Además es una de las únicas en ofrecer servicios personalizados. La cadena podría despuntar más con un sistema de búsqueda y gestión de los resultados más completos.

En segunda posición se encuentran À Punt y la Televisión del Principado de Asturias (TPA). La televisión valenciana, al ser de nueva creación, tiene como principal punto fuerte el contenido ofrecido, mientras que la televisión asturiana obtiene buenas puntuaciones en las características del audiovisual, ofreciendo distintas opciones de visionado, calidad de vídeo y descripción del contenido bastante completo.

Televisión de Galicia y Telemadrid ocupan el tercer lugar. La televisión gallega alcanza puntuaciones altas en cuanto al contenido, la recuperación por navegación y las opciones de limitación/filtro. Telemadrid, en cambio, tiene poca cobertura temporal y no ofrece contenidos de producción ajena, pero tiene como puntos fuertes el lenguaje de interrogación, la gestión de los resultados y los vídeos/páginas relacionadas.

En la parte central de la clasificación, en cuarto y quinto lugar, se encuentran Canal Sur, IB3 y Televisión de la Región de Murcia. La televisión andaluza ofrece contenidos de distinta producción, buena recuperación por navegación, lenguaje de interrogación y gestión de los resultados, pero su cobertura temporal y las opciones de gestión del vídeo son muy pobres. IB3 y Televisión de Murcia, por su parte, obtienen resultados medios en buena parte de los indicadores pero tienen en común que permiten descargar los vídeos, un aspecto que solo hemos podido observar en estas televisiones. Otro aspecto característico de la cadena murciana es que cuenta con dos repositorios distintos para contenidos actuales y archivo.

Aragón Televisión y Televisión Canaria ocupan el sexto lugar. La televisión aragonesa podría alcanzar mayor puntuación, ya que dispone de sistema de búsqueda avanzada y opciones de limitación pero no funcionan. Así que los usuarios deben conformarse con usar la búsqueda simple. La Televisión Canaria es la única que

denomina ‘Multimedia’ al servicio ‘a la carta’ y además no lo ubica en un menú principal. También tiene peculiaridades en cuanto a las características del audiovisual: es la única que cuenta con un reproductor de la plataforma Youtube para visionar los vídeos y no ofrece descripción de los contenidos.

En las dos últimas posiciones se encuentran Euskal Telebista (ETB) y Castilla-La Mancha Media (CMM). Los puntos débiles de las dos televisiones son el sistema de consulta y la gestión de los resultados. ETB cuenta con puntuaciones altas en cuanto al contenido, pero el sistema de consulta no admite lenguaje de interrogación ni permite gestionar los resultados. CMM, por su parte, tiene una cobertura temporal baja y no tiene un sistema de búsqueda para el servicio ‘a la carta’.

De estas doce televisiones de la FORTA analizadas, seis de los servicios ‘a la carta’ ya se habían estudiado en un trabajo anterior de los autores (Anton, Guallar, 2014) con datos recopilados en 2012. La evaluación de todas ellas, siete años después, permite observar la evolución de los servicios ‘a la carta’: cambios positivos y negativos, novedades implantadas, carencias que persisten... Concretamente se trata de TVC, Canal Sur, TVG, ETB, Telemadrid y Canal 9. Éste último caso es especial, puesto que Canal 9 cerró en 2013 y en 2018 reabrió como À Punt, una cadena nueva. Para tener una visión global, es interesante comparar el ranking de televisiones resultante de la asignación de puntuaciones en 2012 y 2019:

Resultados 2012	Resultados 2019
1. Televisió de Catalunya (TVC)	1. Televisió de Catalunya (TVC)
2. Canal Sur	2. À Punt
3. Euskal Telebista (ETB)	3. Telemadrid Televisión de Galicia (TVG)
4. Televisión de Galicia (TVG)	4. Canal Sur
5. Telemadrid Canal 9	5. Euskal Telebista (ETB)

Televisió de Catalunya sigue en cabeza, pero las otras televisiones han acortado distancias y variado posiciones. La televisión catalana ya obtenía buena puntuación en muchos de los indicadores y en estos años ha experimentado muy pocos cambios. El aspecto más remarcable es la pérdida de información proporcionada al usuario. Si en 2012 destacaba por ofrecer ayuda sobre el sistema de búsqueda y preguntas más frecuentes sobre el servicio ‘a la carta’, actualmente solo se ofrece una mínima explicación sobre el archivo.

À Punt se sitúa en segunda posición, al contrario que Canal 9, que ocupaba el último lugar. La nueva televisión valenciana ha incorporado elementos que Canal 9 no ofrecía, como las opciones de gestión del vídeo o páginas y vídeos relacionados. Pero aún tiene aspectos que mejorar, ya que respecto a la cadena anterior ofrece un sistema de consulta y gestión de los resultados más pobre.

Telemadrid es la cadena que ha experimentado mayor ascenso. De ocupar la última posición ha pasado al tercer lugar. La televisión madrileña ha sufrido un cambio positivo en cuanto al sistema de consulta. En 2012 únicamente permitía la recuperación de los contenidos mediante índices o rankings, mientras que actualmente ha incorporado una búsqueda simple que admite distintos lenguajes de interrogación y con opciones de gestión de los resultados. Además, es la única que ofrece vídeos relacionados temáticamente con el contenido que el usuario está visionando.

Televisión de Galicia ha escalado una posición. La cadena ha experimentado cambios en el sistema de búsqueda. En 2012 el servicio ‘a la carta’ disponía de búsqueda simple, mientras que ahora la búsqueda no es genérica para todo el repositorio, sino que se limita a un programa en concreto. Aún así, la televisión gallega ha mejorado la ubicación del acceso al archivo, ofrece más tipos de producción y dispone de más opciones de navegación y limitación/filtro.

Canal Sur baja dos posiciones en el nuevo ranking. La televisión andaluza era la más potente en cuanto a sistema de consulta, ya que ofrecía búsqueda simple y completa. Actualmente solo dispone de una búsqueda simple de *Google* genérica para todo el sitio web y también ha perdido opciones de limitación/filtro. Aún así, ha mejorado la ubicación del acceso al archivo, admite más lenguajes de interrogación y es la única de las seis que ha incorporado la identificación de los términos de búsqueda.

En último lugar se sitúa Euskal Telebista. La televisión vasca apenas ha experimentado cambios en su servicio ‘a la carta’, se mantiene casi igual que en 2012, pero las otras cadenas han incorporado mejoras y por ello ETB ha descendido dos posiciones.

Los resultados de esta comparativa se pueden observar la tabla 2.

Tabla 2. Resultados de la comparativa del análisis de las televisiones en 2012 y 2019

	TVC 2012	TVC 2019	Canal Sur 2012	Canal Sur 2019	Canal 9 2012	À Punt 2019	TVG 2012	TVG 2019	ETB 2012	ETB 2019	Telemadrid 2012	Telemadrid 2019	Total / puntuación máx. 2012	Total / puntuación máx. 2019
A. Aspectos generales														
A1. Denominación	---	---	---	---	---	---	---	---	---	---	---	---	---	---
A2. Ubicación	3	3	2	3	2	3	2	3	2	3	3	3	14/18	15/18
A3. Información del archivo	3	1	0	0	0	0	0	0	0	0	0	0	3/18	1/18
A4. Coste	---	---	---	---	---	---	---	---	---	---	---	---	---	---
A5. Servicios personalizados	---	1	---	0	---	1	---	0	---	0	---	0	---	2/6
B. Contenido														
B1. Géneros	3	3	3	3	3	3	3	3	3	3	3	3	18/18	18/18
B2. Origen	3	3	3	3	3	3	2	3	3	3	2	2	16/18	17/18
B3. Cobertura temporal	2	2	1	1	1	3	2	2	2	2	1	1	9/18	11/18
B4. Actualización	3	3	3	3	3	3	3	3	3	3	3	3	18/18	18/18
B5. Televisión en directo	---	1	---	1	---	1	---	1	---	1	---	1	---	6/6
B6. Publicidad	---	0	---	1	---	2	---	1	---	0	---	2	---	6/12
C. Sistema de consulta														
C1. Tipos de consulta	1	1	2	1	1	1	1	1	1	1	0	1	6/18	6/18
C2. Lenguaje de interrogación	3	2	1	2	1	1	1	1	1	0	3	2	7/18	9/18
C3. Recuperación por navegación	3	3	3	3	2	2	2	3	3	3	3	2	16/18	16/18
C4. Opciones de limitación/filtro	0	1	2	0	1	0	1	2	0	0	0	0	4/18	3/18
C5. Ayuda	1	0	0	0	0	0	0	0	0	0	0	0	1/6	0/6
D. Páginas de resultados														
D1. Gestión de los resultados	0	0	2	2	2	0	0	0	0	0	0	2	4/18	4/18
D2. Campos de cada registro	3	3	2	1	3	3	2	2	2	2	2	2	14/18	13/18
D3. Agrupación por categorías	0	0	0	0	0	0	0	0	0	0	0	0	0/6	0/6
D4. Ident. términos relacionados	0	0	0	1	0	0	0	0	0	0	0	0	0/6	1/6
E. Características audiovisuales														
E1. Información básica	3	3	3	3	2	3	3	3	2	3	3	3	16/18	18/18
E2. Descripción del contenido	2	1	2	1	2	1	2	1	2	1	2	1	12/18	6/12
E3. Opciones de visionado	3	2	1	1	1	1	1	1	2	1	1	1	9/18	7/18
E4. Calidad del visionado	---	1	---	0	---	0	---	0	---	0	---	0	---	1/6
E5. Opciones de gestión del vídeo	3	2	2	1	0	2	2	2	2	2	2	1	11/18	10/18
E6. Vídeos y páginas relacionadas	2	2	2	2	0	2	2	2	2	2	2	3	10/18	13/18
Total / puntuación total máxima	41/54	38/58	34/54	33/58	27/54	35/58	29/54	34/58	30/54	30/58	27/54	34/58		

8 Conclusiones

Los resultados del presente estudio muestran que los servicios ‘a la carta’ de las televisiones autonómicas aún tienen un nivel de desarrollo medio.

De los 58 puntos que las televisiones pueden alcanzar en el total de indicadores, las cadenas analizadas consiguen entre 27 y 38 puntos. Ninguna de ellas, por lo tanto, se acerca al máximo. Televisió de Catalunya (TVC), À Punt, la Televisión del Principado de Asturias (TPA), Televisión de Galicia (TVG) y Telemadrid ocupan las tres primeras posiciones del ranking. En la parte central encontramos Canal Sur, IB3, Televisión de la Región de Murcia, Aragón Televisión y Televisión Canaria. Cerrando la lista, en las dos últimas posiciones, Euskal Telebista (ETB) y Castilla-La Mancha Media (CMM).

Es indudable la importancia que las televisiones dan a su repositorio, ubicando su acceso en un lugar prioritario del sitio web de la cadena y, en muchos casos, dedicándole una sección propia en la página principal. Pero, en general, las televisiones se centran en el contenido y dejan de lado como consultarlo. Se da por hecho que el usuario tendrá interés por los últimos contenidos, a los que podrá acceder fácilmente mediante los variados índices o rankings que ofrecen. Es por ello que los sistemas de búsqueda y las páginas de resultados son tan básicos. Para complementar la recuperación por navegación, las televisiones se limitan a ofrecer una búsqueda simple, sin opciones de limitación, ni gestión de los resultados, y tampoco identificación de los términos buscados, un elemento indispensable de cualquier motor de búsqueda. Estas carencias se han convertido en un problema enquistado de los servicios ‘a la carta’. Hace siete años las televisiones analizadas presentaban estos mismos puntos débiles y solo en un caso se ha observado una mejora.

Como ya se ha comentado, los indicadores relativos al contenido son los que obtienen puntuaciones más altas (actualización, televisión en directo, géneros cubiertos). Aún así, sería interesante poner a disposición de los usuarios contenidos de producción ajena, aunque sea por un tiempo limitado a causa de los derechos, y ampliar la cobertura temporal. El hecho de incorporar contenidos retrospectivos permitiría dar mayor difusión y revalorizar el fondo documental. En este sentido, cabe tener en cuenta que cuanto más contenido se incorpore al repositorio en línea, más aumenta la necesidad de contar con un buen motor de búsqueda. En cuanto a los nuevos indicadores incorporados al estudio, dos de ellos demuestran que los servicios ‘a la carta’ aún tienen mucho camino por recorrer. Se trata de ‘servicios personalizados’ y ‘calidad del visionado’. Dos aspectos muy potentes que solo tres televisiones tienen mínimamente desarrollados.

Por lo tanto, tras el análisis, podemos afirmar que los servicios ‘a la carta’ de las televisiones autonómicas tienen un nivel de desarrollo medio con puntos fuertes y débiles generalizados.

Por último, cabe señalar algunas líneas de posibles investigaciones futuras en relación con este trabajo. Por una parte, sería conveniente aplicar esta metodología a

muestras de análisis diferentes, como por ejemplo otras televisiones de España como pueden ser la televisión pública estatal y las televisiones privadas, o a televisiones de otros países, y en estos casos, realizar estudios comparativos entre las mismas. Asimismo, sería también interesante realizar estudios cualitativos para conocer la opinión de los responsables de estos servicios acerca del funcionamiento de los mismos, así como estudiar cual es la participación de los departamentos de Archivo y Documentación de las empresas televisivas en estos servicios.

Referencias

- Anton, L. y Guallar, J. (2014). Análisis de los archivos audiovisuales en internet de las televisiones autonómicas españolas. *Revista Española de Documentación Científica*, 37(1), e033. Recuperado de <https://doi.org/10.3989/redc.2014.1.1044>
- Caldera Serrano, J. y Arranz-Escacha, P. (2012). *Documentación audiovisual en televisión*. Barcelona: Editorial UOC.
- Cascón-Katchadourian, J., Ruiz-Rodríguez, A. A. y Alberich-Pascual, J. (2018). Revisión, análisis y evaluación de sistemas para la gestión de activos multimedia en organizaciones. *Revista Española de Documentación Científica*, 41 (1). Recuperado de <https://doi.org/10.3989/redc.2018.1.1481>
- De Mon Martín, A. y Guallar, J. (2014). Análisis de las videotecas de las principales televisiones en España. *Cuadernos de Documentación Multimedia*, 25. Recuperado de <http://revistas.ucm.es/index.php/CDMU/article/view/47472>
- Díaz-Campo, J. (2014). Las cadenas de televisión españolas en Internet: un estudio sobre la calidad de sus sitios web. *Estudios sobre el Mensaje Periodístico*, 20, (1), 273-287. Recuperado de <http://revistas.ucm.es/index.php/ESMP/article/view/45219>
- Giménez-Rayó, M. (2012). La documentación audiovisual en televisión en el mundo 2.0: retos y oportunidades. *Trípodos*, 31, 79-97. Recuperado de http://www.tripodos.com/index.php/Facultat_Comunicacio_Blanquerna/article/view/39
- Giménez-Rayó, M. y Guallar, J. (2014). Centros de documentación en televisión y productos documentales. *El Profesional de la Información*, 23(1), 13-25. Recuperado de <http://dx.doi.org/10.3145/epi.2014.ene.02>
- Hidalgo Goyanes, P. (2013). Patrimonio audiovisual en televisión. En: Marcos Recio, Juan Carlos (Ed.). *Gestión del patrimonio audiovisual en medios de comunicación* (pp. 53-80). Madrid: Editorial Síntesis.
- Inarejos, L. y Guallar, J. (2015). Centros de documentación de televisiones en Catalunya: Estudio de BTV, RTVE, TVC y 8tv. *Cuadernos de Documentación Multimedia*, 26, 48-65. Recuperado de http://dx.doi.org/10.5209/rev_CDMU.2015.v26.50629
- Linares, J., Codina, L., Abadal, E. y Guallar, J. (2016). Periodismo en bases de datos y buscabilidad de la información: Protocolo de análisis. *Hipertext.net*, 14. Recuperado de <http://eprints.rclis.org/30468/>

- López-de-Quintana-Sáenz, E. (2014). Rasgos y trayectorias de la documentación audiovisual: logros, retos y quimeras. *El Profesional de la Información*, 23(1), 5-12. Recuperado de <http://dx.doi.org/10.3145/epi.2014.ene.01>
- Pedraza-Jiménez, R., Codina, Ll. y Guallar, J. (Ed.). (2016). *Calidad en sitios web: método de análisis general, e-commerce, imágenes, hemerotecas y turismo*. Barcelona: Editorial UOC.
- Perpinyà Morera, R. y Cid-Leal, P. (2018). Los portales de archivos españoles: transparencia, interoperabilidad y orientación a los usuarios. *Revista Española de Documentación Científica*, 41 (3). e212. Recuperado de <https://doi.org/10.3989/redc.2018.3.1507>

Tecnologías del habla: nuevas oportunidades para los archivos de televisión

Virginia Bazán-Gil¹, Eduardo Lleida², Carmen Pérez³, Manuel Gómez⁴ y Alberto de Prada⁵

¹ ORCID [0000-0003-4920-2212](https://orcid.org/0000-0003-4920-2212). Fondo Documental, RTVE, España.
virginia.bazan@rtve.es

² ORCID [0000-0001-9137-4013](https://orcid.org/0000-0001-9137-4013). Instituto de Investigación en Ingeniería de Aragón.
Universidad de Zaragoza, España
lleida@unizar.es

³ ORCID [0000-0001-5128-5905](https://orcid.org/0000-0001-5128-5905). Innovación y Estrategia Tecnológica, RTVE, España.
carmen.perez@rtve.es

⁴ ORCID [0000-0001-5128-5905](https://orcid.org/0000-0001-5128-5905). Área de Desarrollo Digital, RTVE, España.
manuel.gomez@rtve.es

⁵ ORCID [0000-0002-9140-5745](https://orcid.org/0000-0002-9140-5745). Fondo Documental, RTVE, España.
alberto.deprada@rtve.es

Resumen. La abundancia de contenidos audiovisuales y la dificultad creciente para identificar y describir esos contenidos de forma eficiente ha convertido la Inteligencia Artificial en un objeto de deseo para los archivos de televisión. Los futuros procesos de generación automática de metadatos en los archivos se fundamentarán en tres tecnologías complementarias: visión artificial, tecnologías del habla y procesamiento del lenguaje natural. La aplicación de estas tecnologías no solo facilitará el acceso a un volumen creciente de contenidos audiovisuales, sino que además permitirá alcanzar un nivel de detalle en el análisis hasta ahora impensable en los archivos de televisión. En este nuevo horizonte, las funciones esenciales de los documentalistas se verán una vez más alteradas, ya que los procesos automatizados requerirán de su implicación en las fases de entrenamiento de los algoritmos en los que se basan estas tecnologías, así como en el control de la calidad de los datos generados de forma automática. En este contexto surge Iberspeech Challenge 2018, un reto tecnológico impulsado por la Cátedra RTVE Universidad de Zaragoza, puso a disposición de la comunidad científica más de 500 horas de contenidos audiovisuales permitiendo a distintos grupos de investigación, nacionales e internacionales, poner a prueba sus algoritmos en tres tareas diferenciadas: transcripción de habla a texto (speech to text), diarización de hablantes y diarización multimodal. Los resultados obtenidos han demostrado las dificultades tecnológicas que todavía deben superarse tanto en la transcripción voz a texto como en la diarización, tanto de hablante como multimodal.

Palabras clave: Archivos de television, Inteligencia Artificial, Tecnologías del habla, Español.

Abstract. As the number of audiovisual contents to be identified and analysed has been increasing since the last few years, and as the resources available to face this situation are decreasing, Artificial Intelligence has become a desired tool for television archives. The future automatic metadata extraction workflows will be based in three complementary technologies: artificial vision, speech technologies and natural language processing. The use of these technologies will allow us to access a greater number of contents as well as to increase the granularity in the analysis. The role of the documentalist will be modified once again. Training algorithms and data validation will be two new relevant tasks for these professionals. In this new scenery, the RTVE University of Zaragoza Chair promoted the Iberspeech 2018 Challenge. This technological challenge made more than 500 hours of audiovisual content in Spanish available to the scientific community. Iberspeech 2018 also allowed the different national and international research groups to test their algorithms in three different tasks: speech to text, speaker diarization and multimodal diarization. The results obtained have shown the technological difficulties that still have to be overcome. These results should also be addressed from the user's perspective in order to answer questions connected to the degree of error tolerance in automatic transcription within three different areas: edition, broadcasting and archive.

Keywords: Television archives, Artificial Intelligence, Speech Technologies, Spanish

1 **Inteligencia Artificial para el análisis de contenidos audiovisuales**

La abundancia de contenidos audiovisuales y la dificultad creciente para identificar y describir estos contenidos de forma eficiente ha convertido a la Inteligencia Artificial en un objeto de deseo para los archivos de televisión (Bazán Gil & Guerrero Gómez-Olmedo, 2018). Los futuros procesos de generación automática de metadatos en los archivos se fundamentarán en tres tecnologías complementarias: visión artificial, tecnologías del habla y procesamiento del lenguaje natural. La visión artificial se ocupa del reconocimiento de imágenes, la agrupación y segmentación de escenas y el seguimiento de objetos y personas. Las tecnologías del habla, por su parte, permiten tanto la conversión voz a texto como un tratamiento de los hablantes y sus emociones, permitiendo su identificación y segmentación. De forma complementaria, el procesamiento del lenguaje natural permite el reconocimiento de entidades y temas tratados en un discurso y facilita la creación de resúmenes y descripciones aportando

así nuevos metadatos para el archivo (Lleida, 2018). En otras palabras, la visión artificial proporciona la descripción de la capa de imagen, especialmente relevante en los materiales originales en los que el sonido es únicamente ambiente; las tecnologías del habla permiten convertir la capa de audio en un texto que adquiere relevancia tanto en la fase de edición y montaje de un programa, como en las fases posteriores de emisión y difusión a través de la Web (subtitulado); y finalmente, el procesamiento del lenguaje natural contribuye a enriquecer con metadatos el material en la fase de archivo, generando puntos de acceso en forma de entidades para la recuperación.

2 Las tecnologías del habla y su aplicación en televisión

De las tecnologías mencionadas en el epígrafe anterior, las del habla, ya han sido adoptadas por las cadenas de televisión. Gracias a ellas es posible la generación de subtítulos para garantizar la accesibilidad de las personas con discapacidad auditiva a los contenidos audiovisuales. Más del 94% de los contenidos emitidos por RTVE están subtitulados. Los subtítulos, generados por el Área de Accesibilidad están disponibles tanto en la emisión en TDT como en la redifusión en la Web y, en un futuro muy próximo, se incorporarán al Archivo como un punto de acceso más a los contenidos, facilitando con ello las búsquedas por palabras y la recuperación de secuencias concretas dentro de un programa determinado.

Aún más, la transcripción de voz a texto puede realizarse sobre los rodajes originales que se utilizan en la edición de un programa de televisión, evitando a los redactores la tediosa labor de transcribir manualmente las entrevistas. En esta línea, el Área de Innovación y Estrategia Tecnológica y el Fondo Documental RTVE trabajan, junto al equipo del programa Crónicas, en un proyecto piloto para la transcripción automática de entrevistas con sistemas de transcripción de voz a texto. Este proyecto no sólo supondrá una optimización del tiempo dedicado a la producción del programa, sino que aumentará sustancialmente la accesibilidad de un material, los rodajes, que habitualmente cuentan con escasos metadatos asociados cuando llegan al archivo.

El uso de esta tecnología, tanto en la producción como en la emisión, garantiza la creación de una capa de información relevante para la preservación y reutilización de los contenidos. Es fundamental, no solo generar los datos, sino establecer además las sinergias entre áreas y los flujos de trabajo necesarios para que estos datos acompañen a los contenidos en su viaje hacia el archivo. En este sentido son relevantes ya que constituyen puntos de acceso a los contenidos preservados.

Como veremos con más detalle, las principales tecnologías implicadas en estos procesos son la transcripción de voz a texto (Speech to text) y la diarización o segmentación y agrupación por hablantes. Los sistemas de voz a texto nos permiten conocer qué se ha dicho mientras que la diarización nos ayudará a identificar quién habla, cuándo habla y qué aspecto tiene si aplicamos, además, técnicas de reconocimiento facial tal y como muestra la figura 1.

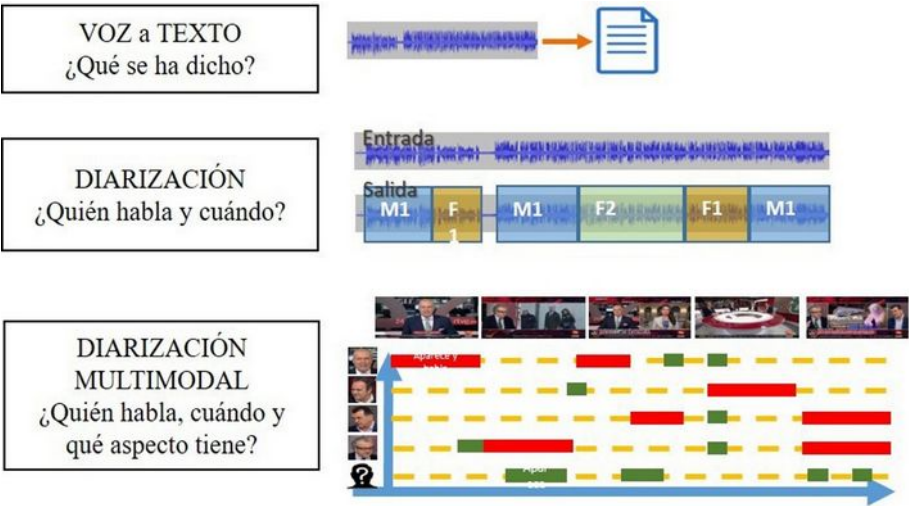


Figura 1. Tecnologías del habla aplicadas a archivos de televisión

3 Iberspeech 2018: el reto tecnológico de la Cátedra RTVE - Universidad de Zaragoza

En este contexto, en el que la Inteligencia Artificial brinda a los archivos de televisión nuevas oportunidades, la Corporación RTVE y la Universidad de Zaragoza formalizaron en julio de 2017 la creación de una Cátedra cuyo principal objetivo es la realización de actividades de formación, investigación, estudio y divulgación de las Tecnologías de la Información y de las Comunicaciones relacionadas con el Big Data y su aplicación al análisis de contenidos audiovisuales y sonoros (Cátedra RTVE Universidad de Zaragoza, 2017). En ella, junto al equipo de la Universidad de Zaragoza colaboran distintas áreas de RTVE como son: el Fondo Documental, el área de Innovación y Estrategia Tecnológica y el Área de Desarrollo Digital.

Como parte de sus actividades, y con el objetivo de impulsar la investigación en tecnologías del habla en español, en mayo de 2018 la Catedra lanzó el RTVE Iberspeech Challenge 2018 (RTVE, 2018). Un reto tecnológico que puso a disposición de la comunidad científica más de 500 horas de contenidos emitidos junto a sus correspondientes subtítulos, además de los subtítulos de los programas de producción propia emitidos por el Canal 24h a lo largo de 2017 (Lleida, Ortega, Miguel, Bazán, Pérez, Zotano, et al., 2018)

El conjunto de datos liberado, conocido como base de datos RTVE2018, incluye cerca de una veintena de programas de distintos géneros y temática, producidos y emitidos por RTVE entre 2015 y 2018. Estos programas presentan distintas dificultades desde el punto de vista de las tecnologías del habla como son: la

diversidad de acentos del español, la superposición de diálogos, el habla espontánea, la variabilidad acústica, el ruido de fondo o el vocabulario específico, factores todos ellos que inciden en el rendimiento de los sistemas de transcripción.

Sobre este conjunto de programas, los 22 grupos participantes debían detectar automáticamente, etiquetar y transcribir segmentos de habla. Para ello se definieron tres tareas distintas: Voz a texto, diarización y diarización multimodal. En las tareas de voz a texto y diarización los participantes podían optar por dos modalidades de participación: entrenamiento cerrado y abierto. El entrenamiento cerrado suponía entrenar el sistema únicamente con los datos proporcionados en la base RTVE2018, mientras que el entrenamiento abierto permitía utilizar cualquier otra base de datos documentada. En la tarea de diarización multimodal, los participantes contaban con imágenes y vídeos adicionales de cada una de las personas que debían reconocer.

3.1 Transcripción de voz a texto

En esta tarea los sistemas de reconocimiento debían transcribir cerca de 39 horas de emisión correspondientes a los programas: “Al filo de lo imposible” (AFI), “Arranca en Verde” (AV), “Dicho y hecho” (DH), “España en Comunidad” (EC), “La mañana” (LM), “La tarde en 24h. Tertulia” (LT24HTer), “Latinoamérica en 24H” (LA24H) y “Saber y ganar” (SG) (Lleida, Ortega, Miguel, Bazán, Pérez, Gómez, et al., 2018b). El rendimiento de los sistemas se calculó en función de la tasa de error por palabras (WER) comparando la transcripción automática con su transcripción real:

$$WER = \frac{N_{INS} + N_{BOR} + N_{SUS}}{N}$$

donde N_{INS} es el número de inserciones (palabras reconocidas que no se han pronunciado), N_{BOR} es el número de borrados (palabras no reconocidas pero que se han pronunciado), N_{SUS} es el número de sustituciones (palabras reconocidas equivocadamente) y N el número total de palabras a reconocer.

Esta tarea supone segmentar los documentos de audio en función de los hablantes y agrupar todos los segmentos correspondientes a un mismo hablante, sin conocer a priori ni el número de hablantes ni su identidad. Los participantes pusieron a prueba sus sistemas con aproximadamente 22 horas de emisión correspondientes a los programas “España en Comunidad”, “La mañana”, “La tarde en 24h. Tertulia” y “Latinoamérica en 24H” (Ortega et al., 2018). En este caso, el rendimiento de los sistemas se midió en función del Error de Diarización (DER) comparando la segmentación obtenida por los sistemas con una segmentación real supervisada. El DER contabiliza la fracción de tiempo que no se atribuye correctamente a un hablante específico:

$$DER = \frac{T_{PER} + T_{FA} + T_{HAB}}{T_{VOZ}}$$

donde T_{PER} es la cantidad de voz considerada como no voz (silencio, música, etc.), T_{FA} es la cantidad de no voz considerada como voz, T_{HAB} es la cantidad de voz asignada a un hablante equivocado y T_{VOZ} la cantidad total de voz a evaluar.

3.2 Diarización de hablantes

Esta tarea supone segmentar los documentos de audio en función de los hablantes y agrupar todos los segmentos correspondientes a un mismo hablante, sin conocer a priori ni el número de hablantes ni su identidad. Los participantes pusieron a prueba sus sistemas con aproximadamente 22 horas de emisión correspondientes a los programas “España en Comunidad”, “La mañana”, “La tarde en 24h. Tertulia” y “Latinoamerica en 24H”(Ortega et al., 2018). En este caso, el rendimiento de los sistemas se midió en función del Error de Diarización (DER) comparando la segmentación obtenida por los sistemas con una segmentación real supervisada. El DER contabiliza la fracción de tiempo que no se atribuye correctamente a un hablante específico:

$$DER = \frac{T_{PER} + T_{FA} + T_{HAB}}{T_{VOZ}}$$

donde T_{PER} es la cantidad de voz considerada como no voz (silencio, música, etc.), T_{FA} es la cantidad de no voz considerada como voz, T_{HAB} es la cantidad de voz asignada a un hablante equivocado y T_{VOZ} la cantidad total de voz a evaluar.

3.3 Diarización multimodal

La diarización multimodal permite segmentar contenidos audiovisuales en función de los hablantes y enlazar los segmentos a una misma voz y a una misma cara. En este caso los personajes (39) eran previamente conocidos. (Lleida, Ortega, Miguel, Bazán, Pérez, Gómez, et al., 2018a) Los sistemas se pusieron a prueba sobre 4 horas de emisión de los programas “La tarde en 24h. Tertulia” y “La mañana”, y su rendimiento se midió en función de la tasa de Error de Diarización (DER) en la segmentación de hablantes y de caras.

4 Resultados Iberspeech 2018

4.1 Transcripción de voz a texto

En esta tarea, y para la condición de entrenamiento abierto, participaron 7 equipos internacionales con 14 sistemas de reconocimiento, basados tanto en sistemas comerciales, como tecnologías propietarias (Jorge et al., 2018) o en tecnologías open source como Kaldi (Povey et al., 2011) o DeepSpeech 2 (Amodei et al., 2016). El conjunto de horas empleado para el entrenamiento de los sistemas osciló entre las 109 y las 3800 horas.

En la figura 2 se muestra la distribución de tasas de error por programas para los 14 sistemas evaluados (Iberspeech, 2018). Como era de esperar existe una gran dependencia del tipo de programa. Programas con entrevistas o conversaciones espontáneas y con audio en exteriores como “Dicho y Hecho”, “La Mañana” o “Arranca en Verde” son los programas que suponen el mayor reto con tasas de error muy por encima del 20% para la mayoría de los sistemas. El sistema mejor evaluado obtiene una tasa de error del 16,45% y el peor evaluado un 35,80%.

La figura 3 muestra la distribución de las tasas de error por programas para el sistema ganador. Cabe destacar que para una gran mayoría de programas mantiene unas tasas de error por debajo del 20% e incluso para un programa como “Latinoamérica en 24H”, con distintos acentos del español, la tasa de error se sitúa por debajo del 10%. Este sistema ha sido entrenado con 3.800 horas de voz transcrita de vídeos de distintas fuentes de español tanto peninsular como de América Latina.

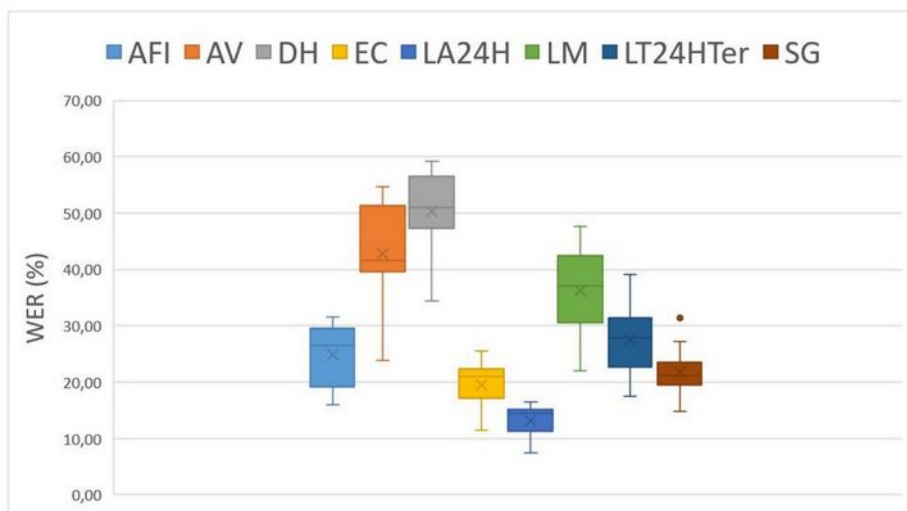


Figura 2. Distribución de la tasa de error WER por programas para el mejor sistema

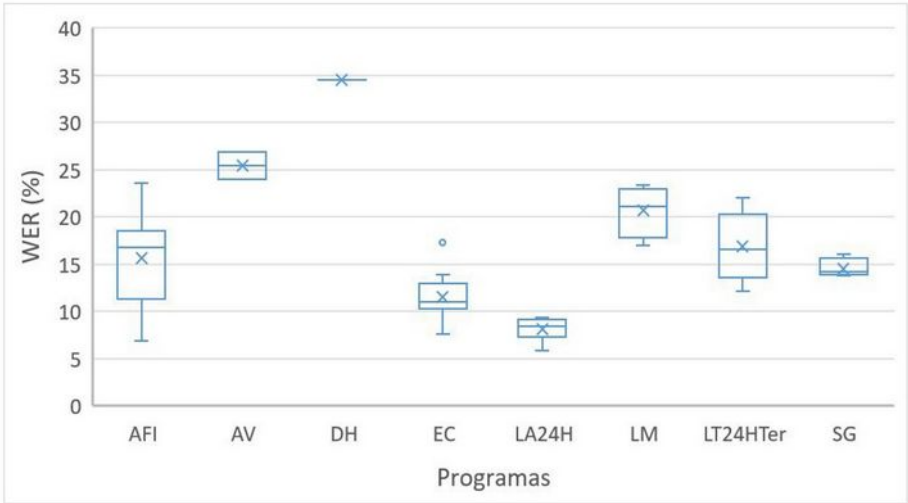


Figura 3. Distribución de la tasa de error WER por programas para el mejor sistema

En la condición de entrenamiento cerrado han participado 3 grupos con 6 sistemas. Las tasas de error conseguidas en esta modalidad se sitúan entre el 19,57% y el 26,66%. Se trataba de una condición más exigente ya que los sistemas contaban únicamente con los subtítulos y no con las transcripciones reales, hecho que ha incidido en el rendimiento al igual que el número limitado de horas de entrenamiento. Finalmente, dos equipos enviaron resultados incluyendo signos de puntuación (puntos y comas) con tasas de error incrementadas un 25% relativo.

4.2 Diarización de hablantes

En el reto de diarización de hablantes han participado 8 equipos internacionales, 6 bajo la condición cerrada y 4 bajo la abierta. Todos los sistemas son propietarios de cada equipo participante (Iberspeech, 2018). En la condición cerrada, donde únicamente se podían utilizar los datos proporcionados en la base de datos RTVE2018, las tasas DER se sitúan entre el 17,27% del mejor sistema y el 39,09% del sistema con peores prestaciones. En la condición abierta las tasas de DER se sitúan entre el 26% y el 31%.

Para el sistema con menor tasa DER, el 13,7% del DER se corresponde con la asignación incorrecta del hablante, es decir, un segmento de voz de un hablante se ha asignado a otro, el 1,1% se corresponde con hablantes no detectados y un 2,5% a segmentos sin hablantes que se han asignado a alguno de ellos. Examinado los errores en la detección del número de hablantes, encontramos que el mejor sistema tiene de

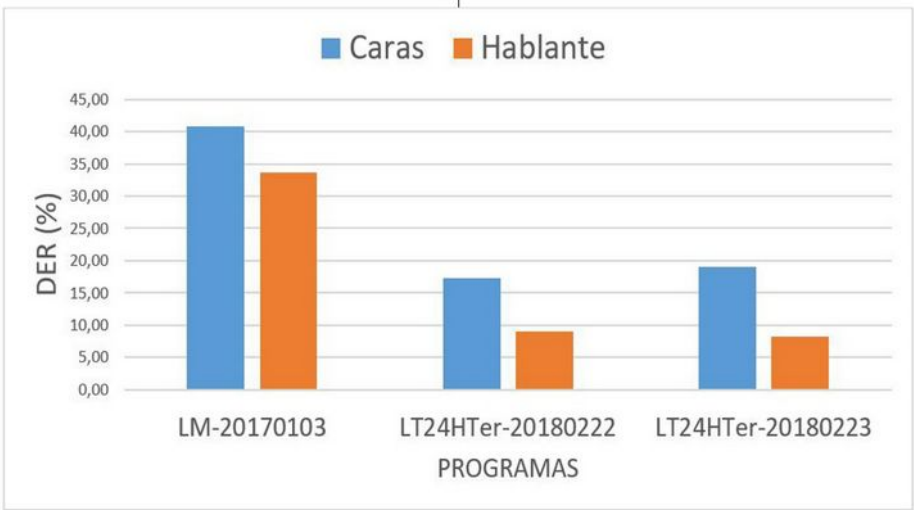


Figura 5. Mejores DER caras y hablante por programa

5 Conclusiones

Uno de los objetivos de la Cátedra RTVE de la Universidad de Zaragoza es la evaluación de tecnologías aplicables al análisis de contenidos audiovisuales. Con este propósito se ha organizado el primer reto a nivel internacional para evaluar sistemas de reconocimiento automático del habla, diarización de hablantes y multimodal para la lengua española en el contexto de programas de televisión de contenido diverso. Se ha conseguido una amplia participación, con un total de 22 de grupos de origen tanto nacional como internacional que han valorado muy positivamente la organización de un reto tecnológico como el propuesto. Los resultados muestran las dificultades tecnológicas que el reto suponía. Las tasas de error en transcripción de voz a texto varían, para el mejor sistema, entre un 6% y un 35% dependiendo del contenido del programa.

En relación a la diarización tanto de hablante como multimodal, los resultados obtenidos reflejan la dificultad en estimar el número de hablantes, la variabilidad en las prestaciones con el tipo de contenido y el mejor comportamiento de la diarización de hablante frente a la de cara.

Como resultado del esfuerzo de organización del reto, se pone a disposición de la comunidad científica una base de datos de contenido audiovisual con más de 550 horas de programación y con los protocolos de entrenamiento, desarrollo y evaluación para la comparativa de sistemas de reconocimiento automático del habla, diarización de hablantes y multimodal.

Estos resultados deben abordarse además desde la perspectiva del usuario, es decir desde la utilidad de las tecnologías para su integración en los flujos de trabajo de una organización como RTVE.

En los últimos 5 años los reconocedores de voz han mejorado de forma espectacular. Aun así, todavía existe una gran variabilidad en la tasa de error, entre un 5 % y un 50%, que limita su utilidad en todos los escenarios posibles: desde el material original hasta los programas emitidos. Los factores que inciden en el buen funcionamiento de los sistemas son múltiples, destacando el entorno acústico, la expresión oral del hablante y el contexto léxico-sintáctico. Tanto los resultados de Iberspeech, como las pruebas realizadas con el equipo del programa “Crónicas”, ponen de manifiesto que donde el oído humano no llega los reconocedores tampoco aciertan. Por ejemplo, las entrevistas en grupo con un único micrófono en las que predominan expresiones locales, muletillas, palabras incorrectas, errores fonéticos, sintácticos y léxicos constituyen un verdadero reto tanto para los reconocedores como para el oído humano.

En estos momentos en RTVE se están poniendo a prueba las tecnologías del habla en tres escenarios distintos: la producción, la emisión (tanto en Televisión como en la Web) y el Archivo. De estas pruebas se obtienen varias conclusiones:

- La importancia no solo de generar metadatos en función de las necesidades de las distintas áreas sino de establecer los mecanismos adecuados para garantizar la reutilización de estos datos en la cadena de producción – emisión – archivado.
- La correlación entre tipo de programa y resultado de la transcripción. En este sentido, y desde el punto de vista del archivo, es necesario comprender el funcionamiento de la tecnología y bajo qué premisas tiene un mayor rendimiento. Este conocimiento debe ponerse en relación con las necesidades reales de trabajo y con las expectativas de las distintas áreas en cuanto la calidad de los resultados.
- El grado de tolerancia respecto a los errores varía en función del escenario para el que se generan los metadatos. Tasas de error por encima del 15% pueden ser tolerables para el archivo cuando se carece de otros datos para la recuperación, pero no son admisibles desde el punto de vista de la emisión de los contenidos.

Para los archivos audiovisuales la incorporación de las tecnologías del habla a sus procesos diarios es más que una prioridad, pero la pregunta es ¿por dónde empezar? En un escenario ideal, en el que estas herramientas estén ya integradas en los flujos de trabajo, el Archivo recibirá el subtítulo generado por el Área de Accesibilidad lo que, junto con los metadatos generados tanto por la producción del programa como por el área de emisiones, proporcionará suficiente información para la recuperación y reutilización de los contenidos emitidos. En lo que al material original se refiere, el archivo recibirá del equipo de producción del programa, las transcripciones con las correcciones realizadas por los redactores.

Pero ¿qué hacer con el material original o emitido preservado hasta ahora con unos metadatos mínimos en el archivo? ¿Qué criterios debemos establecer para su tratamiento con sistemas de reconocimiento automático? Si usamos la tecnología indiscriminada corremos el riesgo de incurrir en un gasto en recursos económicos, tecnológicos y humanos sin la garantía de obtener unos resultados óptimos. No basta con disponer de la tecnología y usarla, es necesario definir primero cómo se va a aplicar. Pero ¿Qué criterios debemos emplear? ¿Los relacionados con su uso futuro como producción, reemisión o comercialización de los contenidos? ¿Aquellos relacionados con la posible fidelidad de la transcripción? En otras palabras ¿Es mejor tener información, aunque no sea exacta, que no tener ninguna?

Estas son las preguntas a las que debemos dar respuesta si queremos integrar de manera eficaz estas tecnologías en el Archivo.

Referencias

- Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... Zhu, Z. (2016). *Deep Speech 2: End-to-End Speech Recognition in English and Mandarin*. Retrieved from <http://proceedings.mlr.press/v48/amodei16.pdf>
- Bazán Gil, V. (2018). El renacimiento de los archivos: inteligencia artificial y semántica aplicada a la descripción de contenidos audiovisuales. Retrieved January 9, 2019, from https://es.slideshare.net/Artium_Vitoria/ix-encuentros-de-centros-de-documentacin-de-arte-contemporneo-en-artium-virginia-bazn-gil
- Bazán Gil, V., & Guerrero Gómez-Olmedo, R. (2018). Descripción automática de archivos audiovisuales: NeuralTalk, un modelo de video2text aplicado al archivo de RTVE Cita recomendada. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, (41). <https://doi.org/10.1344/BiD2018.41.7>
- Cátedra RTVE Universidad de Zaragoza. (2017). Cátedra RTVE de la Universidad de Zaragoza. Retrieved January 9, 2019, from <http://catedrartve.unizar.es/>
- Iberspeech. (2018). Iberspeech 2018. In *Iberspeech 2018*. Barcelona. Retrieved from https://www.isca-speech.org/archive/IberSPEECH_2018/
- Jorge, J., Martínez-Villaronga, A., Golik, P., Giménez, A., Albert Silvestre-Cerdà, J., Doetsch, P., ... Sanchis, A. (2018). MLLP-UPV and RWTH Aachen Spanish ASR Systems for the IberSpeech-RTVE 2018 Speech-to-Text Transcription Challenge. In *Iberspeech 2018* (pp. 257–261). Barcelona. <https://doi.org/10.21437/IberSPEECH.2018-54>
- Lleida, E. (2018). Tecnologías para el análisis y metadado de contenidos audiovisuales. Retrieved January 9, 2019, from http://www.rtve.es/contenidos/documentos/instituto/4_Jornada_Archivos_tv.pdf
- Lleida, E., Ortega, A., Miguel, A., Bazán, V., Pérez, C., Gómez, M., & De Prada, A. (2018a). *Albayzin Evaluation: IberSPEECH-RTVE 2018 Multimodal Diarization Challenge*.

- Lleida, E., Ortega, A., Miguel, A., Bazán, V., Pérez, C., Gómez, M., & De Prada, A. (2018b). *Albayzin Evaluation: IberSPEECH-RTVE 2018 Speech to Text Transcription Challenge*.
- Lleida, E., Ortega, A., Miguel, A., Bazán, V., Pérez, C., Zotano, M., & De Prada, A. (2018). *RTVE2018 Database Description*.
- Ortega, A., Viñals, I., Miguel, A., Lleida, E., Bazán, V., Pérez, C., ... De Prada, A. (2018). *Albayzin Evaluation: IberSPEECH-RTVE 2018 Speaker Diarization Challenge*.
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... Vesely, K. (2011). The kaldi speech recognition toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, 1–4. <https://doi.org/10.1017/CBO9781107415324.004>
- RTVE. (2018). La Cátedra RTVE en la Universidad de Zaragoza presenta su primer reto tecnológico a la comunidad científica - RTVE.es. Retrieved January 9, 2019, from <http://www.rtve.es/rtve/20180521/catedra-rtve-universidad-zaragoza-presenta-su-primer-reto-tecnologico-comunidad-cientifica/1737360.shtml>

Audio-Visual Semantics: propuesta de una ontología para la descripción de secuencias audiovisuales

Juan-Antonio Pastor-Sánchez¹, Tomás Saorín²,
Virginia Bazán³, Manuel Escribano⁴ y María-José Baños-Moreno⁵

¹ ORCID [0000-0002-1677-1059](https://orcid.org/0000-0002-1677-1059). Departamento de Información y Documentación, Universidad de Murcia, España
pastor@um.es

² ORCID [0000-0001-9448-0866](https://orcid.org/0000-0001-9448-0866). Departamento de Información y Documentación, Universidad de Murcia, España
tsp@um.es

³ ORCID [0000-0003-4920-2212](https://orcid.org/0000-0003-4920-2212). Radio Televisión Española, España
virginia.bazan@rtve.es

⁴ ORCID [0000-0003-2521-7030](https://orcid.org/0000-0003-2521-7030). VSN Video Stream Networks, España
mescribano@vsn.es

⁵ ORCID [0000-0001-9137-1330](https://orcid.org/0000-0001-9137-1330). ODILO, España
mbm41963@um.es

Resumen. El presente trabajo aborda la descripción de los aspectos conceptuales de contenidos audiovisuales mediante la ontología Audio-Visual Semantics (AVS) que permite representar acciones, características e interacciones entre entidades y/o elementos con una granularidad multinivel. Para ello se ha partido de una aproximación en el que las piezas audiovisuales están compuestas por secuencias. En dichas secuencias es posible identificar diferentes sucesos sobre los que pueden elaborarse descripciones. Los sucesos están compuestos por una serie de elementos tales como agentes, acciones, el objeto de dichas acciones u otros sucesos. Se contempla el uso de cualificadores (especificando su alcance y su valor) para definir cualidades o atributos de los diferentes elementos que intervienen en un suceso. Los agentes, acciones, objetos y cualificadores no se definen de forma extensiva en la propia ontología, sino que son referenciados como conceptos de vocabularios SKOS, lo que hace innecesario alterar la ontología AVS y permite definir relaciones semánticas entre conceptos y el etiquetado multilingüe. La ontología AVS se encuentra en proceso de desarrollo y en la actualidad se está procediendo a su validación mediante la descripción de piezas audiovisuales y la verificación de los resultados obtenidos.

Palabras clave: Contenidos Audiovisuales, Ontologías, Descripción semántica, SKOS

Abstract. This paper shows the description of the conceptual aspects of audiovisual content through the Audio-Visual Semantics ontology (AVS) that allows to represent actions, characteristics and interactions between entities and / or elements with a multilevel granularity. For this, it is considered that the audiovisual pieces are composed of sequences. In these sequences it is possible to identify different happenings that can be described. Happenings are composed of a series of elements such as agents, actions, targets or other events. The use of qualifiers (specifying their scope and value) is contemplated to define qualities or attributes of the different elements that intervene in an event. The agents, actions, objects and qualifiers are not defined extensively in the ontology itself, but are referenced as concepts of SKOS vocabularies, which avoids modifying the AVS ontology and allows to define semantic relations between concepts and multilingual labeling. Currently, the AVS ontology is in development and in the validation phase through the description of audiovisual pieces and the verification of the results obtained.

Keywords: Audiovisual Contents, Ontologies, Semantic Description, SKOS

1 Introducción

La tecnología digital actual ha facilitado en gran medida la creación y difusión de material audiovisual en múltiples ámbitos: entretenimiento, noticias, publicidad, educación, investigación, etc. Además, la mayor disponibilidad de estos contenidos en Internet ha aumentado su reutilización y redistribución. La producción y consumo de vídeo y podcast se ha generalizado, sin que sea necesario el uso de servicios o elementos tecnológicos complejos. Cualquier usuario puede grabar, editar y distribuir un documento audiovisual en Internet (YouTube, iVoox, redes sociales) utilizando únicamente un smartphone.

La descripción de escenas es un aspecto básico del trabajo diario de los documentalistas. Periodistas, editores y personal de archivo utilizan esta herramienta para una descripción detallada de todo el material audiovisual para preparar noticias, documentales, describir eventos deportivos, etc. Estas tareas hasta el momento suelen realizarse mediante la descripción de escenas utilizando texto libre. Igualmente es posible acceder a todo tipo de contenidos audiovisuales desde cualquier dispositivo, así como utilizar plataformas comerciales de *streaming* para su consumo.

Todo ello ha tenido un impacto significativo en la gestión y distribución de documentos audiovisuales por parte de los productores y operadores (Evain, Matton y Vaervagen, 2017), así como en los procesos de preservación digital (Plank, 2018; Corrado y Moulaison, 2016; Evens y Hautekeete, 2011). Igualmente, esta situación ha provocado que los sistemas de recomendación en las plataformas de contenidos audiovisuales en streaming estén cobrando cada día mayor importancia (Hallinan y Stri-phas, 2016).

La interoperabilidad en los procesos de gestión y acceso de documentos audiovisuales ha sido abordada por otros trabajos (Höffernig y Bailer, 2009; Dimoulas, Veglis y Kalliris, 2015). En este sentido, para acceder a un contenido audiovisual de manera eficiente se precisa un procesamiento previo de su semántica (Höffernig et al., 2011). Esta tarea no es sencilla, debido a la propia naturaleza de este tipo de documentos. Mientras que para las personas resulta relativamente sencillo describir y categorizar objetos, imágenes o sonidos en función de su significado, para las máquinas es algo ciertamente complejo. De hecho, los contenidos audiovisuales no son fácilmente procesables debido a los diferentes elementos que aparecen y la dinámica entre los mismos.

Son abundantes los trabajos que han abordado el uso de tecnologías semánticas en este ámbito, especialmente en aquellos aspectos relacionados con la edición, producción, distribución, programación y emisión de contenidos audiovisuales integrándose con MPEG-7 que permite una inclusión de aspectos semánticos básicos (Fourati, Jedidi y Gargouri, 2014; Hunter y Nack, 2001; Isaac y Troncy, 2004; Raymond et al., 2010). Por su parte, la European Broadcasting Union (EBU) han elaborado la ontología EBUCore que ofrece un conjunto de elementos de metadatos para la descripción de aspectos técnicos y estructurales¹ así como una serie de vocabularios SKOS que pueden utilizarse para la descripción de recursos².

El presente trabajo aborda la descripción de los aspectos conceptuales del contenido que trascienden la mera identificación de entidades o elementos contextuales tales como personas, organizaciones, lugares, etc. Esto permitiría plantear consultas y procesos que vayan más allá de la búsqueda y reutilización de contenidos audiovisuales únicamente a partir de un análisis y explotación de simples descripciones textuales. Para ello se propone la ontología Audio-Visual Semantics (AVS) que permite representar acciones, características e interacciones entre dichas entidades y/o elementos con una granularidad multi-nivel. Dicha propuesta es el resultado de la primera fase de un proyecto de Investigación y Desarrollo financiado Centro de Desarrollo Tecnológico Industrial, E.P.E., una entidad pública empresarial, dependiente del Ministerio de Ciencia, Innovación y Universidades. En dicho proyecto participan Video Stream Network (VSN), Radio Televisión Española (RTVE) y la Universidad de Murcia

1 Puede consultarse la ontología EBUCore en:

<https://www.ebu.ch/metadata/ontologies/ebucore/>

2 Dichos vocabularios están disponibles en: <https://www.ebu.ch/metadata/ontologies/skos/>

2 Metodología

El objetivo general del proyecto de investigación en el que se enmarca el desarrollo de la ontología AVS es el desarrollo de un software para que el proceso de descripción de contenidos audiovisuales pueda realizarse de forma rápida, en un contexto multilingüe, centrado en la semántica de los contenidos y con un incremento en la precisión de las tareas de búsqueda y recuperación de contenidos audiovisuales.

El sistema debe permitir la definición de declaraciones completas y semánticamente significativas con un alto grado de formalización y no vinculado a un idioma específico. Un requisito indispensable es el uso de estándares abiertos y ampliamente extendidos. La descripción de escenas debe ser flexible e incluso cercana al estructuras sintácticas sencillas y generales. En este sentido, sería posible identificar sujetos, acciones y objetos de dichas acciones con la suficiente expresividad semántica para modelar una amplia variedad de declaraciones.

El objetivo último del proyecto es dotar al Gestor de Contenidos Audiovisuales de VSN de la capacidad para catalogar y buscar contenido multimedia usando descripciones semánticas. Para ello se apoyará en AVS que define con precisión el dominio de las descripciones que se pueden construir para representar cualquier escena.

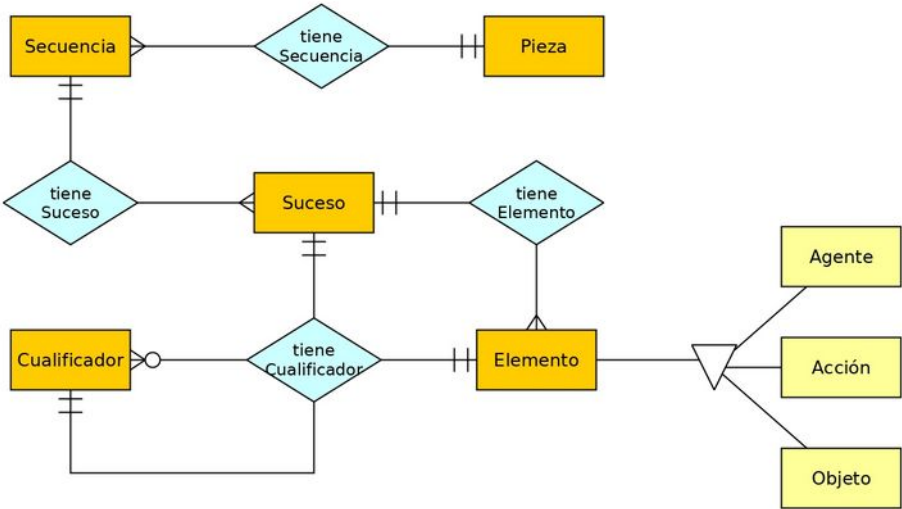


Figura 1. Modelo conceptual. Fuente: elaboración propia.

La metodología para la elaboración de la ontología AVS parte de un modelo conceptual en el que las piezas audiovisuales están compuestas por secuencias. En dichas secuencias es posible identificar y describir diferentes sucesos. Los sucesos están compuestos por una serie de elementos que pueden vincularse con 1) agentes que realizan una acción, 2) las acciones realizadas y 3) el objeto u objetos de dichas acciones.

También es posible que dichas vinculaciones sean con otros sucesos, por ejemplo, cuando una persona hace declaraciones sobre un hecho determinado, o cuando un suceso es causa de otro, etc. Para dotar de una mayor flexibilidad al modelo, es posible definir cualificadores sobre los sucesos, elementos y sobre los propios cualificadores. Este mecanismo de cualificación permite adecuar el nivel de detalle de las descripciones a las necesidades específicas de los contenidos o de los sistemas en los que se realizan.

Existe una gran diversidad de propuestas metodológicas para la elaboración de ontologías. Para la construcción de AVS se han seguido los principios metodológicos de Mendonça y Soares (2017) que se resumen en los siguientes pasos realizados:

1. Especificación de los objetivos y requisitos de la ontología.
2. Obtención del conocimiento mediante el análisis de las descripciones de minutos de piezas de contenidos audiovisuales y visualización de las mismas.
3. Conceptualización de la ontología a través de la identificación de los conceptos, relaciones y atributos.
4. Formalización e implementación de la ontología OWL especificando sus clases, propiedades y axiomas.
5. Evaluación y validación de la ontología mediante su aplicación en la descripción semántica de secuencias de piezas audiovisuales específicas.
6. Documentación de la ontología.

Cabe resaltar dos de los requisitos que se tuvieron en cuenta para el desarrollo de la ontología. El primero de ellos se relaciona con la necesidad de que la ontología debía ser de fácil uso en un entorno de uso a través de una herramienta de descripción minutada de los documentos audiovisuales. Es decir, el operador o documentalista debería poder utilizar esta ontología de un modo totalmente transparente y a través de una interfaz lo más sencilla posible. Lo anterior implicaba ofrecer una estructura de clases y propiedades lo más sencilla posible.

El segundo requisito tiene en cuenta los aspectos relacionados con el mantenimiento de la ontología. Considerando lo anterior se debía tener en cuenta que la identificación de nuevas entidades (personas, organizaciones, lugares) acciones, características, etc, podría darse durante el proceso de descripción. Por lo tanto, la incorporación de las mismas no debía modificar la ontología, sino que debía realizarse en una serie de vocabularios controlados. La gestión de los mismos se realizaría de forma dinámica, añadiendo nuevos elementos durante el proceso de descripción y con mecanismos de autocompletado para la búsqueda y selección de los mismos.

3 Resultados

AVS³ se basa en la definición de un conjunto relativamente reducido de elementos. Un punto a tener en cuenta en la definición de clases es la jerarquía “Clip → Secuencia → Suceso”, que establece la dinámica de descripción de un contenido audiovisual. Según este principio las piezas audiovisuales (avs:Clip) se dividen en secuencias o escenas (avs:Sequence) en las que se identifican eventos o sucesos que tienen lugar durante las mismas (avs:Happening). La ontología también incluye una serie de propiedades para la definición de intervalos que permitan localizar temporalmente las secuencias y los sucesos dentro de la piezas audiovisuales.

Por su parte, los sucesos se conciben como la combinación de elementos (avs:Element) que pueden ser agentes, acciones y objetos de las acciones. En vez de crear una jerarquía de clases a partir de los elementos, se han utilizado una serie de propiedades (avs:hasAgent, avs:hasAction y avs:hasTarget) que definen la naturaleza del vínculo entre un elemento y un concepto de un vocabulario SKOS, determinando de esta forma el tipo de elemento que interviene en el suceso.

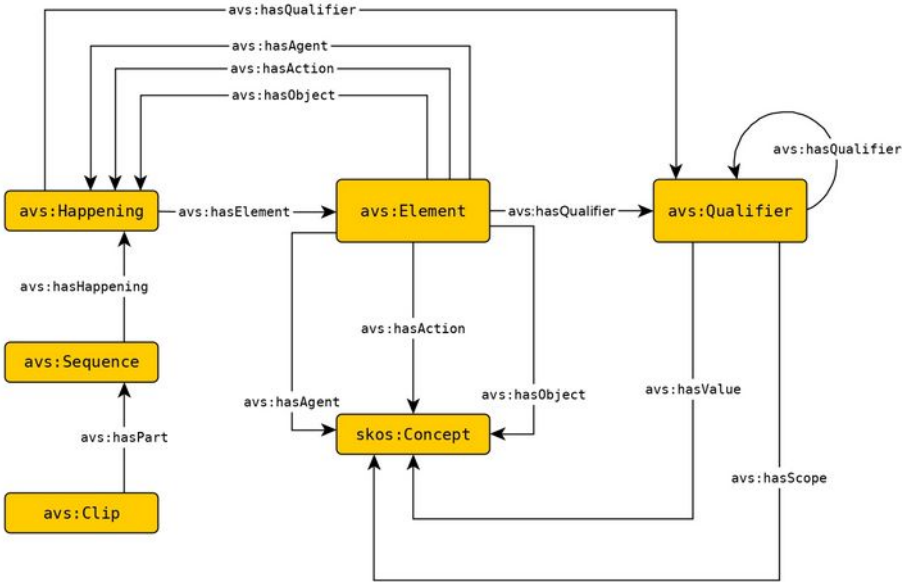


Figura 2. Elementos de la ontología AVS. Fuente: elaboración propia.

El uso de cualificadores (avs:Qualifier) permite definir cualidades o atributos de los diferentes elementos que intervienen en un suceso. Para ello los cualificadores tienen dos componentes que se definen mediante sus correspondientes propiedades:

3 La ontología estará disponible en <http://purl.org/umu/avs> y actualmente se encuentra en la fase de validación final, tras la cual será documentada.

- El alcance del cualificador (avs:hasScope) que permite identificar la característica que se está cualificando (lugar del suceso, vestimenta de una persona, estado de ánimo, situaciones meteorológicas, momento del suceso, etc).
- El valor del cualificador (avs:hasValue) que permite seleccionar un valor concreto para el alcance del cualificador (amarillo, vestido, enfadado, tormenta eléctrica, amanecer, etc).

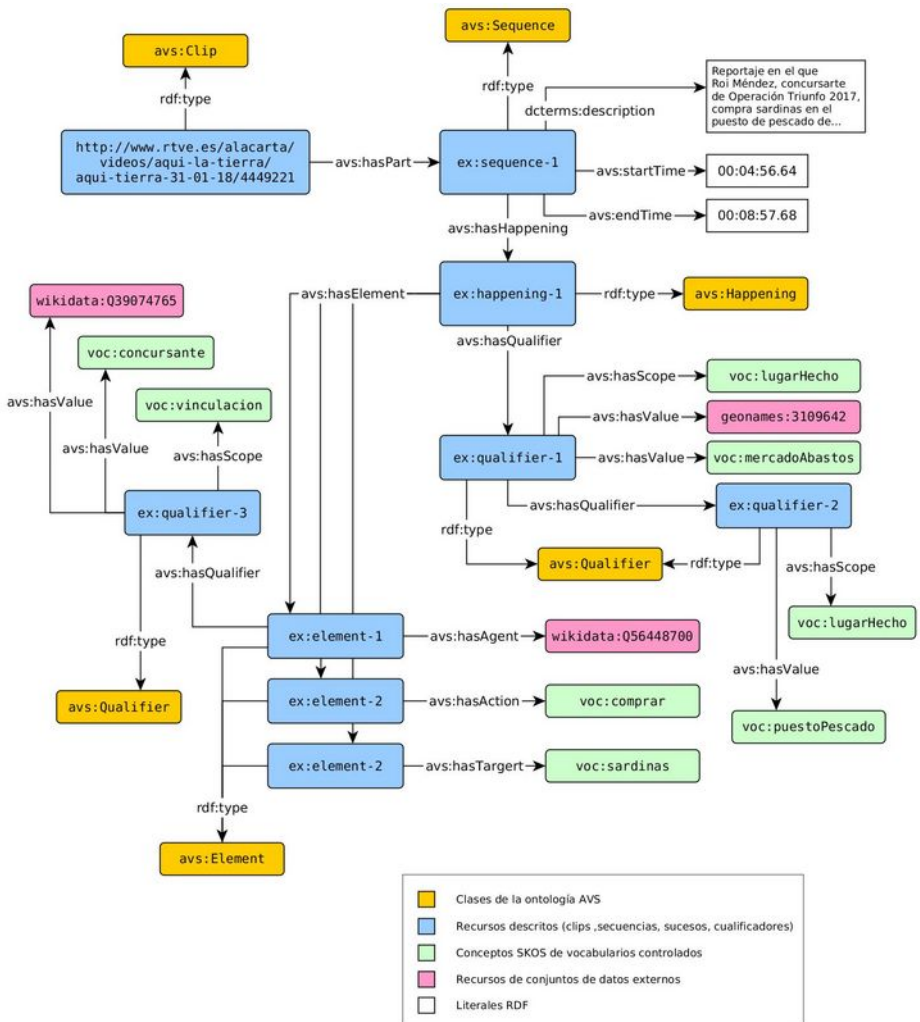


Figura 3. Elementos de la ontología AVS. Fuente: elaboración propia.

El ejemplo de la figura 3 delimita una secuencia dentro de un clip. Al mismo tiempo se identifica un suceso específico, que se cualifica mediante el recurso `avs:qualifier-1` para indicar la ubicación en la que se desarrolla el suceso mediante la intersección del concepto “Mercado de abastos” con el elemento de Geonames correspondiente a Santiago de Compostela. Dicho cualificador es a su vez cualificado para establecer dicha ubicación con mayor detalle. El suceso tiene asociados tres elementos: agente que realiza la acción (el elemento de wikidata correspondiente a Roi Méndez), acción realizada (comprar) y el objeto de la acción (sardinas). El elemento correspondiente al agente se cualifica para definir su vinculación como concursante de Operación Triunfo 2017 (elemento de Wikidata). Este ejemplo también muestra como es posible reutilizar conjuntos de datos externos en la descripción de los sucesos de una secuencia (en este caso Wikidata y Geonames).

El uso de SKOS es un aspecto relevante de AVS. Los agentes, acciones, objetos y cualificadores no se definen de forma extensiva en la propia ontología, sino que son referenciados como conceptos de vocabularios SKOS. Los vocabularios pueden organizarse en torno a esquemas de conceptos o colecciones SKOS para representar grupos de conceptos para lugares, acciones, alimentos, organizaciones, personas, programas de televisión, roles, etc. Incluso, la cualificación permitiría definir cuando el elemento objeto de un suceso es receptor directo o indirecto de la acción. De este modo se consigue una mayor flexibilidad durante la aplicación de la ontología en la representación de diferentes atributos, autoridades, lugares, etc. La definición de nuevos valores únicamente requerirán la creación de un nuevos conceptos SKOS. De este modo, se evita alterar la ontología AVS y se reutilizan las diversas relaciones que ofrece SKOS para la organización jerárquica y asociativa de conceptos, así como las propiedades para la desambiguación terminológica y el etiquetado multilingüe. AVS también permite reutilizar recursos de conjuntos de datos externos como Wikidata, Geonames, VIAF, etc.

4 Conclusiones y líneas de trabajo futuras

La ontología AVS tiene una estructura sencilla y adecuada para la descripción rápida de la semántica de los aspectos conceptuales de secuencias de documentos audiovisuales. En la actualidad se está procediendo a la evaluación y validación de la ontología mediante su aplicación en la descripción de piezas audiovisuales facilitadas por RTVE y la verificación de los resultados obtenidos. También se están realizando consultas SPARQL sobre los conjuntos de datos generados para evaluar la viabilidad de la ontología en los procesos de búsqueda. Por el momento no se disponen de datos suficientes para asegurar si la metodología utilizada es válida tanto para un entorno de contenidos audiovisuales de carácter general (ficción, informativos, documentales) o si es más adecuada para colecciones más homogéneas desde el punto de vista de su contenido.

En la actualidad, la recuperación de objetos audiovisuales digitales se realiza principalmente a partir de descripciones textuales de los mismos y la aplicación de técnicas de recuperación de información donde la precisión y la exhaustividad son afectados por los clásicos problemas relacionados con el lenguaje natural: sinonimia, polisemia, multilingüismo, ambigüedad semántica, etc. La formalización de la descripción del contenido mediante ontologías es una solución de gran interés para evitar dichos problemas. En este sentido, AVS aporta un alto grado de formalización de dichas descripciones que además pueden aplicar técnicas de inferencia y descubrimiento de datos para mejorar los procesos de búsqueda de contenidos.

Por otro lado, la aplicación de vocabularios SKOS supone una solución válida para incorporar nuevos elementos de descripción sin que sea preciso definir una compleja estructura de clases y subclases sujeta a las nuevas necesidades identificadas durante el proceso de descripción. Por su parte, el uso adecuado de etiquetas preferentes y alternativas permite agilizar los procesos de descripción y recuperación y aporta una mayor coherencia terminológica. También es posible utilizar relaciones jerárquicas y asociativas para organizar adecuadamente los elementos de los vocabularios. Otra ventaja que aporta el uso de SKOS es la posibilidad de compartir y reutilizar los vocabularios entre diferentes operadores y de integrar su aplicación en herramientas de búsqueda.

El paso más inmediato será analizar la operatividad de la ontología en una herramienta de descripción de documentos audiovisuales. La interacción entre datos, interfaz y usuario debe realizarse del modo más ágil posible. Dentro de este punto también se contemplará la generación y gestión de los vocabularios SKOS durante el proceso de descripción. La búsqueda y recuperación secuencias concretas también deberá ser diseñada cuidadosamente para que el planteamiento y ejecución de consultas resulte una tarea sencilla.

Son interesantes las propuestas de trabajos que apuntan la integración de las tecnologías semánticas en los sistemas de recomendación (Sotelo, Juayek y Scuoteguazza, 2013; Sotelo y Juayek, 2015). Algunas posibles líneas de trabajo podrían ser la integración con dichas propuestas, así como con otras para la representación de narrativas transmedia (Pastor y Saorín, 2018) y la reutilización/mapeado de otros vocabularios y ontologías (p.e. Dublin Core, EBUCore) para la anotación o enriquecimiento de las descripciones.

Referencias

- Corrado, E.M. & Moulaison Sandy, H.L. (2016). *Archiving Conference, Archiving 2016 Final Program and Proceedings*, pp. 161-166.
<https://doi.org/10.2352/issn.2168-3204.2016.1.0.161>

- Dimoulas, C., Veglis, A., & Kalliris, G. (2015). *Audiovisual hypermedia in the semantic Web. En: Encyclopedia of Information Science and Technology* (3ª edición), pp. 7594-7604. IGI Global.
- Evain, J.P., Matton M. & Vaervagen, T. (2017) Wikipedia and DBpedia for Media - Managing Audiovisual Resources in Their Semantic Context. En: van Erp M. et al. (eds) *Knowledge Graphs and Language Technology ISWC 2016*, pp. 41-56. Lecture Notes in Computer Science, vol 10579. Springer, Cham. https://doi.org/10.1007/978-3-319-68723-0_4
- Evens, T., & Hautekeete, L. (2011). Challenges of digital preservation for cultural heritage institutions. *Journal of Librarianship and Information Science*, 43(3), 157-165. <https://doi.org/10.1177/0961000611410585>
- Fourati, M., Jedidi, A. & Gargouri, F. (2014), Towards a Semantic Multi-modalities Description of Audiovisual Documents. En *Proceeding ISM '14 Proceedings of the 2014 IEEE International Symposium on Multimedia*, 259-262. <https://doi.org/10.1109/ISM.2014.28>
- Hallinan, B., & Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society*, 18(1), 117-137. <https://doi.org/10.1177/1461444814538646>
- Höffernig, M., & Bailer, W. (2009). Formal metadata semantics for interoperability in the audiovisual media production process. En *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe 2009)*. http://ceur-ws.org/Vol-539/paper_6.pdf
- Höffernig M., Bailer W., Nagler G. & Mülner H. (2011) Mapping Audiovisual Metadata Formats Using Formal Semantics. In Declerck et al. (eds) *Semantic Multimedia. SAMT 2010*. Lecture Notes in Computer Science, vol 6725. Springer, Berlin, Heidelberg. <https://pdfs.semanticscholar.org/5c30/8ceaa43bdd9c061522d9245ae0e510ddea5.pdf>
- Hunter, J. & Nack, F. (2001). An overview of the MPEG-7 description definition language (DDL). *IEEE Trans. Circuits Syst. Video Techn*, 11, 765-772. <https://pdfs.semanticscholar.org/635a/6745f57ba892f28d854c84f8f54c39fa746f.pdf>
- Isaac, A.; Troncy, R. (2004). Designing and Using an Audio-Visual Description Core Ontology. *Workshop on Core Ontologies in Ontology Engineering at EKAW'04*, Whittlebury. <https://pdfs.semanticscholar.org/bca6/00fb958d5b97709fd383c14f25b14ea2ddd9.pdf>
- Mendonça, F.M., & Soares, A.L. (2017). Construindo ontologias com a metodologia ontoforinfoscience: uma abordagem detalhada das atividades do desenvolvimento ontológico. *Ciência da Informação*, 46(1), p. 43-59. <http://revista.ibict.br/ciinf/article/view/4013/3713>
- Pastor-Sánchez, J.A., Saorín, Tomás (2018). A Conceptual model for an OWL ontology to represent the knowledge of transmedia storytelling. En *Proceedings of the Fifteenth International ISKO Conference*, pp. 511-520. <http://hdl.handle.net/10760/33908>

- Plank, M. (2018). Managing Born-Digital Audiovisual Media. *International Association of Sound and Audiovisual Archives (IASA) Journal*, (47), 61–67.
<http://journal.iasa-web.org/pubs/article/view/56>
- Raimond Y., Scott T., Oliver S., Sinclair P. & Smethurst M. (2010) Use of Semantic Web technologies on the BBC Web Sites. En: Wood D. (eds) *Linking Enterprise Data*, pp. 263-283. Springer, Boston, MA.
- Sotelo, R., Juayek, M. & A. Scuoteguazza (2013). A comparison of audiovisual content recommender systems performance: Collaborative vs. semantic approaches. En *2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1-5. DOI: 10.1109/BMSB.2013.6621791
- Sotelo, R. & Juayek, M. (2015). Incidence of specific semantic characteristics on the performance of recommender systems of audiovisual content. En *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1-4. DOI: 10.1109/BMSB.2015.7177277

Organização do conhecimento durante o processo de investigação: utilização do *ATLAS.ti* em duas teses de Doutoramento

Luis Corujo¹, Jorge Revez² e Carlos Guardado da Silva³

¹ ORCID [0000-0003-4411-2453](https://orcid.org/0000-0003-4411-2453). Centro de Estudos Clássicos, Faculdade de Letras, Universidade de Lisboa, Portugal; CEIS20, Universidade de Coimbra, Portugal.
luiscorujo@campus.ul.pt

² ORCID [0000-0002-3058-943X](https://orcid.org/0000-0002-3058-943X). Centro de Estudos Clássicos, Faculdade de Letras, Universidade de Lisboa, Portugal; CEIS20, Universidade de Coimbra, Portugal.
jrevez@campus.ul.pt

³ ORCID [0000-0003-1490-8709](https://orcid.org/0000-0003-1490-8709). Centro de Estudos Clássicos, Faculdade de Letras, Universidade de Lisboa, Portugal.
carlosguardado@campus.ul.pt

Resumo. Nas diferentes etapas do ciclo de vida da investigação, a geração de novo conhecimento espoleta a necessidade de executar tarefas de organização desse conhecimento. Ao mesmo tempo, em cada uma das etapas, a própria geração de conhecimento depende do apoio de tarefas de organização do conhecimento. Este trabalho pretende discutir a organização do conhecimento realizada no curso dos processos de investigação científica. Com base em duas experiências de investigação distintas, na área da Ciência da Informação, que consubstanciam este como um estudo multicaso, é problematizado o papel do *software ATLAS.ti* nos processos de organização do conhecimento. Este estudo parte do questionamento da existência de interrelações entre a Investigação Científica, as Ferramentas Tecnológicas e a Organização do Conhecimento, focando as potencialidades de uma ferramenta, que permite trabalhar os dados empíricos e executar a sua análise. Visa conhecer o grau de intervenção nos processos de organização do conhecimento subjacentes à investigação e o contributo geral, deste tipo de ferramentas tecnológicas, para a produção científica. Os casos relatados demonstram a existência de interrelações, uma vez que o *ATLAS.ti* permite trabalhar os dados empíricos e executar a sua análise, plasmando e facilitando mecanismos de organização do conhecimento, enquanto processo intelectual iterativo. O produto, ou seja a base de dados, que vai sendo construída a partir e na unidade hermenêutica, juntamente com os memorandos que relatam a evolução do processo, podem ser usados por outros para chegar a outras contextualizações, isto é, a um conhecimento diferente daquele que foi atingido por quem construiu o primeiro (tipo de) conhecimento. Mostra-se, assim, que esta utilização ajuda a manejar e a articular os dados

durante o processo investigativo, o que fará emergir uma teoria, concretizando-se como novo conhecimento, que poderá dar origem a estudos futuros.

Palavras-chave: Organização do Conhecimento; Investigação qualitativa; *Software*; *Computer Assisted/Aided Qualitative Data Analysis Software (CAQDAS)* ; *ATLAS.ti*.

Abstract. At the different stages of the research life cycle, the generation of new knowledge prompts the need to perform organizational tasks of this knowledge. At the same time, in each of the stages, the knowledge generation itself depends on the support of knowledge organization tasks. This work intends to discuss the organization of knowledge during scientific research processes. Based on two distinct research experiences, in Information Science area, which consubstantiate this as a multi-case study, the role of ATLAS.ti software in the processes of knowledge organization is problematized. This study starts from the questioning of the existence of interrelationships between Scientific Research, Technological Tools and Knowledge Organization, focusing on the potentialities of a tool that allows the empirical data to be worked out and analyzed. It aims to know the degree of intervention in the processes of organization of knowledge underlying research and the general contribution of this type of technological tools to scientific production. The reported cases demonstrate the existence of interrelations, since ATLAS.ti allows to work the empirical data and execute its analysis, shaping and facilitating mechanisms of knowledge organization, as iterative intellectual process. The product, that is, the database, which is being constructed from and in the hermeneutical unit, together with the memos that report the evolution of the process, can be used by others to reach other views, that is, to a different knowledge of the one who was struck by the one who built the first (kind of) knowledge. It is shown, therefore, that this use helps to manage and articulate the data during the investigative process, which will emerge one theory, materializing as new knowledge, that may give rise to future studies.

Keywords: Knowledge Organization; Qualitative Research; Software; Computer Assisted/Aided Qualitative Data Analysis Software (CAQDAS); *ATLAS.ti*.

1 Introdução e Objetivos

O ciclo de vida da investigação científica é composto por diferentes momentos e sucessivas etapas, sendo um processo complexo e multilinear. Em cada uma das suas etapas, a geração de novo conhecimento espoleta a necessidade de executar tarefas de organização desse conhecimento. Ao mesmo tempo, em cada uma das etapas, a própria geração de conhecimento depende do apoio de tarefas de organização do conhecimento.

Este trabalho pretende discutir a organização do conhecimento realizada no curso dos processos de investigação científica. Com base em duas experiências de investigação distintas, na área da Ciência da Informação, é problematizado o papel do *software ATLAS.ti* nos processos de organização do conhecimento. Esta ferramenta é um dos produtos disponíveis no mercado para a análise de dados, principalmente de natureza qualitativa. As aplicações *CAQDAS* (*Computer Assisted/Aided Qualitative Data Analysis Software*) permitem trabalhar os dados empíricos e executar a sua análise, plasmando e facilitando mecanismos de organização do conhecimento, enquanto processo intelectual iterativo.

O presente estudo assenta no questionamento da existência de interrelações entre a Investigação Científica, as Ferramentas Tecnológicas e a Organização do Conhecimento, e discute a utilização do *software ATLAS.ti*. Este permite trabalhar os dados empíricos e executar a sua análise. Visa-se conhecer o seu grau de intervenção nos processos de organização do conhecimento subjacentes à investigação e o contributo geral, deste tipo de ferramentas tecnológicas, para a produção científica.

2 Estado da Arte

2.1 O ciclo de vida da investigação científica na perspectiva da Organização de Conhecimento

A Organização do Conhecimento (OC) interessa-se por «descrever, representar, arquivar e organizar documentos», bem como por «representar documentos, assuntos e conceitos», quer por seres humanos quer por aplicações informáticas (Hjørland, 2008, 2016, p. 475), e a forma de partilhar o conhecimento (Gnoli, 2010).

Por outro lado, o ciclo de vida da investigação científica é composto por diferentes etapas, desde a ideia inicial até à avaliação do impacto dos resultados. Nas etapas centrais, um elemento essencial é o processo de organização do conhecimento, que se desenvolve no curso da investigação. Um dos aspetos-chave da OC consiste, precisamente, nos processos de organização do conhecimento (POC) (Hjørland, 2016, p. 475), que integram a análise de conteúdo, como se concretiza no âmbito dos dois casos relatados de investigação qualitativa, quando da organização do conhecimento em sistemas de conceitos. Processo este que requer, para além da figura central do investigador, o recurso à tecnologia da informação, designadamente a aplicação *ATLAS.ti*. Sendo os dois projetos de investigação realizados no âmbito de doutoramento em Ciência da Informação, este pode ser definido como um campo de aplicação da OC, entendendo-se esta como uma ciência ou uma subdisciplina da Ciência da Ciência, como a definiu Dahlberg (2014 *apud* Dodebei, 2014, p. 4).

Dentre as distintas abordagens no âmbito da OC, aquelas podem classificar-se, segundo Hjørland, em cinco tipos: práticas e intuitivas, com recurso a critérios de praticidade ou intuitivos (Leydesdorff, 2006, p. 602); baseadas em consenso

(científico e educacional); facetadas analíticas; cognitivas e baseadas no utilizador; e epistemológicas ou de análise de domínio. Esta última abordagem destaca-se pelo facto de entender a classificação de qualquer objeto, qualquer documento ou domínio de múltiplas perspectivas igualmente corretas (Hjørland, 2016, pp. 476–477), dependendo a informação do ponto de vista de uma comunidade específica (Hjørland, 2002, p. 116) ou de um investigador. A análise de domínio permite «avaliar o que é realmente importante ou significativo em um determinado campo científico», podendo identificar-se, por exemplo, tendências e pensamentos dominantes (Gutierrez Castanha & Wolfram, 2018, p. 15). Neste sentido, assume especial relevância para a OC particularmente em estudos de natureza epistemológica, em processos sociais de produção e uso da informação (Guimarães, 2014) e em POC, facilitando, inclusive, a elaboração de teoria (Smiraglia, 2015).

De outra perspetiva, a maioria das classificações, tendo por base *softwares*, assenta, sobretudo, nos princípios da literatura (de referência) e da enumeração de classes, em oposição aos sistemas facetados, isto é, sem fundamentação teórica (Hjørland, 2016, p. 479).

Quando considerados os campos científicos sobre os quais se debruça a OC, eles são distintos, englobando, a título de exemplo, a dimensão física e digital das bibliotecas, dos arquivos e dos museus, que possuem, por regra, sistemas específicos de organização da informação e do conhecimento. Todavia, independentemente da diversidade de campos, encontra-se, ou é desejável encontrar-se, uma base teórica comum da OC (Ørom, 2003).

Em bases de dados de classificação bibliográfica, a que se recorreu na segunda Investigação (veja-se infra), a OC consiste, em sentido restrito, no «desenho de registos bibliográficos e sistemas de vocabulários controlados», enquanto, em um sentido mais amplo, se ocupa de «como o conhecimento é organizado nos diferentes domínios e como pode ser usado para a Recuperação da Informação» (Hjørland, 2016, p. 481).

Por outras palavras, de que modo se organiza o conhecimento na sociedade? Tendo presentes os dois casos concretos, de que modo se organizou o conhecimento no âmbito dos processos, em contextos concretos e distintos de investigação? Assentando ambas as tentativas de classificação na literatura, são igualmente perspectivas de organização do conhecimento social, assumindo particular relevância para a classificação das respetivas áreas disciplinares (Hjørland, 2016, p. 482). Por sua vez, tendo por base teorias académicas, pode designar-se este tipo de classificação por intelectual, por oposição à classificação social (Hjørland, 2016, p. 482). No fundo, são ambas construções de sistemas conceptuais distintos, construídos a partir de campos sociais diversos, independentemente do recurso a tecnologias de informação e do seu tipo.

2.2 Ferramentas Tecnológicas de Apoio à Organização de Conhecimento

Na perspectiva do *ba* (Nonaka & Konno, 1998), conceito japonês que pode ser traduzido por “espaço” partilhado para desenvolver relações, a criação do conhecimento está dependente do contexto. Este contexto pode ser físico (ex.: escritórios e espaços espalhados pela organização), virtual (ex.: e-mail, teleconferência) ou mental (ex.: experiências, ideias e ideais partilhados), mas integra sempre conhecimento, que é adquirido através das experiências individuais ou reflexões de outrem.

A participação no *ba* significa estar envolvido na criação de conhecimento, diálogo, adaptação e definição de práticas, e simultaneamente transcender as perspectivas ou limites do indivíduo. Se o conhecimento for separado do *ba*, deixa de ser Conhecimento e passa a ser Informação, que pode ser comunicada para lá do *ba*. Os autores referem, também, que muitos tipos de sistemas de informação suportam o *ba* e possibilitam a criação do conhecimento, nomeadamente os que integram funções de troca e organização do conhecimento, como repositórios eletrónicos, comunicação por e-mail, colaboração e simulação. Autores como Almeida (2004), Carvalho (2006) e Corujo (2017) demonstram as potencialidades das ferramentas tecnológicas de apoio à OC, desde o uso de aplicações informáticas para registo e desenho de mapas mentais ou do conhecimento, o *brainstorming* como elemento de estimulação da criatividade, a Internet, as Intranets, o *groupware* identificado com a produção partilhada de ficheiros/documentos em plataformas virtuais, repositórios para a armazenagem e partilha de ficheiros, correio eletrónico para troca de correspondência, a utilização de plataformas e aplicações de Mensagem instantânea ou chats e, ainda, as reuniões por teleconferência via Web.

Com base nestes estudos, aventa-se a hipótese de que aplicações para análise qualitativa assistida por computador (em inglês *Computer Assisted/Aided Qualitative Data Analysis Software* – CAQDAS) têm potencialidades que as identificam como ferramentas tecnológicas de apoio à OC. Estas aplicações oferecem ferramentas que auxiliam na investigação qualitativa, como por exemplo, na análise de uma transcrição, codificação e interpretação de texto, abstração recursiva, análise de conteúdo, análise de discurso, metodologia de amostragem teórica, etc. Algumas implicações metodológicas do uso dos CAQDAS centram-se nos debates sobre a validade, muito embora esta relação possa ser vista a partir dos objetivos de um processo de validação. Portanto, um resultado pode ser provisoriamente válido, desde que todas as medidas prévias necessárias sejam tomadas para evitar erros (Carvajal Llamas, 2001).

3 Estudo Multicaso

Este é um estudo multicaso, suportado em dois casos experimentados de investigação, no âmbito da realização de duas teses de doutoramento. Para Yin, «o estudo de caso é uma investigação empírica que investiga um fenómeno contemporâneo (o "caso") em profundidade e em seu contexto de mundo real, especialmente quando os limites entre o fenómeno e o contexto puderem não ser claramente evidentes. (...) A investigação do estudo de caso enfrenta a situação tecnicamente diferenciada em que existirão muito mais variáveis de interesse do que pontos de dados, e, como resultado conta com múltiplas fontes de evidência, com os dados precisando convergir de maneira triangular, e como outro resultado beneficia-se do desenvolvimento anterior das proposições teóricas para orientar a coleta e a análise de dados» (Yin, 2015, pp. 17–18).

De acordo com Amado, os estudos de caso «podem ser de natureza quantitativa, de natureza fenomenológica e interpretativa, ou mista (os que conciliam o uso de técnicas e instrumentos próprios das abordagens qualitativas e quantitativas)» (2014, p. 121). Mesmo não estando preocupado com a generalização, «nos estudos de caso de investigação, a intenção do investigador vai para além do conhecimento desse valor intrínseco do caso, visando concetualizar, comparar, construir hipóteses ou mesmo teorizar; contudo, o ponto de partida desses processos é a compreensão das particularidades do caso ou dos casos em estudo» (Amado, 2014, p. 124).

Apresentando «um intento de exploração», «tentam descobrir problemáticas novas, renovar perspetivas existentes ou sugerir hipóteses fecundas» (Bruyne, Herman, & Schoutheete, 1977, p. 225). Permitem o «estudo em profundidade de casos particulares, isto é, numa análise *intensiva*, empreendida numa única ou em algumas organizações reais. O estudo de caso reúne informações tão numerosas e tão detalhadas quanto possível com vistas a apreender a totalidade de uma situação» (Bruyne et al., 1977, pp. 224–225).

Os estudos de caso são desprezados por alguns autores, sobretudo pelo exagero do esforço requerido no aprofundamento da investigação e por oferecerem uma base curta para se poder fazer uma generalização a partir do caso. Ora, como sublinha Yin, «os estudos de caso, como os experimentos, são generalizáveis às proposições teóricas e não às populações ou aos universos. Nesse sentido, o estudo de caso, como o experimento, não representa uma "amostragem" e ao realizar o estudo de caso, sua meta será expandir e generalizar teorias (generalização analítica) e não inferir probabilidades (generalização estatística)» (2015, p. 22). De uma forma simples, a «generalização analítica consiste em uma cuidadosa declaração, teoria ou proposição teórica. A generalização pode tomar a forma de uma lição aprendida, uma hipótese de trabalho ou outro princípio que se acredite aplicável a outras situações (não apenas outros "casos similares")» (Yin, 2015, p. 72).

Por último, para a OC, a partir da informação recolhida, em um caso através da técnica da entrevista estruturada e, no outro, da pesquisa documental sobre avaliação da informação de arquivo eletrónica em revistas de arquivística, procura-se investigar

o recurso ao *software ATLAS.ti* para a análise sistemática de dados qualitativos, de forma a compreender fenómenos complexos ocultos em dados não estruturados.

4 As investigações científicas

4.1 Descrição da Investigação 1: Análise de Conteúdo

No curso de uma investigação sobre a relação entre as bibliotecas e a investigação científica, o *software ATLAS.ti* foi utilizado para gerar conhecimento a partir do material empírico recolhido pela técnica de entrevista estruturada.

No esquema adotado nesta investigação, em métodos mistos, os dados recolhidos pelas entrevistas foram utilizados para o diálogo analítico com os dados recolhidos nas fases prévias, de análise documental e de inquérito por questionário.

Apesar do processo de organização da informação obtida ter beneficiado da adoção da técnica de entrevista, que recolhe os dados em modo estruturado, os dados brutos permaneciam desorganizados, à luz da fundamentação teórica, gerada a partir da revisão da literatura, que deveria servir como um esquema de suporte à análise. Foi assim necessário, para a inteligibilidade da informação, a utilização de uma ferramenta de tratamento e análise. A aplicação informática utilizada permitiu a realização de uma análise do conteúdo, de forma mais eficiente e eficaz. Para tal, os dados foram codificados, utilizando uma lista prévia de categorias, conferindo à informação obtida uma grelha de leitura, a partir da qual se extraíram os resultados.

4.2 Descrição da Investigação 2: Teoria Fundamentada

O Método da Teoria Fundamentada tem enformado uma investigação que se iniciou em 2017 e que continua em curso, e que pretende abordar a avaliação de informação de arquivo eletrónica. Partindo da recolha da produção científica em revistas da especialidade, balizada entre 2006 e 2016, e utilizando uma escala de Likert para selecionar os textos pertinentes, tem-se procedido, até ao momento, à codificação dos resumos dos textos selecionados. As estratégias analíticas deste método são inerentemente comparativas e interativas, uma vez que este método orienta os investigadores a fazerem comparações constantes e sistemáticas, e a envolverem ativamente os dados e a teoria emergente durante todo o processo de investigação (Charmaz & Bryant, 2008). As ferramentas de análise incluem a amostragem teoricamente induzida, a codificação sistemática, o Método da Comparação Constante (MCC), as Perguntas Sensibilizadoras, os Memorandos, a Literatura Especializada e os Programas Informáticos (Freitas, 2012), e têm sido utilizadas no âmbito deste estudo. A codificação é feita tendo por base as perguntas sensibilizadoras iniciais e as que vão surgindo ao longo do processo empírico. Os códigos constituem incidências, e permitem a redução dos dados para níveis cada vez mais conceptuais, mais

abstratos, com o recurso ao MCC. Consiste, assim, em um processo de “vai-vem”, em espiral, até perceber qual das categorias emergentes será o eixo central da problemática de modo a produzir a integração teórica, com o apoio da literatura especializada. Neste sentido, os memorandos constituem as peças essenciais da descrição das tarefas e do processo de reflexão durante a análise, permitindo também a validação da investigação. Os programas informáticos, como neste caso o *ATLAS.ti*, têm permitido efetuar a codificação, a redução dos dados pela comparação constante, e o registo dos memorandos e comentários, que relatam todo o processo de investigação.

5 O *ATLAS.ti*: um exemplo de CAQDAS

O *ATLAS.ti* é uma aplicação informática usada principalmente, mas não exclusivamente, na investigação qualitativa ou na análise de dados qualitativos, tendo sido desenvolvida preliminarmente no seio da Universidade Técnica de Berlim, entre os anos 1989 e 1992. A primeira versão comercial do *ATLAS.ti* foi lançada em 1993, estando atualmente na versão 8 para PC e MAC, e já também em ambiente virtual (*cloud*), iOS e Android.

As raízes metodológicas do *ATLAS.ti* encontram-se, mas não lhes estão restritas, na teoria fundamentada, na análise de conteúdo e na eliciação de conhecimento. O objetivo do *ATLAS.ti* é ajudar os investigadores a descobrir e a analisar de forma sistemática fenómenos complexos ocultos em dados não estruturados (texto, multimédia, geoespaciais). A aplicação fornece ferramentas que permitem localizar, codificar e anotar os resultados encontrados nos dados primários, avaliar a sua importância e visualizar as relações, muitas vezes complexas, entre eles.

O *ATLAS.ti* consolida grandes volumes de documentos e mantém registo de todas as notas, apontamentos, códigos e memorandos em todos os campos, que exigem um estudo detalhado e análise do material primário, que pode ser texto, imagem, áudio, vídeo e dados geográficos.

Além disso, fornece ferramentas analíticas e de visualização destinadas a permitir novas visões interpretativas sobre os dados (Silver & Lewins, 2014).

6 Utilização do *ATLAS.ti* no âmbito das duas investigações à luz da Organização de Conhecimento

No âmbito da primeira investigação, esta teve início com a recolha e a integração dos dados numa folha de cálculo do *Microsoft Excel*, sendo posteriormente importados para uma base de trabalho (*Unidade Hermenêutica*) sediada no programa informático *ATLAS.ti* (versão 7.5.18), ferramenta vocacionada para o estudo de dados de natureza qualitativa. O conjunto de dados obtidos foi depois lido e analisado, de forma a tirar partido das potencialidades analíticas desta ferramenta. No *ATLAS.ti*, a importação

dos dados transformou cada entrevista em um documento primário (*primary document*), reunindo as perguntas e as respostas e individualizando-as. Neste sentido, as 13 entrevistas foram, em primeiro lugar, agrupadas em duas famílias de documentos distintas: os investigadores e os bibliotecários. Foi usada a função *Primary Doc Family Manager* para a reunião dos documentos, com a possibilidade de, mais tarde, executar um filtro por família e, assim, proceder à segmentação da informação.

Foi depois codificado o conteúdo de cada entrevista, utilizando a lista de 63 categorias geradas/agrupadas para a construção do questionário. Esta lista foi importada para a Unidade Hermenêutica de forma a facilitar o processo de codificação. Neste processo, a cada trecho ou citação (*quotation* na linguagem do programa) foi associado um ou mais códigos oriundos da referida lista. Em cada código fica então visível o número de vezes que foi utilizado (representando a fundamentação ou *groundedness*) e o número de ligações com outros códigos (representando a densidade ou *density*), ainda que não tenhamos utilizado esta última funcionalidade de interligação entre códigos.

Foram encontrados dois obstáculos no processo de codificação, que são resultado da técnica escolhida: por um lado, as respostas foram geralmente telegráficas (com muitos *sim* e *não*), com pouco conteúdo ou sem qualquer tipo de justificação ou fundamentação; por outro lado, a lista utilizada para a análise e para a codificação nem sempre teve a versatilidade necessária para abranger algumas dimensões semânticas das respostas. De forma a ultrapassar estes escolhos, optou-se por codificar também a percepção positiva ou negativa dos diferentes trechos contidos nas respostas, com o ensejo de poder relacionar esta dicotomia com as populações em análise – investigadores e bibliotecários –, bem como com os códigos associados aos diferentes excertos ou citações. Este trabalho adicional permitiu perceber o tom geral de cada entrevista e criar algumas formas de apresentação de resultados.

O segundo caso relatado recorreu à mesma versão do *ATLAS.ti* do caso anterior, tendo sido importados para a Unidade Hermenêutica dois documentos primários, cada um constituído pelos resumos dos textos considerados com maior pertinência e com probabilidade de pertinência, referenciados como nível 5 e 4 pela escala de Likert. A primeira fase, de codificação aberta, começou pela codificação dos elementos pré-textuais, seguindo as perguntas sensibilizadoras e prosseguiu pela leitura dos resumos e marcação de citações que originaram códigos *in vivo* (*Code In Vivo* na linguagem do programa), em que a designação do código replica a citação. Alguns destes códigos *in vivo* têm sido reduzidos a códigos com designações mais abstratas à medida que se torna evidente a existência de redundâncias. No caso específico da procura da terminologia substantiva, recorreu-se à ferramenta de codificação automática (*autocoding* na linguagem do programa), para procurar as ocorrências dos termos escolhidos nos documentos primários e codificá-los. De notar que o *ATLAS.ti* possibilita a utilização de um mesmo código em várias citações, permitindo verificar a recorrência e a fundamentação, e ainda a criação de super-códigos, que reúnem dentro

de si vários códigos que podem ser considerados aspetos, propriedades e dimensões dos super-códigos.

A investigação encontra-se atualmente na segunda fase, em que os códigos abertos e os códigos pré-textuais, num total de 718, estão a ser reunidos por categorias ou Famílias (*Families* na linguagem da aplicação) com designações suficientemente abrangentes, e estão a produzir-se memorandos que refletem as perceções que impactam na decisão de alocação de cada um dos códigos. Estas famílias (atualmente 22) estão também a ser agrupadas em categorias mais abstratas através da ferramenta de criação de Super Famílias (*Super Family Tool* na linguagem da aplicação). Pretende-se brevemente avançar, ainda nesta fase, para a verificação de interligações entre os códigos, recorrendo ao gestor de ligações entre códigos (*Code-Link Manager* na linguagem da aplicação, e entre as famílias, o que aumentará a densidade da análise, por meio do gestor de redes (*Network Manager* na linguagem do programa). A intenção desta é a descoberta do eixo – categoria central - sobre o qual giram as restantes categorias no contexto da avaliação de informação de arquivo. De referir que as tarefas e as reflexões analíticas têm sido registadas nos memorandos e nos campos de comentários dos códigos, super-códigos, famílias e super-famílias.

7 Discussão

Os casos indicados permitem evidenciar ganhos derivados da utilização do *software ATLAS.ti*, uma vez que permitem lidar com enormes volumes de dados de forma facilitada, o que se traduz em uma poupança de tempo, que pode ser investido pelos investigadores nas tarefas de análise. Muito embora as aplicações informáticas não tomem decisões conceptuais, elas cumprem a função de auxiliar a análise, neste caso, a análise textual, correspondendo a uma sequência lógica de procedimentos de manipulação de dados.

O *ATLAS.ti* permite aplicar funções analíticas em termos da gestão, da leitura e relativamente aos dados, e desenvolver a representação por intermédio de códigos, super-códigos, famílias e super-famílias, bem como definir relações entre estes elementos através de interligações e redes, e proceder à anotação de todas estas operações e reflexões.

Outro aspeto pertinente é a clareza e a possibilidade de sistematização, no âmbito da aplicação/adaptação no *software*, das etapas definidas pelas metodologias utilizadas, o que concorreu para decisões uniformes. Tal aspeto foi reforçado pela utilização extensiva dos memorandos e dos apontamentos nos campos de comentários que, juntamente com a possibilidade de produção automática de relatórios, são elementos essenciais para a reflexão analítica e a garantia da validação da investigação.

No primeiro caso indicado, a análise de conteúdo das entrevistas, realizada com o recurso à aplicação *ATLAS.ti*, permitiu acelerar o processo de tratamento e organização da informação, gerando, de forma eficiente, o conhecimento necessário

para a interlocução entre os dados novos e os dados anteriores, possibilitando a emergência da comparação. O conhecimento gerado pela utilização do *software* foi integrado na base de conhecimentos da própria investigação, genericamente confirmando os resultados prévios. Sem a aplicação, o tempo despendido teria sido maior e os resultados mais frágeis.

No segundo caso relatado, para além dos aspetos já indicados, o *ATLAS.ti* tem tido um papel determinante na aplicação do MCC e no desenvolvimento de códigos cada vez mais abstratos, contribuindo para a redução e a saturação dos dados, com o fim de encontrar um eixo final forte, sobre o qual se efetua a integração teórica das outras categorias, dando origem a uma teoria fundamentada pelos dados manipulados nesta aplicação informática.

Assim, pode afirmar-se que, ao contribuir com ganhos para a investigação científica, ferramentas tecnológicas como o *ATLAS-ti* trazem também ganhos para os contextos de Organização de Conhecimento (Figura 1).

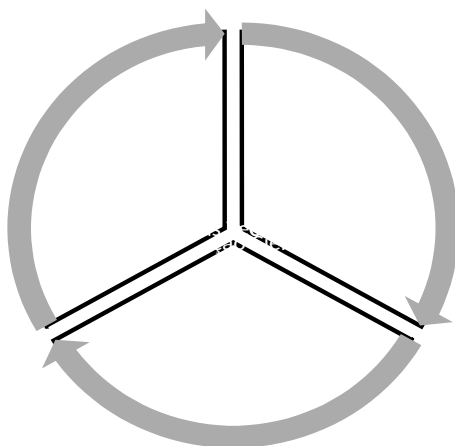


Figura 1. Esquema de sistematização das interrelações entre a Organização do Conhecimento, a Investigação Científica, e as Ferramentas Tecnológicas.

8 Conclusão

Os casos relatados permitem demonstrar a existência de interrelações entre a Investigação Científica, as Ferramentas Tecnológicas e a Organização do Conhecimento, dado que o *software ATLAS.ti* permite trabalhar os dados empíricos e executar a sua análise, plasmando e facilitando mecanismos de organização do conhecimento, enquanto processo intelectual iterativo.

O produto, ou seja a base de dados, que vai sendo construída a partir e na unidade hermenêutica, juntamente com os memorandos que relatam a evolução do processo, podem ser usados por outros para chegar a outras contextualizações, isto é, a um

conhecimento diferente daquele que foi atingido por quem construiu o primeiro (tipo de) conhecimento. Mostra-se assim que esta utilização ajuda a manejar e a articular os dados durante o processo investigativo, o que faz emergir uma teoria, concretizando-se como novo conhecimento, que pode dar origem a estudos futuros.

Uma vez que os casos relatados somente utilizaram o *ATLAS.ti*, um único exemplo de ferramenta tecnológica CAQDAS, considera-se que uma das vias futuras será a ampliação desta investigação a casos de investigação qualitativa ou mista que utilizaram outras aplicações de tipo CAQDAS e mesmo outros *softwares* de tipo estatístico, no âmbito de estudos de carácter quantitativo ou misto.

Referencias

- Almeida, M. A. P. N. de. (2004). *La situación de la gestión del conocimiento en portugal estudio exploratorio y prácticas gubernamentales, académicas y empresariales* (Tese de Doutoramento). Universidade de Extremadura, Cáceres.
- Amado, J. (2014). *Manual de investigação qualitativa em educação*. Disponível em: <https://digitalis.uc.pt/handle/10316.2/35271>
- Bruyne, P. de, Herman, J., & Schoutheete, M. de. (1977). *Dinâmica da pesquisa em ciências sociais: os pólos da prática metodológica*. Rio de Janeiro: Livraria Francisco Alves.
- Carvajal Llamas, D. (2001). Herramientas Informáticas Para El Análisis Cualitativo. *Nómaditas (Col)*, (14), 252–259. Disponível em: <http://www.redalyc.org/articulo.oa?id=105115268019>
- Carvalho, R. B. de. (2006). *Intranets, portais corporativos e gestão do conhecimento: análise das experiências de organizações brasileiras e portuguesas* (Tese de Doutorado). Universidade Federal de Minas Gerais, Belo Horizonte.
- Charmaz, K., & Bryant, A. (2008). Grounded Theory. Em L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 374–377). SAGE Publications.
- Corujo, L. (2017). O projeto IES: análise do processo na ótica da gestão do conhecimento - estratégias de gestão e conversão do conhecimento (estudo de caso). *Brazilian Journal of Information Science: Research Trends*, 11(1). <https://doi.org/10.5016/brajis.v11i1.5994>
- Dodebei, V. (2014). *13a ISKO International Conference, 19-22 Maio 2014, Cracóvia, Polónia: Relatório da participação da ISKO-Brasil* (p. 9). Disponível em: <https://docplayer.net/883836-Isko-brasil-sociedade-brasileira-para-a-organizacao-do-conhecimento.html>
- Freitas, M. C. V. de. (2012). Grounded Theory como Método de Investigação em Arquivologia: Subsídios Teóricos e Práticos. Em M. L. P. Valentim (Ed.), *Estudos Avançados em Arquivologia* (pp. 107–137). Marília, SP: UNESP.

- Gnoli, C. (2010). Classification transcends Library Business. *Knowledge Organization*, 37, 223–229.
- Guimarães, J. A. C. (2014). Análise de domínio como perspectiva metodológica em organização da informação. *Ciência da Informação*, 43(1). Disponível em: <http://revista.ibict.br/ciinf/article/view/1415>
- Gutierrez Castanha, R. C., & Wolfram, D. (2018). The domain of knowledge organization: A bibliometric analysis of prolific authors and their intellectual space. *Knowledge Organization*, 13–22. <https://doi.org/10.5771/0943-7444-2018-l-13>
- Hjørland, B. (2002). Principia Informatica. Foundational Theory of Information and Principles of Information Services. Em H. Bruce, R. Field, P. Ingwersen, & P. Vakkari (Eds.), *Emerging frameworks and methods: proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4)* (pp. 109–121). Disponível em: <https://repository.arizona.edu/handle/10150/105735>
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2–3), 86–101. <https://doi.org/10.5771/0943-7444-2008-2-3-86>
- Hjørland, B. (2016). Knowledge Organization. *Knowledge Organization*, 43(6), 475–485.
- Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, 57(5), 601–613. <https://doi.org/10.1002/asi.20322>
- Nonaka, I., & Konno, N. (1998). The Concept of “Ba”: Building a Foundation for Knowledge Creation: *California Management Review*, 40(3), 40–54. <https://doi.org/10.2307/41165942>
- Ørom, A. (2003). Knowledge Organization in the domain os Art Studies: History, Transition and conceptual changes. *Knowledge Organization*, 30, 128–143.
- Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide* (2ª). Disponível em: <https://uk.sagepub.com/en-gb/eur/using-software-in-qualitative-research/book238134>
- Smiraglia, R. P. (2015). *Domain analysis for knowledge organization: tools for ontology extraction*. Waltham: Elsevier Chandos Pub.
- Yin, R. K. (2015). *Estudo de caso: planejamento e métodos* (5.ª ed.). Porto Alegre: Bookman.

A interdisciplinaridade como elemento de convergência no contexto patrimonial entre o Mundo da vida e o Mundo dos sistemas

Marco Donizete Paulino da Silva¹, Luciana de Souza Gracioso² y Maria da Graça Melo Simões³

¹ORCID [0000-0003-1401-7180](https://orcid.org/0000-0003-1401-7180). Claretiano Centro Universitário-Pólo São Carlos, Brasil.
marco_donizete@yahoo.com.br

²ORCID [0000-0002-6320-4946](https://orcid.org/0000-0002-6320-4946). Departamento de Ciência da Informação, Universidade Federal de São Carlos, Brasil.
luciana@ufscar.br

³ORCID [0000-0001-5575-0261](https://orcid.org/0000-0001-5575-0261). Faculdade de Letras da Universidade de Coimbra, Portugal.
gsimoesuc@gmail.com

Resumen. Con base en la relación de consonancia entre los conceptos de Mundo de la Vida/Mundo de los Sistemas, Epistemología General/Epistemología Regional, propone observar los grados de convergencia entre discursos científicos - en el orden de capital científico/ propiedad intelectual de dominios científicos -, mientras que objetos y en el sentido de confirmar la proposición, evalúa los usos y apropiaciones conceptuales de los términos Interdisciplinariedad e Interdisciplinaria en objetos textuales de orden reflexivo-filosófico, científicos, localizados en el período temporal de 1972-2011. Describe estas circunstancias y análisis, concluyendo la eficiencia de una nueva metodología (método de análisis hermenéutico-dialéctico de matriz conceptual-HMC) y la evidencia de convergencia de los usos de los términos citados como foco de análisis en los ámbitos analizados.

Palabras clave: Interdisciplinariedad; Acción Comunicativa; propiedad intelectual.

Abstract. Based on the relation of consonance between the concepts of World of Life/Systems World, General Epistemology/Regional Epistemology, it proposes to observe the degrees of convergence between scientific discourses - in the order of scientific capital/intellectual property of scientific domains - as objects Interdisciplinarity is considered as intermediation technology and, therefore, with the purpose of confirming this proposition, evaluates the conceptual uses and appropriations of the terms Interdisciplinarity and Interdisciplinary in reflexive-philosophical, scientific objects, located in the period time period of 1972-2011. It describes these circumstances and analyzes,

concluding the efficiency of a new methodology (Hermeneutic-Dialectical Analysis Method of Conceptual Matrix-HMC) and the evidence of convergences of the uses of terms cited as focus of analysis in the domains analyzed.

Keywords: Interdisciplinarity; Communicative Action; intellectual property.

1 Introdução

Para elaboração desse trabalho, partimos das interrelações de Silva (2018) sobre quatro conceitos, nas seguintes condições: Epistemologia Geral/Epistemologia Regional, de Robert Blanché (1978) e de Hilton Japiassu (1976), como construção social do conhecimento, respectivamente, em âmbito global, e em domínios científicos; e, Mundo da Vida/Mundo dos Sistemas, de Jürgen Habermas (1987), respectivamente definidos como a teia de relações estabelecidas como pano de fundo para interações sociais em níveis diversos, e as estruturas sistêmicas que se concretizam nas formas institucionais ou organizacionais do arranjo social.

Também consideramos: em Wachowicz (2013), o impacto das Tecnologias de Informação e Comunicação (TICs) em situações de aumento na democratização do acesso, pela *Internet*, e de difusão na diversificação de: “[...] padrões culturais, saberes e de memórias coletivas [...]” (p. 221) - situando as condições comunicativas da sociedade contemporânea como reflexo da expansão e multiplicação da variedade de artefatos e processos comunicativos, circulantes em espaços heterogêneos da ciência e da sociedade -; e, em Radomsky (2012), o conceito de propriedade intelectual, localizando-o no contexto patrimonial, enquanto produção científica/capital científico, estabelecido por noções de comunidade/instituição, espacial e temporalmente localizadas em domínios/campos de conhecimento, mas compartilhado, na atualidade, em domínios extra-científicos (WEB).

Nessas condições, visões e argumentações elaboradas enquanto capital simbólico de índole intelectual, evidenciam seu valor de patrimônio/recurso informacional, alimentando o campo científico por meio da discussão em torno dos temas debatidos entre domínios disciplinares. No que tange à sustentabilidade, o desafio dessa situação interativa seria garantir tanto a promoção do desenvolvimento do patrimônio cultural de um domínio quanto estabelecer relações das transformações temáticas a partir da prática interpretativa de domínios transversais (apropriações), em fluxos discursivos potencializados por sistemas de informação (em nível textual discursivo) organizados a partir de abordagem analítica pertinente, proposta aqui por meio da Análise de Domínio que, na perspectiva de Hjørland (2002), pode ser aplicada na ordem do campo de Organização do Conhecimento para distinção entre tais comunidades, em estudos dos seus processos de comunicação científica em níveis: críticos e epistemológicos; estruturais e institucionais.

Assim, o objetivo desse trabalho foi discutir as trocas comunicacionais no âmbito do mundo dos sistemas/epistemologia regional (domínios científicos) e do mundo da

vida/epistemología geral (caráter agregador da WEB e dos ambientes digitais) – conceitos utilizados na perspectiva da Teoria da Ação Comunicativa de Jürguen Habermas (1987) -, concebendo-se a interdisciplinaridade como chave principal (tecnología) de intermediação, com uso do instrumento de categorização PMEST como base de axiomatização dos conceitos em uso, nos domínios estudados. Contextualizando-se tal objetivo em relação aos espaços de institucionalização da memória (museus, bibliotecas e arquivos) pela percepção de que tanto o mundo dos sistemas (no âmbito da Ciência) quanto no mundo da vida (WEB como espaço de troca comunicativa) repercutem a determinação de critérios sobre o que se preservar, estabelecendo-se metas e parâmetros para definição dos “itens” de valor histórico-social, e patrimonial, no contexto de uma Sociedade da Informação, cada vez mais dinâmica e multifacetada na produção, e uso, de recursos informacionais – ambiente abordado à luz do conceito de Esfera Pública, também de Habermas (1984).

2 **Análise Hermenêutico-dialética e PMEST**

O método empregado, de análise hermenêutico-dialética (Minayo, 2002), teve como objeto os conceitos de Interdisciplinaridade e Interdisciplinar, baseando-se na Teoria do Conceito, de Dahlberg (1978), para desconstrução/axiomatização dos seus usos terminológicos em unidades textuais selecionadas dos extratos teóricos que abordaram o tema da Interdisciplinaridade. O instrumento PMEST, de Shiyali Ramamrita Ranganathan (Campos; Gomes, 2003; Borem Lima, 2004), foi aplicado em função auxiliar combinada, resultando um conjunto de procedimentos inovativos - que denominamos Método de Análise Hermenêutico-dialética de Matriz Conceitual (HMC) -, conformados na consideração: de cada título e subtítulo (seções e subseções do documento) como facetas (pontos de vista) do conteúdo/tema; dos níveis (títulos e subtítulos) de abordagem (Facetas ou subfacetas do mesmo conteúdo), admitindo como subfacetas extras, desdobramentos temáticos não explicitados por subtítulos.

O critério de seleção do *corpus* priorizou extratos em que se discutissem sentidos conceituais de interdisciplinaridade e interdisciplinar, entendendo-se que as concordâncias ou discordâncias entre as abordagens de cada autor representariam graus de eficiência comunicacional entre extratos (sistemas discursivos de base científico-filosófica), distintos entre si pelo caráter disciplinar de cada domínio representado/defendido por cada autor, representação e defesa consideradas como aspecto epistemológico da comunicação.

Em razão desses interesses e condições, nosso *corpus* compôs-se em duas modalidades/etapas, distinguidas em: Fase 1, relatório CERI (Centre for

Educational Research and Innovation, 1972) sobre congresso de 1970: abordagens teóricas de Apostel (1972), Henckhausen (1972), Boisot (1972), Jantsch (1972), Lienerowicz (1972), e Piaget (1972); e, Fase 2, excertos reflexivos diversos: apropriações de Fazenda (2008, 2011), Klein (2001, 2010), Pombo (1993, 2003) e Japiassu (1976, 1994), produzidos após 1970. O quadro para coleta e análise dos enunciados foi construído em programa Excel, com um cabeçalho dividido em: Unidade/referência, Conteúdo enunciativo (traduzido, quando em outro idioma), Intervalo entre páginas, Facetas PMEST –representadas graficamente pelas iniciais do termo Faceta e da categoria respectiva (em inglês), resultando: Personalidade (FP), Materialidade (FM), Energia (FE), Espaço (FS). Na Faceta Tempo, consideramos, apenas, o ano de produção de cada texto.

3 Resultados e Discussões

Por meio dos investimentos nesses procedimentos obtivemos os resultados apresentados pelos Quadros 1, 2, 3 e 4, que contém, respectivamente, contraposição das interpretações de convergências dos uso dos conceitos de Interdisciplinaridade e Interdisciplinar nas unidades textuais de Apostel 1 e 2 (1972), Henckhausen (1972), Jantsch (1972), e Piaget (1972), em relação à Japiassu (1976, 1994), Fazenda (2008, 2011), Klein (2001, 2010), e Pombo (1993, 2003).

Entendemos que a lógica de convergências de usos dos termos Interdisciplinaridade e Interdisciplinar, em suas facetas PMEST, se estabeleceram por meio de concordâncias - mesmo que acidentais - acerca do emprego do termo, cabendo, entretanto, a consideração de uma interpretação dos níveis e condições dessas concordâncias para se conferir eficiência comunicativa entre os domínios, e as definições conceituais do tema entre si, atribuindo-lhes uma consequente funcionalidade comunicativa dos enunciados (mundo dos sistemas) e seu fluxo no contexto do mundo da vida (WEB), identificando-se em:

- Apostel 1 (1972) (Quadro 1): em Klein (2010), o sentido original das FS OECD (Organisation for Economic Co-operation and Development)/Vocabulário, como espaços de discussão da questão léxica em perspectiva orientadora.

- Apostel 2 (1972) (Quadro 1): identificadas nas FM Pesquisa, FE Ser/Poder/Promover, e FS Pesquisa/Educação - registradas em três autores (Japiassu, 1976; Klein, 2001; Fazenda, 2008), nas seguintes condições de sentido: originalmente relacionados à pesquisa, enquanto estado/identidade/racionalidade, enquanto ferramenta/instrumento promocional; e, disseminadamente teorizado, como prática aplicada e inovadora, constitutiva de processo auxiliar na reestruturação de programas, em caráter de instrumento empírico, no contexto histórico e cultural.

Quadro 1. Convergências de Apostel 1 e 2 (1972) em extratos de Klein (2010, 2001), Japiassu (1976) e Fazenda (2008).

Apostel (1)			
FP	Faceta convergente	Sentido original	Sentido em Klein (2010)
Ambas	FS OECD/Vocabulário	Espaços de discussão Movimento cognitivo definidor	Instrumento de intermediação Caracterização de relações gêneros/espécies (perspectiva léxica)
Apostel (2)			
FP	Faceta convergente	Sentido original	Sentido em Japiassu (1976)
Interdisciplinar	FM Pesquisa	Componente de ação racional de política de pesquisa	Prática aplicada Componente inovador Elemento multifacetado
Interdisciplinar	FE Ser	Estado Identidade promotora da pesquisa científica	Elemento constitutivo ou em processo de efetivação
Interdisciplinar	FS Pesquisa	Exigência Ferramenta conceitual Incremento Propriedade da pesquisa	Instrumento auxiliar do processo de reestruturação de programas de pesquisa
FP	Faceta convergente	Sentido original	Sentido em Klein (2001)
Interdisciplinar	FE Poder/Promover	Ferramenta Qualidade de promoção bilateral	Instrumento de trabalho e de engajamento
FP	Faceta convergente	Sentido original	Sentido em Fazenda (2008)
Interdisciplinaridade	FS Educação	Instrumento conceitual	Prática empírica, histórica e culturalmente contextualizada

- Jantsch (1972) (Quadro 2): os sentidos originais das FS Ciência/Sistema/Disciplina, FE Definir e FM Forma, com o caráter de coordenação de espaços científicos capacitados em proporcionar feedbacks entre sistemas humanos e sociais, se mantiveram e se ampliaram ao considerar tal caráter de atuação em espaços disciplinares, a partir de uma axiomática comum identificada em Japiassu (1976, 1994), Pombo (1993) e Fazenda (2008) – resultando sentidos de tecnologia social aos termos, em função de uma organização hierárquica inovadora e disciplinarmente enriquecedora, gerenciadora de relações disciplinares. Tais sentidos, em Fazenda (2008), predominaram como função pedagógica (campo da Educação), elaborados ao nível de fenômeno curricular (aluno e docente) potencializador dessas formações e, conseqüentemente – já que ambas as categorias atuavam no contexto social –, científica e socialmente convergentes.

Quadro2. Convergências de Jantsch (1972) em extratos de Japiassu (1976, 1994), Pombo (1993) e Fazenda (2008).

Jantsch			
FP	Faceta convergente	Sentido original	Sentido em Japiassu (1976)
Interdisciplinar	FS Ciência	Coordenação de espaços científicos	Tecnologia social com capacidade de organização de níveis hierárquicos
Interdisciplinaridade	FS Sistema	Feedbacks entre sistemas humanos e sociais	Função finalista Elemento gerenciador da organização das relações em níveis disciplinares
FP	Faceta convergente	Sentido original	Sentido em Japiassu (1994)
Interdisciplinar	FS Disciplina	Capacidade de coordenação de espaços disciplinares	Qualidade inovadora Exigência de reorganização
FP	Faceta convergente	Sentido original	Sentido em Pombo (1993)
Interdisciplinaridade	FE Definir	Axiomática comum que possibilita convergência disciplinar	Instrumento de enriquecimento mútuo entre disciplinas
FP	Faceta convergente	Sentido original	Sentido em Fazenda (2008)
Interdisciplinaridade	FE Definir	Axiomática comum que possibilita convergência disciplinar	Fenômeno Questão curricular Questão de formação docente
FP Interdisciplinar	FM Forma	Ação de coordenação nos espaços científicos	Potencializadora da formação Potencializadora de convergência científica e social

- Heckhausen (1972) (Quadro 3): as FM Tipo e Tipologia foram enfatizadas como instrumentos auxiliares, com função integradora (perspectivas aplicada e léxica), tanto em Japiassu (1976) quanto em Klein (2010).

Quadro3. Convergências de Heckhausen (1972) em extratos de Japiassu (1976) e Klein (2010).

Heckhausen			
FP	Faceta convergente	Sentido original	Sentido em Japiassu (1976)
Ambas	FM Tipo	Instrumentos auxiliares na integração de espaços	Propriedades de articulação entre campos materiais diversos
FP	Faceta convergente	Sentido original	Sentido em Klein (2010)
Ambas	FM Tipologias/Gênero/Espécie	Instrumentos auxiliares na integração de espaços	Instrumento de intermediação Caracterização de relações gêneros/espécies (perspectiva léxica)

- Piaget (1972) (Quadro 4): a FM Formas/Relações/Tipo, em Japiassu (1976), assumiu o sentido de funções operacionais relativas a questões estruturais. Tais apontamentos, a nosso ver, situam a prática reflexiva desses autores/teóricos, como adjacentes ao campo de atuação a que se inscreveram enquanto pesquisadores, mantendo-se alinhadas aos princípios e fundamentos disciplinares, mas também promovendo – pelo menos no caso da axiomatização via PMEST e, por meio do interesse interdisciplinar – concordância entre sentidos e interesses de desenvolvimento mútuo a outras disciplinas, com pontos de convergência e graus de comunicabilidade entre unidades enunciativas.

Quadro4. Convergências de Piaget (1972) em extrato de Japiassu (1976).

Piaget			
FP	Faceta convergente	Sentido original	Sentido em Japiassu (1976)
Ambas	FM Formas/Relações/ Tipo	Formas de ligações estruturais	Funções operacionais de relações estruturais

4 Considerações Finais

As trocas comunicacionais no âmbito do mundo dos sistemas/epistemologia regional (representado pelo plano científico) e do mundo da vida/epistemologia geral (representado pelo caráter agregador da WEB), concebidas na perspectiva da Teoria da Ação Comunicativa, demonstraram a interação ativa entre o mundo dos sistemas e o mundo da vida, sobretudo pelo aspecto de convergência entre sentidos outorgados na apropriação dos conceitos de interdisciplinaridade e interdisciplinar.

Essa possibilidade, a nosso ver, tornou-se possível pela condição de uma esfera pública de intersecção, coordenada, no presente trabalho pelo uso do instrumento PMEST na intermediação da linguagem, cabendo considerar que, para que tal coordenação se desse em grau eficiente, requereu dos analistas uma axiomatização conceitual dos termos, promovendo a justaposição dos sentidos de cada um em relação aos domínios de apropriação. Um processo analítico de fundo interdisciplinar com potenciais aprofundamentos futuros.

Considerando os desafios da construção intersubjetiva do patrimônio (enquanto objeto de investigação no horizonte da Organização do Conhecimento), entendemos tal construção como um potencial institucional e social, constituída a partir de correlações e sobreposições dinâmicas, assimétricas e interdisciplinares, sendo esta interdisciplinaridade um recurso catalisador (ou tecnologia) do capital científico acumulado, que serviria para comunicação eficiente, convergente, de sedimentações estabilizadas (propriedade intelectual) ao longo dos discursos teóricos enunciados entre domínios, componentes sistêmicos e mundos epistemológicos.

Referencias

- Centre for Educational Research and Innovation [CERI]. (1972). *Interdisciplinarity: problems of teaching and research in universities*. Paris (France): Centre for Educational Research and Innovation.
- Boisot, M. (1972). Discipline and interdisciplinarity. In: Apostel, L., Berger, G., Briggs, A. & Michaud, G. *Interdisciplinarity: problems of teaching and research in universities*, (Cap. 1, pp. 89-97). Paris (France): Centre for Educational Research and Innovation.
- Borem Lima, G. A. (2004). O modelo simplificado para análise facetada de Spiteri a partir de Ranganathan e do Classification Research Group (CRG). *Información, Cultura y Sociedad*, 11, 57-72. Disponível em: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17402004000200003
- Campos, M. L. A. & Gomes, H. E. (2003). Organização de domínios de conhecimento e os princípios rangathanianos. *Perspectivas em Ciencia da Informação*, 8(2), 150-163. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/366>
- Dahlberg, I. (1978). Teoria do conceito. *Ciência da Informação*, 7(2), 101-107. <http://revista.ibict.br/ciinf/article/view/115/115>
- Fazenda, I. C. A. (2008). Interdisciplinaridade-transdisciplinaridade: visões culturais e epistemológicas. In: FAZENDA, I. C. A. et al. (Org). *O Que é interdisciplinaridade?* (Cap. 2, pp. 17-28). São Paulo: Cortez.
- Fazenda, I. C. A. (2011). Relacionamento crítico dos aspectos teóricos com as diretrizes legais. In: Fazenda, I. C. A. *Integração e interdisciplinaridade no ensino brasileiro: efetividade ou ideologia*. (Cap. 5, pp. 131-146). São Paulo: Edições Loyola.

- Habermas, J. (1984). *Mudança estrutural da esfera pública*. Rio de Janeiro: Tempo Brasileiro.
- Habermas, J. (1987). *Teoría de la acción comunicativa*, I. Madri: Taurus Humanidades.
- Heckhausen, H. (1972). Discipline and interdisciplinarity. In: Apostel, L., Berger, G., Briggs, A. & Michaud, G. *Interdisciplinarity: problems of teaching and research in universities* (Cap. 1, pp. 83-89). Paris: Centre for Educational Research and Innovation.
- Jantsch, E. (1972). Towards interdisciplinarity and transdisciplinarity in education and innovation. In: Apostel, L., Berger, G., Briggs, A. & Michaud, G. *Interdisciplinarity: problems of teaching and research in universities* (Cap. 1, pp. 97-121). Paris: Centre for Educational Research and Innovation.
- Japiassu, H. (1976). *Interdisciplinaridade e a patologia do saber*. (Cap. 1, pp. 39-113). Rio de Janeiro: Imago Editora.
- Japiassu, H. (1994). *A questão da interdisciplinaridade*. Seminário Internacional sobre Reestruturação Curricular. (Palestra). Porto Alegre: Secretaria Municipal de Educação.
- Klein, J. T. (2001). Interdisciplinarity and the prospect of complexity: the tests of theory. *Issues in Integrative Studies*, 19, 43-57.
- Klein, J. T. (2010). A taxonomy of interdisciplinarity. In: FRODEMAN, R. et al. *The Oxford handbook of interdisciplinarity*. (Cap. 2, pp. 15-30). Pondicherry: SPI Publisher Services.
- Lienerowicz, A. (1972). Mathematic and transdisciplinarity. In: Apostel, L., Berger, G., Briggs, A. & Michaud, G. *Interdisciplinarity: problems of teaching and research in universities* (Cap. 1, pp. 121-127). Paris: Centre for Educational Research and Innovation.
- Minayo, M. C, S. (2002). Hermenêutica-dialética como caminho do pensamento social. In: Minayo, M. C, S. & Deslandes, S. F. (Orgs). *Caminhos do pensamento: epistemologia e método*. (Cap. 3, pp. 83-107). Rio de Janeiro: Editora Fio Cruz.
- Piaget, J. (1972). The epistemology of interdisciplinary relationships. In: Apostel, L. & Berger, G., Briggs, A. & Michaud, G. *Interdisciplinarity: problems of teaching and research in universities* (Cap. 1, pp. 127-139). Paris: Centre for Educational Research and Innovation.
- Pombo, O. (1993). A interdisciplinaridade: conceito, problemas e perspectivas. In: Pombo, O.; Levy, T. & Guimarães, H. *A Interdisciplinaridade: reflexão e experiência*. (Cap. 1, pp. 8-14). Lisboa: Revista Aumentada.

- Pombo, O. (2003). Epistemologia da Interdisciplinaridade. *Anais do Seminário Internacional Interdisciplinaridade, Humanismo, Universidade*. Porto, Portugal.
- Radomsky, G. F. W. (2012). Problemas e tensões entre as noções de produção, propriedade intelectual e cultura. *Horizontes Antropológicos*, Porto Alegre, 18(37), 155-183. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-71832012000100007
- Silva, M. D. P. (2018). *A interdisciplinaridade como tecnologia de Ação Comunicativa: uma análise hermêutico-dialética*. Tese de doutorado, Universidade Federal de São Carlos, São Carlos, SP, Brasil.
- Wachowicz, M. (2013). Direitos culturais e saberes: o reconhecimento de um direito de propriedade intelectual de natureza difusa. *Políticas Culturais em Revista*, 1(6), p. 220-234. Disponível em: <https://portalseer.ufba.br/index.php/pculturais/article/view/8307>

Filosofia da Linguagem Pragmática como aporte à representação da memória coletiva

Luciana de Souza Gracioso¹, Marco Donizete Paulino da Silva², Maria da Graça Melo Simões³ y Samanta do Prado⁴

¹ ORCID [0000-0002-6320-4946](https://orcid.org/0000-0002-6320-4946). Departamento de Ciência da Informação, Universidade Federal de São Carlos, Brasil.
luciana@ufscar.br

² ORCID [0000-0003-1401-7180](https://orcid.org/0000-0003-1401-7180). Claretiano Centro Universitário-Pólo São Carlos, Brasil.
marco_donizete@yahoo.com.br

³ ORCID [0000-0001-5575-0261](https://orcid.org/0000-0001-5575-0261). Faculdade de Letras da Universidade de Coimbra, Portugal.
gsimoesuc@gmail.com

⁴ ORCID [0000-0002-7386-3419](https://orcid.org/0000-0002-7386-3419). Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, Universidade Federal de São Carlos, Brasil.
samanta_prado@hotmail.com

Resumen. La investigación verificó que a medida que la filosofía del lenguaje Pragmático (FLP) de L. Wittgenstein, ha sido considerado en la investigación de ciencias de la información (IC) del portugués, analizando con mayor precisión los trabajos que tienen la intersección de esta filosofía de base para la representación social de las memorias colectivas, en sistemas de información. Por lo tanto, se desarrolló una revisión de la literatura de las bases de datos bibliográficas/bibliotecas digitales/repositorios institucionales brasileños y portugueses, a saber: BRAPCI, BENACIB y BDTD (brasileño) y RCAAP (portugués). La escasez de la producción científica en la perspectiva epistemológica en portugués CI era evidente y por lo que es imposible analizar la influencia de los estudios informativos FLP en este país. En el recorte brasileño, se sumaron aproximadamente 92 investigaciones que mencionan la FLP en alguna perspectiva, siendo los trabajos de Luisa Rocha, aquellos que se fundamentaron en tal filosofía para tratar cuestiones relacionadas con la representación y memorias colectivas, evidenciando la validez de este abordaje en estudios informativos de esta naturaleza.

Palabras clave: Filosofía del lenguaje; pragmática; epistemología; representación de la información.

Abstract. The investigation verified that as the Philosophy of the Pragmatic Language (PPL) of L. Wittgenstein has been considered in the Portuguese-language Information Science (CI) investigations, analyzing in a more precise way the works that present the intersection of this philosophy as a foundation

for the social representation of collective memories, in information systems. For this purpose, a literature review was developed based on a bibliographical survey in the databases/digital libraries/institutional repositories in Brazil and Portugal: BRAPCI, BENACIB, and BDTD (Brazilian) and RCAAP (Portuguese). The scarcity of scientific production from an epistemological perspective in the Portuguese IC was evidenced making it impossible to analyze the influence of PPL in the informational studies in this country. In the Brazilian clipping, there were approximately 92 studies that mention PPL in some perspective, and the works of Luisa Rocha are those that were based on such philosophy to deal with questions related to representation and collective memories, showing the validity of this approach in the informational studies of this nature.

Keywords: Philosophy of language; pragmatic; epistemology; representation of information.

1 Introdução

Ludwig Wittgenstein (1889-1951), filósofo austríaco, defendeu em seu segundo momento filosófico, marcado pela produção da obra *Investigações Filosóficas* (1979), a imprevisibilidade na elaboração dos sentidos da linguagem, sinalizando que os sentidos se concretizam momentaneamente durante um jogo de linguagem (ação de comunicação específica e vinculada com uma forma de vida). O recorte desse trabalho se deu a partir da Filosofia da Linguagem Pragmática (FLP), considerando-se que a mesma é uma escola filosófica que se interessa pela experiência da linguagem em uso, em diferentes contextos de comunicação, marcados por variações e heterogeneidades logicamente imprevisíveis, assim como é uma abordagem assumida em diferentes áreas do conhecimento, enquanto enfoque elucidativo dos comportamentos de usos da linguagem (na *internet*, por exemplo). Contemporaneidade confirmada se observarmos a temática assumida no *30th International Wittgenstein Symposium*, ocorrido em Kirchberg, em 2007: *Philosophy of the Information Society*.

Assim, considera-se possível defender que a proposta wittgensteiniana, em relação ao seu entendimento sobre a FLP, é relevante e profícua para se pensar universo dos estudos informacionais, especialmente aqueles mediados por tecnologias, e que por sua vez tenham como propósito representar memórias constituídas e representadas coletivamente, assumindo-se, nesta perspectiva, tal universo informacional constituído de práticas coletivas de publicação, compartilhamento e indexação social de representações de memórias documentais no domínio da Ciência da Informação (CI).

Pinho (2017, p. 3), no contexto dos estudos sobre Organização do Conhecimento, observa que: “[...] os estudos terminológicos, terminográficos, de instrumentos de representação do conhecimento, bibliométricos, além de muitos outros, têm permitido compreender um domínio ou uma comunidade discursiva em momentos determinados historicamente. Com isso, podemos compreender e vislumbrar passado, presente e

futuro”. O conjunto de trabalhos apresentado no contexto desta obra, torna explícito o lugar central da Organização do Conhecimento enquanto disciplina preocupada em salvaguardar a memória, individual e coletiva, função solidificada na medida em que instrumentaliza, e dá condições de acesso às representações dos conhecimentos produzidos e registrados em âmbitos diversos.

A FLP foi introduzida na CI brasileira por Rendón Rojas (1996) e Gonzalez de Gomez (1996). Nesse contexto, ambos os autores partilham da defesa da Pragmática como ponto de partida prévio para os estudos informacionais, uma vez que os sujeitos e suas relações sociais (e de troca de informações) são estabelecidos a partir da linguagem em uso. Essas duas publicações demarcam o início desta corrente nos estudos epistemológicos da informação no Brasil, e na América Latina (Saldanha & Gracioso, 2014). Já, no plano internacional, Revez e Borges (2018) apresentam um panorama de produção científica que confirma a escassez de publicações sobre o assunto FLP na CI e, especificamente no contexto português, tal lacuna é reconhecida pela pesquisa de Borges, Freitas e Oliveira (2019).

A partir desse preâmbulo, consideramos que a abordagem pragmática da linguagem pela FLP colaboraria com a CI para se pensar os desafios da representação social das memórias coletivas (contextualmente localizada em sistemas e ambientes virtuais e abertos de informação) – e por isso indicamos como objetivo geral deste trabalho: averiguar em que medida a FLP tem sido considerada nas investigações da CI luso-brasileira, sendo este recorte sugerido para ampliar as fronteiras de investigação sobre o tema na CI em Língua portuguesa. Em decorrência disso, os seguintes objetivos específicos foram traçados: a) identificar e descrever o conteúdo das publicações científicas publicadas, dando ênfase a análise das produções publicadas no Brasil e em Portugal, sobre a FLP, na CI; b) categorizar e analisar as teorias e definições apresentadas nos trabalhos brasileiros e portugueses, em relação FLP, estabelecendo relações e críticas sobre suas aplicações, quando explicitadas; c) verificar e analisar, nos trabalhos investigados, como a FLP contribui para a reflexão, construção de recursos e instrumentos de tratamento/mediação da informação que promovam representação social das memórias coletivas em sistemas, ou redes, de informação.

A metodologia empregada para se atingir os objetivos foi a de análise de conteúdo clássica (Bauer, 2008), aplicada sobre os textos selecionados para identificação de presença, e categorização, da FLP nos estudos desenvolvidos, selecionados a partir de revisão de literatura em bases de dados/bibliotecas digitais/repositórios institucionais brasileiros e portugueses, compostos de: no contexto brasileiro, Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), repositório BENACIB, e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) (que fornecem, respectivamente, artigos publicados em periódicos científicos, trabalhos completos publicados em anais de evento, e teses e dissertações); e, no context Português, Repositório Científico de Acesso Aberto de Portugal (RCAAP) (que concentra todas as tipologias de fonte de informação). A estratégia de busca predominante em todas estas bases foi: “Wittgenstein” (enquanto

representante da abordagem) e, a partir dos resultados alcançados, foram se estabelecendo os filtros para CI.

2 Memória Coletiva

Para Halbwachs (1990), o processo de compartilhamento de experiências e valores com um grupo de referência é decisivo na constituição da memória, resultando, portanto, que tais resgates e reconstruções sejam influenciados pelo confronto de testemunhos, e pontos de vista, diferentes (movimentos que se estabelecem a partir do uso da linguagem, via jogos de linguagem). Estes confrontos podem, a partir desse prisma, ser tão internos aos indivíduos (confrontamento das experiências subjetivas do presente com as do passado) quanto interpessoais (confronto entre as percepções de um indivíduo em relação ao outro).

O ato de evocar a memória dá a ideia de que é preciso que as lembranças sejam comuns, compartilhadas com aqueles grupos sociais dos quais os indivíduos participam. Isso só é possível a partir do momento em que este indivíduo deixa vestígios de lembranças, permitindo que suas memórias sejam acessadas e usadas pelo grupo de que faz parte. O que tomamos como nota é que estas memórias só são ativadas se tivermos contato com o grupo, considerando tanto a importância do grupo, da ação e dos usos da linguagem para manutenção dessas memórias quanto a importância de sistemas abertos para indexação social, para sobrevivência das memórias coletivas (lembranças coletivizadas). É preciso, portanto, que as lembranças sejam constantemente reconstruídas e reconhecidas pelos grupos e comunidades, caso contrário, elas deixam de existir.

Observa-se que o grupo social (o grupo de referência, o coletivo de que o indivíduo faz parte) possui, segundo Halbwachs (1990), um papel importante no que se refere à atualização, e à complementação, dos testemunhos individuais. Trata-se, segundo Wittgenstein, do compartilhamento das formas e das regras de vida. Ou seja, os grupos sociais nos quais o indivíduo interage são os portadores da memória que foi construída, fruto das próprias relações estabelecidas entre os componentes desses grupos, o que resulta na construção de lembranças partilhadas, assim como de uma intrincada teia de experiências, vivências, observações, nostalgias, aprendizados; relações que são consequência de ver e perceber o mundo ao redor.

O que se tem, a partir dessas premissas, é que o indivíduo participa de vários grupos sociais, consequentemente, ele possui diversas memórias e, neste cenário coletivizado, base pela qual situamos a *Web* pragmática como ambiente que potencializa (pela interação e colaboração) o acesso e a participação de sujeitos informacionais em uma multiplicidade de grupos sociais, interferindo na construção de suas memórias coletivas e individuais. Assim, há uma nova dinâmica de produção de memórias coletivas sendo gerada, produção alicerçada em usos de tecnologias colaborativas mediadoras.

Dessa premissa é que se concatena grupos de referência, formas de vida, memórias coletivas e as tecnologias mediadoras da organização do conhecimento no ambiente da *Web*, considerando-o ambiente/motor contemporâneo deste movimento,

justificando uma análise inicial da presença da FLP na CI brasileira e portuguesa, procurando evidenciar se este campo tem embasado, em alguma medida, a compreensão sobre os fluxos que regem a construção e circulação das memórias coletivas.

3 Seleção e Análise

Tendo-se como ponto de partida os resultados de Borges, Freitas e Oliveira (2019), que apontam existência de uma lacuna considerável de pesquisa de natureza teórica na CI portuguesa, prossegue-se no recorte específico para análise da produção científica sobre a FLP, especificamente a desenvolvida por Wittgenstein, nas bases de dados selecionadas, nas seguintes condições: na BDTD brasileira, total aproximado de 510 mil teses e dissertações, com 304 (0,059608 %) pesquisas sobre o filósofo, dentre as quais, somente seis (1,97%) estavam relacionadas diretamente à CI; já na RCAAP, base portuguesa, aproximadamente 1 milhão e setecentos mil registros de naturezas diversas, 212.937 teses de doutorado e 642.793 dissertações de mestrado, sendo desses 855.730 trabalhos identificados 434 (0,050717%) teses e dissertações relacionadas ao filósofo, sendo que destes, somente 133 são frutos de trabalhos portugueses (301 são pesquisas brasileiras que também disponibilizadas nesta base), não sendo identificados trabalhos específicos sobre Wittgenstein na CI portuguesa, as explicitando o interesse no mesmo pelas investigações brasileiras.

Um relato minucioso sobre os estudos do filósofo no Brasil é apresentado por Darley Dall'Agnol (2008), sendo um dos capítulos de Arley Ramos Moreno (2008) dedicado a tratar da construção de uma pragmática filosófica. Observa-se, complementarmente, que os trabalhos de Wittgenstein foram introduzidos na Filosofia brasileira somente em meados de 1950, mas ganharam espaço em diferentes níveis de graduação e pós-graduações.

No que tange aos limites desse trabalho, na BRAPCI foram identificadas 10 pesquisas relacionadas a Wittgenstein, não sendo identificada nenhuma diretamente relacionada às questões de memória, ainda que parte deles fossem dedicados a elementos relacionados ao tema e à Organização do Conhecimento. É importante sinalizar que a BRAPCI também indexa periódicos portugueses, mas dentre os trabalhos identificados, todos foram publicados em fontes originalmente brasileiras. Especificando o recorte da amostra para trabalhos publicados em anais de eventos científicos em CI (que comportam a apresentação de trabalhos brasileiros, mas também portugueses), constatamos que durante o ENANCIB (Encontro Nacional de Pesquisa em Ciência da Informação) foram apresentadas 74 pesquisas que mencionam, em alguma perspectiva, o filósofo da linguagem. Esse movimento ascendente nos permite prospectar uma continuidade na produção de pesquisas que transitam no horizonte da FLP, que, potencialmente, tem a capacidade de gerar, cada vez mais, novos agentes epistêmicos sobre o assunto.

As primeiras publicações no ENANCIB, contemplando a filosofia de Wittgenstein, datam de 2001. Posteriormente, em 2003, foram publicados seis trabalhos; em 2005, um; em 2006, dois; em 2007, oito; em 2008, cinco; em 2009, dois; em 2010, três; em 2011, nove; em 2012, 11; em 2013, cinco; em 2014, 12; em 2015, seis; em 2016, 11. Para análise dos autores que tiveram publicações no ENANCIB, com menções a Wittgenstein, optamos por espelhar de modo descritivo, apenas aqueles que tiveram quatro trabalhos ou mais apresentados no evento ao longo do período. Evidenciando-se a predominância de trabalhos produzidos por Gustavo Silva Saldanha (13), Dra. Maria Nélida Gonzalez de Gomez (9), Luaciana de Souza Gracioso e Luisa Maria Rocha (4), que foram suas orientandas da Profa. Nélida.

Também se verificou a presença do filósofo austríaco nos trabalhos publicados nos anais da ISKO Brasil, desde sua primeira edição em 2011, distribuindo-se tais publicações na seguinte ordem: 2012 (0); 2013 (3); 2015 (03); 2017 (2). Observando-se não haver trabalhos dessa envergadura para as questões relacionadas a memória, optamos por desenvolver uma análise mais detalhada desta produção, decidindo lançar, em pesquisas futuras, abordagens comparativas com outros capítulos da ISKO. De momento, observa-se a seguinte distribuição de pesquisas sobre Wittgenstein na ISKO, categorizadas por grupos de trabalho (GTs), nas seguintes condições: GT1 *Estudos Históricos e Epistemológicos da Informação* (41), confirmando-se, assim, a presença majoritária em pesquisas que fundamentam o campo; GT2 *Organização e Representação do Conhecimento* (28), o que reforça nosso pressuposto de que a filosofia da linguagem wittgensteiniana propicia elementos que podem subsidiar o desenvolvimento não só de teorias, mas também de práticas relacionadas a organização e representação da informação; e, GT9 *Museologia, Patrimônio e Informação* (5), relacionadas em alguma medida, à memória.

Com o intuito de melhor contextualizar os dados apresentados, no que diz respeito a produção científica brasileira e portuguesa, de modo geral, foi verificado que no Brasil, segundo a Plataforma Lattes (com mais de três milhões de registros de currículos de pesquisadores brasileiros e internacionais), somam-se 853 pesquisadores que indicaram como assunto de suas pesquisas “Wittgenstein”, cabendo enfatizar, contudo, que no âmbito da CI brasileira, quando se cruzou essa busca por assunto e área de atuação, foram recuperados 19 pesquisadores que indicaram o filósofo como tema.

Em Portugal, segundo a Plataforma DeGóis - desenvolvida e mantida pelo MEC (Ministério da Educação e Ciência), por meio da FCT (Fundação para a Ciência e a Tecnologia) -, que reúne os currículos de pesquisadores portugueses, foi possível identificar que dos 28.574 registros existentes até 2019, 53 produções científicas indicam o filósofo em algum aspecto de seu conteúdo. No entanto, dos pesquisadores identificados, não foi localizada nenhuma relação com a área de CI.

Estes dados nos servem como sinalização sobre o estado da arte em relação ao engajamento de pesquisadores com o filósofo, no âmbito brasileiro e português, evidenciando que: pela Plataforma DeGóis os pesquisadores em Wittgenstein correspondem a 0,1855% do total de pesquisas registradas na Base, não sendo possível constatar, entretanto, a partir da análise individual de cada registro, a

vinculação destes pesquisadores e de suas pesquisas com a área de CI; e, pela Plataforma Lattes, este percentual é de 0,0244% (percentual bem aquém do verificado nas pesquisas portuguesas), cabendo, no entanto, a ressalva de que ao se estabelecer o recorte para pesquisas sobre Wittgenstein na CI brasileira, constata-se que 2,2274% das pesquisas desenvolvidas sobre o filósofo foram feitas por pesquisadores atuantes nesta área.

4 Discussões

A partir do montante de trabalhos sistematizados, não foram identificados especificamente pesquisas que dessem conta de apresentar pontualmente a FLP proposta por Wittgenstein, enquanto fundamento para ações de Organização do Conhecimento envolvendo práticas voltadas a representação de memórias coletivas. Neste sentido, alguns objetivos específicos propostos, assim como seus respectivos métodos não foram desenvolvidos. Contudo, um destaque pode ser dado aos trabalhos produzidos e apresentados por Luisa Maria Rocha, no âmbito do GT9 do ENANCIB, que não se dedica propriamente a Organização do Conhecimento como ponto de partida, mas que, ao apresentar elementos que dizem respeito ao arranjos de museus para promoção de interação e apropriação da informação, perpassam também por esta disciplina.

Assim, especialmente pelo trabalho intitulado *Pragmatica: uma nova matriz para os museus de ciência?* (2014), apresenta-se clara articulação das ideias de Jogos de Linguagem, Formas de Vida e Regras de Vida, propostas por Wittgenstein, que podem ampliar a matriz gnosiológica da CI, sobretudo pela ordem museológica, e no que diz respeito aos estudos das comunidades, das coletividades e das possibilidades de suas representações em ambientes de informação. Rocha (2014) não toca no tema da memória coletiva como objeto determinado de investigação, mas ao se dedicar ao Museu, dedica -se também ao seu local de organização e produção.

Ao se analisar este espaço de memória à luz da filosofia pragmática de Wittgenstein, sugere-se na “cultura material” um espaço de inserção passível de ampliação: “[...] uma vez que a variabilidade de uso das representações no mundo da vida possibilita o alargamento do significado de forma a se tornarem importantes nos jogos de linguagem partilhados pelas diferentes pessoas e suas formas de vida. (Rocha, 2014, p. 4).

Ao concluir suas investigações, a autora enfatiza que no plano do discurso museológico, refletir sobre a informação significa trabalhar a transversalidade do seu fluxo na abertura de novos sentidos e caminhos: “[...] Ao deslocar-se o seu lugar para o mundo de vida, é privilegiada a informação em sua heterogenia, a possibilidade de criação de novos sentidos oriundos da própria integração social da informação” (Rocha, 2014, p.18). Havendo, assim, uma defesa clara sobre a importância de procurar evidenciar, cada vez mais, junto aos ambientes informacionais, a integração social enquanto elemento constitutivo das próprias dinâmicas de Organização do

Conhecimento, especialmente nos espaços dedicados a organização e compartilhamento de memórias.

5 Considerações finais

Foi objetivo deste trabalho investigar em que medida a FLP wittgensteiniana se apresenta como aporte teórico à representação da memória coletiva, nos estudos informacionais (CI), especialmente desenvolvidos na comunidade científica brasileira e portuguesa, procurando apresentar um quadro que espelhasse o estado dessa produção no referido contexto.

Os objetivos específicos foram traçados com o intuito de identificar e descrever o conteúdo das publicações científicas publicadas, dando ênfase a análise das produções sobre a FLP, na CI, publicadas no Brasil e em Portugal. Sobre este objetivo, a análise comparada se tornou inviável uma vez que não foram localizadas ainda, produções científicas sobre o filósofo na CI portuguesa, impossibilitando, por conseguinte, o objetivo específico de categorizar e analisar as teorias e definições apresentadas nos trabalhos brasileiros e portugueses em relação FLP, estabelecendo relações e críticas sobre suas aplicações, quando explicitadas.

Já o objetivo de verificar e analisar, nos trabalhos investigados, como a FLP contribui para a reflexão, construção de recursos e instrumentos de tratamento e mediação da informação, promovendo a representação social das memórias coletivas em sistemas ou redes de informação pode ser parcialmente alcançado uma vez os trabalhos desenvolvidos por Luisa Rocha, tocam especificamente no entroncamento sugerido como hipótese deste trabalho, e que diz respeito a defesa de que a FLP oferece subsídios teóricos relevantes para a construção reflexiva e desenvolvimento de um conjunto de práticas que concatenem cada vez mais a natureza coletiva de construção de saberes e memórias, aos ambientes e sistemas voltados a organização do conhecimento.

Referencias

- Bauer, M. W. (2008). Análise de conteúdo clássica: uma revisão. In: Bauer, M. W. & Gaskell, G. *Pesquisa qualitativa com texto, imagem e som: um manual prático* (Cap. 8, pp. 189-217). Rio de Janeiro: Editora Vozes.
- Borges, M. M.; Freitas, M. C. V.; & de Oliveira, S. R. (2019). A Ciência da Informação em Portugal nas primeiras décadas do século xxi: uma abordagem preliminar para uma cartografia iberoamericana. *Bibliotecas. Anales de Investigación*; 15(1), 260-292.
- Dall'Agnol, D. (Org.) (2008). *Wittgenstein no Brasil*. São Paulo: Editora Escuta.
- González de Gómez, M. N. (1996). Da organização do conhecimento às políticas de informação. *INFORMARE - Cadernos do Programa de Pós-Graduação em Ciência da Informação*, 2(2), 58-66.

- Halbwachs, M. (1990). *A memória coletiva*. São Paulo: Editora Revista dos Tribunais Ltda.
- Moreno, A. R. (2008). Filosofia da linguagem e terapia: em direção a uma pragmática filosófica. In: Dall'Agnol, D. (org.) *Wittgenstein no Brasil*. São Paulo: Escuta.
- Pinho, F. A. (2017). Prefácio. In: Pinho, F. A. & Guimarães, J. A. C. (org.) *Memória, tecnologia e cultura na organização do conhecimento*. Recife: Ed. UFPE, 2017.
- Rendón Rojas, M. Á. (1996). Hacia um nuevo paradigma em bibliotecologia. *Transinformação*, 8(3), 17-31. Disponível em: <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/1598>
- Revez, J. & Borges, L. C. (2018). Pragmatic paradigm in information science research: a literature review. In *10th Qualitative and Quantitative Methods in Libraries International Conference* (pp. 1–8). Chania, Crete: ISAST.
- Rocha, L. M. (2014). Pragmatica: uma nova matriz para os museus de ciência? In *14. Encontro Nacional de Pesquisa em Ciência da Informação* (pp. 1-21). Santa Catarina: ENANCIB. Disponível em: <http://enancib.ibict.br/index.php/enancib/xivenancib/paper/viewFile/4582/3705>
- Saldanha, G. S. & Gracioso, L. S. (2014). Filosofia da linguagem e Ciência da Informação na América Latina: apontamentos sobre pragmática e linguagem ordinária. In: Rendón Rojas, M. Á. (Org.). *El problema del lenguaje en la bibliotecología/ciencia de la información/documentación: un acercamiento filosófico-teórico* (pp. 1-32). Cidade do México: UNAM.
- Wittgenstein, L. (1979). *Investigações Filosóficas*. 2o ed. São Paulo: Abril Cultural.

Panorama teórico-analítico-sintético sobre a adoção de facetas no contexto da organização do conhecimento

Márcio Bezerra da Silva¹ y Zeny Duarte de Miranda²

¹ ORCID [0000-0002-0052-7174](https://orcid.org/0000-0002-0052-7174). Faculdade de Ciência da Informação, Universidade de Brasília, Brasil.

marciobdsilva@unb.br

² ORCID [0000-0003-0365-6905](https://orcid.org/0000-0003-0365-6905). Instituto de Ciência da Informação, Universidade Federal da Bahia, Brasil.

zenydu@gmail.com

Resumen. Apresenta um panorama de uso de facetas na organização da informação, entre ambientes físicos e digitais. Fundamenta-se no contexto da organização do conhecimento, encarada como o arcabouço mais importante na teoria da organização da informação, e enfatiza-se na teoria da classificação facetada do indiano Ranganathan (1892-1972). Aborda três pontos básicos de discussão sobre a teoria ranganathiana: as cinco categorias fundamentais, personalidade-matéria-energia-espaco-tempo; a multidimensionalidade sistemática; e os cânones de renques e de cadeias de classes. Caracteriza-se como uma pesquisa exploratória, para identificar ambientes que usam facetas, e bibliográfica. Resulta que a maioria dos modelos de facetação está aplicado em ambientes digitais, que são modelos puramente acadêmico-profissionais, que os sistemas documentários (físicos) foram os suportes de aplicação mais adotados, que se identificaram modelos no período de 1933 até 2017 e que a categoria fundamental [Personalidade] foi a mais citada. Concluiu-se que a presença de facetas em sistemas de organização do conhecimento promulga que um mesmo produto tenha interpretações diferentes em uma mesma taxonomia, formaliza a multidimensionalidade idealizada na teoria da classificação facetada de Ranganathan, mas sem segui-la na íntegra por vezes, adota mnemônicas personalizadas e mostra que os ambientes onde são adotadas variam entre físicos e digitais, das bibliotecas físicas aos repositórios digitais.

Palabras clave: Organização da informação. Organização do conhecimento. Teoria da classificação facetada. Facetas. Categorias fundamentais.

Abstract. Presents an overview of the use of facets in the organization of information, between physical and digital environments. Is based on the context of the organization of knowledge, which is regarded as the most important framework in the theory of information organization, and is emphasized in Ranganathan's theory of faceted classification (1892-1972). Addresses three

basic points of discussion about Ranganathan's theory: the five fundamental categories, personality-matter-energy-space-time; systematic multidimensionality; and the canons of ranks and class chains. Is characterized as an exploratory research, to identify environments that use facets, and bibliographical. The results are that most of the factional models are applied in digital environments, which are purely academic-professional models, that the documentary (physical) systems were the most adopted application supports, that models were identified in the period from 1933 to 2017 and that the fundamental category [Personality] was the most cited. Was concluded that the presence of facets in systems of knowledge organization promulgate that the same product have different interpretations in the same taxonomy, formalizes the multidimensionality idealized in Ranganathan's theory of the faceted classification, but without being followed in full in some cases, adopts custom mnemonics and shows that the environments where they are adopted vary from physical to digital, from physical libraries to digital repositories.

Keywords: Information organization. Knowledge organization. Theory of the faceted classification. Facets. Fundamental categories.

1 Introdução

Organizar pode ser considerado um ato intrínseco ao ser humano. Enquanto uma atividade que vem de longa data, a humanidade procura organizar o que se encontra em sua volta, entre pessoas, animais, objetos e fatos, segundo características específicas. Com o passar dos tempos, a ânsia por organizar o cotidiano passou a convergir, cada vez mais, à necessidade de tomar posse da informação que circunda o homem, ao enxergá-la como um insumo para a tomada de decisão. Neste sentido, caminha-se ao contexto da organização do conhecimento (OC), na qual encontram-se estudos que investigam a multidimensionalidade como um caminho para organizar a informação, inicialmente em ambientes físicos, mas que transcenderam aos digitais, ao acompanhar a dinamicidade imposta pelas mudanças na sociedade humana.

Pensar desta forma é ir ao encontro da teoria da classificação facetada (TCF), do indiano, matemático e bibliotecário Shiyali Ramamrita Ranganathan (1897-1972), ao trazer as cinco manifestações genéricas e distintas para classificar um recurso bibliográfico, comumente chamadas de facetas. De posse do objetivo de “[...] denotar qualquer componente de um domínio [...]” (GOMES; MOTTA; CAMPOS, 2006), as cinco manifestações, instruídas na obra intitulada *Prolegomena to Library Classification*, correspondem a mnemônica PMEST, traduzida em personalidade [P], matéria [M], energia [E], espaço [S] e tempo [T] (RANGANATHAN, 1960, 1963, 1967).

Apesar da compreensão de que a classificação facetada, enquanto uma proposta de categorização que ganhou notoriedade a partir da TCF do indiano Ranganathan, permite que um único assunto seja classificado em hierarquias diferentes a partir de múltiplos pontos de subjetividade, objetiva a estruturação do conhecimento, permite o

mapeamento de uma área de assunto e inclui novos conceitos sem alterar a estrutura do sistema, dúvidas surgem sobre as formas de facetação criadas para a OI, das bibliotecas físicas à contemporaneidade dos ambientes digitais, tanto comerciais, quanto acadêmicos.

Seguindo um percurso metodológico formalizado pelas técnicas de pesquisa exploratória e bibliográfica, e de coleta de dados em livros, artigos científicos, dissertações, teses e *websites* acadêmicos e comerciais, o presente artigo adota como objetivo geral o interesse em apresentar um panorama de uso de facetas na OI, entre ambientes físicos e digitais. Especificamente, verificar suportes físicos e digitais onde os modelos de facetação foram adotados, apresentar a evolução dos ambientes ao longo dos anos, identificar facetas à luz da mnemônica PMEST e elencar um exemplo de classificação facetada entre os modelos identificados.

2 Teoria da Classificação Facetada: pontuações básicas

A TCF baseia-se na ideia de *two-fold infinity*, ou seja, em duas visões, sendo uma pensada no leitor, correspondente as coleções de documentos, e outra pela ótica do universo de ideias a serem organizadas com base nas características dos documentos. Na verdade, Ranganathan defende a multiplicidade de visões de mundo e de necessidades informacionais, além de compreender que a natureza do universo do conhecimento é dinâmica. Esta dualidade permitirá características de flexibilidade e de capacidade de resposta, que inclusive poderá ser usada como medida de sucesso de qualquer sistema (RANGANATHAN, 1967).

Quantos aos postulados, também chamados de categoriais fundamentais (PMEST), referem-se a um termo genérico usado para denotar algum item, como um assunto básico¹ ou um isolado², além de formar renques, termos e números (RANGANATHAN, 1967). Ainda segundo Ranganathan, “cada faceta de qualquer assunto, bem como cada foco nela, é reagida como uma manifestação de uma das cinco **Categorias Fundamentais** [...]” (RANGANATHAN, 1963, p. 1*25, grifo do autor).

A CC, enquanto o primeiro sistema desenvolvido na perspectiva dinâmica, deflagrando a facetação como um novo tipo de classificação bibliográfica (CAMPOS, 2001), adota princípios para a confecção de sistemas multidimensionais e organizar a informação. Neste caso, ao tomar como base o plano das ideias (conceitos, mensagens), pois o cerne discursivo do artigo está no nível conceitual, elencam-se os cânones que afetam diretamente os arranjos taxonômicos, referentes a ordenação dos conceitos e as relações entre si. Entre eles, Gomes, Motta e Campos (2006) destacam

-
- 1 É um elemento sistematizado de ideias correspondentes a um campo (RANGANATHAN, 1967), ou seja, são “[...] áreas mais abrangentes do conhecimento, como Matemática, Agricultura” (CAMPOS, 2001, p. 49).
 - 2 É um elemento adotado para formar um componente de um assunto, mas que não pode ser considerado um assunto em si. Campos (2001) cita a ideia isolada Milho, que ao ser combinado com o assunto básico Agricultura define-se o assunto Agricultura do Milho.

os cânones de renques de classes e de cadeias de classes.

A formação dos renques (*arrays*), também denominados de ciclos (*rounds*), trata-se de classes derivadas de um universo que possui uma única característica de divisão e assim estabelece-se uma ordenação na sequência preferida (RANGANATHAN, 1967). Os renques (figura 1) são arranjos em “[...] séries horizontais. Por exemplo: Macieira e Parreira são elementos da Classe Árvore Frutífera, formada pela característica de divisão – tipo de árvores frutíferas” (CAMPOS, 2001, p. 51).

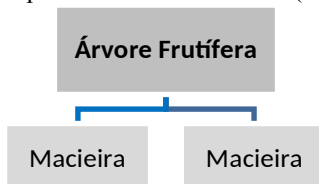


Figura 1. Renque, série horizontal.

A formação das cadeias (*chain*), do mesmo modo chamadas de níveis (*levels*), seria uma sequência formada por classes em um universo sequencial (1, 2, 3 e assim sucessivamente) até um ponto desejado (RANGANATHAN, 1967). Estruturando um arranjo em cadeias (figura 2) formam-se “[...] séries verticais de conceitos [...] conforme a cadeia descendente ou ascendente. Por exemplo: Macieira é um tipo de Árvore Frutífera, que, por sua vez, é um tipo de Árvore. Neste exemplo, observa-se uma cadeia descendente” (CAMPOS, 2001, p. 51).

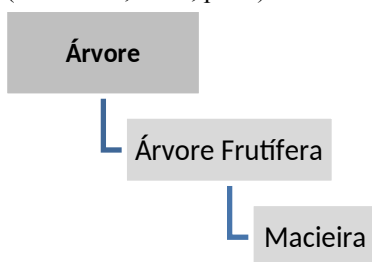


Figura 2. Cadeia, série horizontal.

Quanto mais horizontal for o sentido da taxonomia, maior será a exaustividade nas categorias fundamentais, e, por vezes, de facetas, apresentando características que estejam no mesmo nível de classificação. Por outro lado, quanto mais vertical for a estrutura da OI, mais profunda será o arranjo de conceitos, ou seja, uma especificidade mais acentuada, no sentido do mais amplo (*top down*) ao mais específico (*bottom up*).

3 Resultados da pesquisa

Em um caminho iniciado na organização de acervos físicos e que chegou ao mundo digital, as facetas se tornaram um viés aos sistemas de organização do conhecimento

(SOC) multidimensionais, reconhecidos em sistemas documentários e tesouros, ambos em suporte físico, assim como em banco de dados (BD), *software*, biblioteca digital, lista eletrônica e tesouro digital, gerando a totalidade de 24 modelos de facetação identificados.

Encarados como esquemas que permitem a organização, o gerenciamento e a recuperação de informações (VICKERY, 2008), os SOC têm o seu núcleo de funcionamento lógico-relacional na classificação, vista por Ranganathan (1967, p. 395) como uma tarefa que busca “[...] mapear o universo multidimensional dos assuntos [...]”. Ranganathan enxergava a classificação como uma ação de versatilidade sistêmica e que deveria acompanhar a evolução do conhecimento ao respeitar a dinamicidade e infinitude entre os conceitos, pensamento este que ia de encontro aos esquemas bibliográficos desenvolvidos na década de 1930 diante de suas dificuldades quanto a hospitalidade de novos assuntos no sistema.

A ideia de construir um sistema versátil está representada por 24 (100%) modelos de facetação identificados na literatura e que se concentram em maior número na perspectiva dos ambientes digitais segundo o quantitativo de 54% (13) dos modelos, enquanto que 46% (11) respondeu aos espaços físicos.

Especificando os suportes de aplicação nos 24 modelos de facetação, sete suportes de aplicação foram identificados, sendo cinco de maneira digital e dois no viés físico. Apesar da maior quantidade de ambientes estar direcionada ao âmbito digital, o maior resultado do gráfico um (1) redireciona-se à perspectiva física, pois 36% (8) representa os suportes em sistemas documentários (físicos), realidade comum nas bibliotecas, enquanto que os menores valores corresponderam aos suportes *software* e lista eletrônica, quantificados por 4% (1) cada uma.

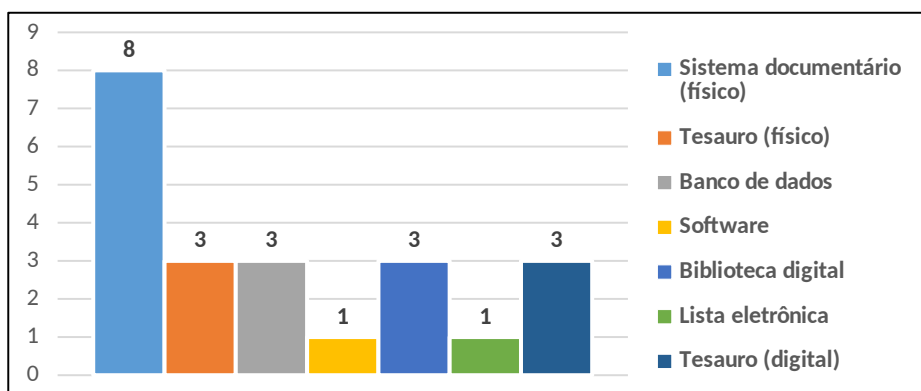


Gráfico 1. Suportes de aplicação.

Quanto ao seguimento evolutivo dos SOC facetados foi possível perceber que a linha se inicia em 1959 e mantém-se até os tempos atuais, mostrando uma nítida continuidade dos estudos iniciados por Ranganathan na década de 1930, bem como marca a transcendência da aplicação desse tipo de sistema aos ambientes digitais

(gráfico 2). Especificamente, os SOC facetados no âmbito digital iniciam-se em 2003, enquanto que os três últimos sistemas identificados no viés físico são de 2006, data que marca a transcendência dos modelos físicos para digitais. Contudo, todos os modelos de facetação correspondem ao escopo acadêmico-profissional, resultado que pode ser compreendido pela cognição de aproximação automática entre a OC e as bibliotecas (físicas e digitais), no sentido de organizar o conhecimento humano no universo da academia, entre livros, periódicos científicos, monografias, dissertações, teses etc.

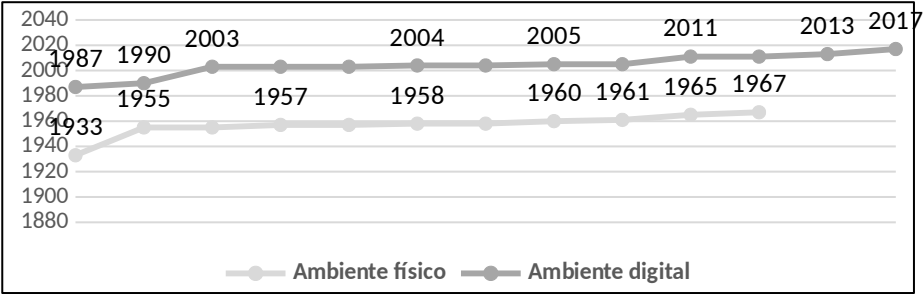


Gráfico 2. Evolução dos ambientes ao longo dos anos.

No que se referem as categorias fundamentais adotadas ao longo da breve evolução (gráfico 2), do ambiente físico ao digital, no total das 163 categorias identificadas e comparadas a mnemônica PMEST, a categoria [P] possui a maior quantidade de citações, correspondendo a 39% (64), enquanto que o menor valor satisfaz a categoria [T], valorada em 7% (11) do total (gráfico 3). Além disso, o resultado por ser explicado tanto pelo conceito de cada categoria na mnemônica ou pela sequência (útil) de apresentação, ou seja, a ordem PMEST.

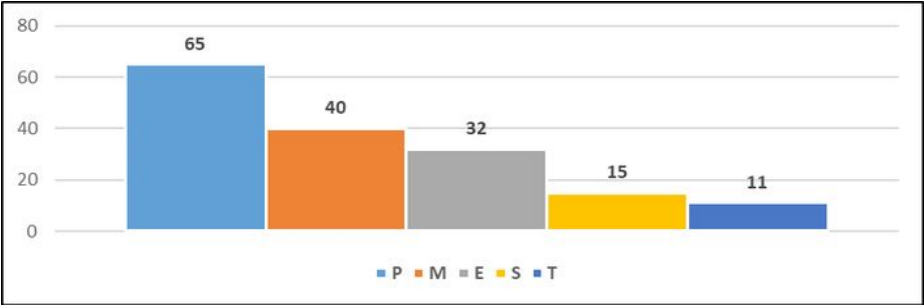


Gráfico 3. Categorias Fundamentais à luz do PMEST³.

Os resultados mostraram que a quantidade preponderante de modelos de facetação foi aplicada em ambientes digitais; que os sistemas documentários (físicos) foram os suportes de aplicação mais adotados; que a realização de modelos facetados no

3 Valores: 39% do total corresponde a [P], 25% a [M], 20% a [E], 9% a [S] e 7% a [T].

contexto da OC iniciou-se em 1933 e mantêm-se até os tempos atuais, sendo que os modelos digitais se iniciaram em 1987, data em que marca a transposição definitiva do físico ao digital, conforme identificações na literatura; que a categoria [P] foi a mais presente na mnemônica PMEST; e a adoção das categorias fundamentais estão em consonância com a sequência útil.

Adotando como exemplo um protótipo de BD (facetado), a sequência útil, como uma espécie de adaptação da utilizada na TFC, está representada pela tipologia [P]; matéria [M], como propriedade (qualidade) e matéria (substância); ação [E]; lugar [S] e tempo [T] (figura 3) (DA SILVA; NEVES, 2016). No protótipo, na classificação do livro “Introdução a Microinformática”, foi gerada a sequência facetada “[P],[M],[M2],[T]”, apresentando um segundo nível de renques entre “[M],[M2]” (matéria-propriedade). Alterando [T] por um novo [P], gerou-se a sequência “[P],[M],[M2],[2P]”, isto é, um segundo cliço de cadeias entre “[P],[2P]”.

Figura 2. Exemplo de classificação facetada.

Fazendo a leitura (básica) da sequência “[P],[M],[M2],[T]”, para classificar a supracitada obra passou-se pela primeira rodada da personalidade [P], primeira rodada da matéria [M], segunda rodada da matéria [M2] e primeira rodada do tempo [T]. Quanto a sequência “[P],[M],[M2],[2P]”, passou-se pela primeira rodada da personalidade [P], primeira rodada da matéria [M], segunda rodada da matéria [M2] e segundo nível da personalidade [2P].

4 Considerações finais

O presente trabalho apresentou um panorama onde destaca a importância da OI para a recuperação da informação, independentemente de o ambiente ser físico ou digital. Para tanto, o cerne aqui problematizado se deu na atividade da classificação, especialmente nos sistemas desenvolvidos com base em facetas. Trata-se de um anseio que encontra subsídios discursivos no campo da CI e desenvolvido em pesquisas realizadas no âmbito da OC. Neste sentido, em uma literatura identificada

no contexto da OC foi possível identificar que a maioria dos modelos de facetação está aplicado em ambientes digitais, que são modelos puramente acadêmico-profissionais e não comerciais, que os sistemas documentários (físicos) foram os suportes de aplicação mais adotados, que foram identificados modelos no período de 1933 (TCF de Ranganathan) até 2017 e a categoria fundamental [P] foi a mais citada entre as dimensões do PMEST.

Concluiu-se que a presença de facetas em SOC promulga que um mesmo produto tenha interpretações diferentes em uma taxonomia navegacional, formalizam a multidimensionalidade idealizada na TCF de Ranganathan, mas sem segui-la na íntegra por vezes, adotam mnemônicas personalizadas e mostram que os *websites* assumiram a facetação como um meio que não se limita a organizar a informação, mas que influencia nas formas de navegação e busca de informação nas interfaces de ambientes digitais.

Referencias

- Campos, m. L. A. (2001). *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: Ed. da UFF.
- Da Silva; M. B; Brito, d. N. (2016). *Revista.Ibero-Americana de Ciência da Informação*, 9(1), 242-257. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/2236/1984>
- Gomes, h. E.; Motta, d. F.; Campos, M. L. de A. (2006). Revisitando Ranganathan: a classificação na rede. *Biblioteconomia, Informação & Tecnologia da Informação (BITI)*. Disponível em: <http://www.conexaorio.com/bitl/revisitando/revisitando.htm#canones>
- Mills, J. (2004). Faceted classification and logical division in information retrieval. *Library Trends* 52(3) Winter, 541-570, Disponível em: <https://www.ideals.illinois.edu/bitstream/handle/2142/1687/Mills541570.pdf?sequence=2&isAllowed=y>
- Ranganathan, S. R. *Colon Classification*. Bombay: Asia Publishing House, 1963.
- Ranganathan, S. R. *Prolegomena to library classification*. Bombay: Asia Publishing House, 1967.
- Ranganathan, S. R. *Reference service*. Bombay: Asia Publishing House, 1961.
- Satija, M. P. (2017). *Colon Classification (CC)*. Encyclopedia of Knowledge Organization. Disponível em: http://www.isko.org/cyclo/colon_classification
- Vickery, B. C. (2008). *On 'knowledge organisation'*. Disponível em: <http://web.archive.org/web/20080404103206/www.lucis.me.uk/knowlorg.htm>

Emanuele Tesauro nos trópicos: caminhos de uma tradução e crítica teórico-histórica em organização do conhecimento

Gustavo Silva Saldanha¹, Tatiana de Almeida² y Maria Helena Teixeira da Silva³

¹ ORCID [0000-0002-7679-8552](https://orcid.org/0000-0002-7679-8552). IBICT; UNIRIO, Rio de Janeiro, Brasil.
gustavosalldanha@ibict.br

² ORCID [0000-0002-1703-0148](https://orcid.org/0000-0002-1703-0148). UNIRIO, Rio de Janeiro, Brasil.
tatiana.almeida@unirio.br

³ ORCID [0000-0001-5330-3361](https://orcid.org/0000-0001-5330-3361). UFF PURO, Rio de Janeiro, Brasil.
maria_helena@id.uff.br

Resumen. A presente reflexão, de natureza teórica, procura compreender o percurso de estudos brasileiros sobre Emanuele Tesauro, a partir da perspectiva epistemológico-histórica e da filosofia da linguagem, levando à (necessária) experiência interpretativa de tradução do teórico. O corpus da proposta está focado no parágrafo introdutório do Índice Categórico, discurso nevralgico que coloca Tesauro no coração da construção histórica da organização do conhecimento e, ao mesmo tempo, na vanguarda dos desafios atuais. O Índice tesauroiano está presente no capítulo terceiro, *Cagioni Instrumentali delle argutezze oratorie simboliche et lapidarie*, de sua obra *Il Cannocchiale Aristotelico*, publicado em 1670, tratado essencial para a compreensão moderna da metáfora como possibilidade de compreensão do mundo através da linguagem. Tesauro considera seu trabalho como uma relação entre oratória, lapidária e simbólica, baseada nos princípios da retórica aristotélica.

Palabras clave: Emanuele Tesauro. Epistemologia histórica. Organização do conhecimento. Tradução.

Abstract. The theoretical research seeks to understand the course of Brazilian studies on Emanuele Tesauro, from the epistemological-historical perspective and the philosophy of language, leading to the (necessary) interpretative experience of translating the theoretical. The proposal's corpus is focused on the introductory paragraph of the Categorical Index, a key discourse that places Tesauro at the heart of the historical construction of the organization of knowledge and, at the same time, at the forefront of current challenges. The Thesaurian Index is present in the third chapter, *Cagioni Instrumentali delle argutezze oratorie simboliche et lapidarie*, from his work *Il Cannocchiale Aristotelico*, published in 1670, an essential treatise for the modern understanding of metaphor as a possibility of understanding the world through language. Thesaurus considers his work as a relationship between oratory, lapidary and symbolic, based on the principles of Aristotelian rhetoric.

Keywords: Emanuele Tesauro. Historical epistemology. Knowledge organization. Translation.

1 Introdução

Após uma série de estudos histórico-epistemológicos, voltados à Organização do Conhecimento (OC), que chamam a atenção para o pensamento de Emanuele Tesauro, teórico do século XVII estudioso da retórica, destaca-se a urgência de se trazer para o meio acadêmico a discussão sobre a importância da tradução, como forma de atualização, de sua obra. A presente reflexão, de natureza teórica, procura compreender o percurso de estudos brasileiros sobre Tesauro, personagem central do Seiscentos para os estudos da linguagem, desenvolvidos centralmente nos últimos anos, a partir da perspectiva epistemológico-histórica e da filosofia da linguagem, levando à (necessária) experiência interpretativa de tradução do teórico.

O corpus da proposta está focado no parágrafo introdutório do Índice Categórico, discurso nevrálgico que coloca Tesauro no coração da construção histórica da OC e, ao mesmo tempo, na vanguarda dos questionamentos e dos desafios de inovação do domínio na contemporaneidade. O Índice tesauroiano está presente no capítulo terceiro, *Cagioni Instrumentali delle argutezze oratorie simboliche et lapidarie*, de sua obra *Il Cannocchiale Aristotelico*, publicado em 1670, tratado essencial para a compreensão moderna da metáfora como possibilidade de compreensão do mundo através da linguagem. Tesauro considera seu trabalho como uma relação entre oratória, lapidária e simbólica, baseada nos princípios da retórica aristotélica.

O estudo da obra e da teoria de Emanuele Tesauro na OC ainda é escasso e inicial. Dessa forma, acreditamos haver lacunas em relação aos fundamentos da OC que poderiam ser contextualmente embasadas, caso fosse facilitada a exploração dos pensamentos de Tesauro. Uma das formas capazes de possibilitar esse aprofundamento teórico e epistemológico é a tradução de sua obra, a qual se pode considerar um modo de interpretação e de atualização, já que se trata de uma obra do século XVII. Tal hermenêutica tesauroiana coloca em questão os fundamentos filosóficos e aplicados da OC em questão, principalmente pela inflexão da presença do pensamento de Aristóteles como central para a contemporânea reflexão do domínio, fonte esta também nuclear em Tesauro.

De acordo com Saldanha e Silveira (2016), os resultados dos estudos tesauroianos em OC, no contexto ibero-americano e mundial, são insignificantes no tocante a sua teorização (e a relação aparentemente “curiosa” entre o sobrenome e o instrumento revolucionário das teorias sobre o “tesauro”, o instrumento e não o assunto histórico). Como exemplo, temos a ausência de uma reflexão direta e crítica sobre a teoria de Emanuele Tesauro em toda a produção do conhecimento no contexto da ISKO, até 2016.

Os estudos identificados, que abordam historicamente os preceitos de Emanuele Tesauro no âmbito da Ciência da Informação, são aqueles de Almeida & Crippa

(2009) e Monteiro & Giraldes (2008). Emanuele é apontado pelos primeiros estudiosos - Almeida & Crippa (2009) - como um notável pensador barroco, divulgador dos preceitos da metáfora. Já Monteiro & Giraldes (2008) indicam a presença de instrumento de linguagem documental, semelhante aos estudados no cerne da OC, na obra do pensador aristotélico.

De algum modo, na última década, estudiosos traçaram um panorama da importância desse autor e de suas ideias para a construção dos fundamentos histórico-epistemológicos do espaço de saberes da OC. Pelo exposto até o momento, levanta-se a proposta de traduzir as ideias de Emanuele Tesauro, como forma de incremento ao estudo crítico-histórico de sua obra. O objetivo se pauta em discutir o desenvolvimento da tradução para o português do Índice Categórico de Emanuele Tesauro, presente na obra *Luneta Aristotélica*, além de analisar a repercussão silenciosa das propostas do Índice Categórico tesauroiano na OC do século XX (com foco em autores pré-determinados).

A prática da pesquisa sobre Tesauro e sua obra em OC parece imediatamente aderente, principalmente pela (co)incidência do nome entre sujeito e instrumento (linguagem documentária). Porém, não apenas a relação imediata se faz concreta, como os caminhos parecem tortuosos, dada sua recepção contemporânea praticamente restrita à linguística e à semiótica. É nesse sentido, entretanto, que o percurso de compreensão da ausência de Tesauro no campo informacional (sobretudo na OC) retoma sob uma figuração incidente mais robusta: justamente entre semiótica e linguística encontramos grandes e relevantes linhas de reflexão no domínio, contudo sem (ainda) constituir um solo seguro para os estudos tesauroianos.

2 Investigações em construção: caminhos para os estudos tesauroianos

Como primeiro passo dos estudos, no contexto da investigação da pluralidade de cenários sociopolíticos de trabalho do que chamaram organização dos saberes, Saldanha e Silveira (2016) investigam via pesquisa teórica a fundamentação retórica do trabalho de Tesauro. As tentativas iniciais estavam em compreender (e destacar) o papel dos estudos retóricos no desenvolvimento de uma OC em seus pressupostos históricos, bem como escavar hipóteses sobre a condição de apagamento do teórico do século XVII.

No plano epistemológico, Saldanha e Souza (2017) colocaram em cena o contexto histórico-teórico do pensamento de Emanuele Tesauro. O foco da pesquisa estava em compreender o desenvolvimento de um conjunto de ações intersubjetivas dedicadas à representação, à criação e à apropriação dos saberes, no escopo dos estudos da linguagem, a partir de Tesauro. O núcleo do trabalho estava o contexto barroco, fonte da teorização tesauroiana. A linguagem e a crítica da linguagem na filosofia barroca abriam, na análise, fontes sensíveis para as futuras discussões da OC, como regimes, políticas e economia política da classificação. Almeida e Saldanha (2017) exploraram os potenciais metametodológicos da OC, apontando para as

relações possíveis entre os métodos analíticos e os métodos discursivos do domínio, potenciais instrumentos teóricos e operacionais que permitiriam investigar o próprio corpo de trabalhos desenvolvido em suas fronteiras. Ao apontar para a linhagem discursiva, os autores adentram o percurso de estudos dos *loci* (lugares) chegando até a relevância epistemológico-histórica de Emanuele Tesauro não apenas para essa dimensão, como também para a construção lógica da OC.

No mesmo ano, Saldanha (2017) procurou demonstrar o potencial de compreensão simbólica do pensamento de Tesauro para a OC, identificando as relações de sua teorização com as modernas figurações sociais e culturais do domínio. A aproximação coloca em cena as ligações possíveis e urgentes entre uma teoria barroca e as aberturas de uma teoria crítica da OC, principalmente relativa às pluralidades culturais.

Em “Who is Tesauro”, Saldanha, Silveira, Crippa e Almeida (2018) aprofundam o questionamento sobre a ausência de Emanuele na OC, indo, neste momento, objetivamente a análise crítica das razões que justificariam o apagamento do pensamento tesauroiano. Em pauta foram identificados pelos autores alguns aspectos centrais para tal ausência, a saber, a chamada vitória do mecanicismo sistêmico a partir de Bacon e seu Novo Organon aplicado ao domínio, o neopositivismo e sua força no campo, principalmente por sua condição estruturada na lógica, a economia política da ciência, que orienta a OC para as demandas do mercado e da indústria.

No mesmo coração dos estudos tesauroianos no Brasil, Menezes (2018), na tese “*Rasum tabulae*: um limiar metafórico-escritural dos estudos da informação,” explora o papel do conceito de retórica de Emanuele Tesauro para a Ciência da Informação, demonstrando a profundidade complexa das relações ontológicas no teórico do século XVII. As potencialidades culturais estão aqui destacadas por Menezes (2018), desdobrando a análise da metáfora na retórica para o papel do conceito de agudeza na teoria de Tesauro.

3 Caminhos da tradução: abrindo cenários para os estudos tesauroianos no mundo lusófono

O caráter metalinguístico do texto de Tesauro evoca não só uma maneira de compreender e restaurar a retórica aristotélica, mas de estabelecer uma prática de escrita, bem como, juntamente com a apresentação do método de tal prática, cheia do entrelaçamento de informações, conceitos, práticas, na forma de comentários. Todas essas possibilidades, ao mesmo tempo, soam como uma “prescrição”, no sentido de explorar formas de escrever, falar e representar como deveria ser feito, uma fonte que existe para todos os grandes índices.

Essa é, em suma, a loucura do Emanuele: ele está usando o Aristóteles lógico (analítico-conceitual), mas sob uma teoria do Aristóteles retórico (o discursivo) para indicar a dinâmica das palavras em relação às coisas. Como dito anteriormente, a parte traduzida do texto de Emanuele foi o Índice Categórico. O método adotado foi tradução palavra por palavra, com a observação de que nos casos em que a tradução

se tornava inaceitável o método era abandonado e tais palavras permaneciam na língua original.

Partindo do ponto de vista que a tradução é um exercício de comunicação entre diferentes línguas, épocas, histórias, enfim, culturas, que consiste em relacionar o significado de uma palavra com outras palavras que não pertencem à mesma cultura, o trabalho foi, e ainda está sendo feito, envolvendo quatro fases:

- Definição do escopo de tradução com foco na relação entre o pensamento de Tesouro e a OC, chegando até o Índice Categórico;
- Primeiro contato com a obra com apropriação da Introdução e do Índice Categórico;
- Tradução propriamente dita, que corresponde a parte criativa da tradução;
- Avaliação da tradução à luz do léxico da OC.

Na segunda fase a dificuldade encontrada dizia respeito ao desenho de algumas letras usadas à época. A terceira etapa foi a mais trabalhosa e criativa da tradução, envolveu a pesquisa entre os significados possíveis para os termos e as escolhas, segundo o tradutor, das mais fiéis, em português, às usadas pelo autor do texto. O apoio nesta fase foram os seguintes dicionários: *Il Nuovo Dizionario Italiano Garzanti* (1991), *Zingarelli minore* (2001) e *Dizionario Portoghese-Italiano/Italiano-Portoghese* (1960).

4 Traduzindo o movimento do Barroco: a dinâmica da OC em Tesouro

Tesouro é fruto inescapável de um pensamento barroco, e assim se faz sua linguagem. Na abertura de seu Índice Categórico, o teórico do Seiscentos diz-nos:

Segredo realmente segredo: nova, e profunda, incansável Mina de Infinitas Metáforas, de símbolos argutos, e de engenhosos Conceitos. (TESAURO, 1670, p.107, tradução de Silva, 2018).

A linguagem barroca de Tesouro escorre sobre forma e conteúdo, sobre texto e contexto. O que parece-nos distante, no entanto, reencontra nossa linha argumentativa no *Trivium* (gramática, retórica e lógica): o oculto a ser revelado (o conhecimento em sua dependência de movimentos anti-entrópicos, dependente dela, da organização) através de “Conceitos” (em maiúscula na visão de Tesouro).

No entanto, a (prematura ou original, a inovadora ou apenas contextual, posto que assim o barroco pensava) posição do pensamento de Tesouro nos coloca, antes dos “Conceitos”, o papel das “Metáforas”, ou ainda “Infinitas Metáforas” que levam à criação, à reprodução e a multiplicação (esta, sim, finita) de “Conceitos”. Desdobra-se daqui o poder da engenhosidade do intelecto na sua compreensão do mundo através deles, metáforas e conceitos.

Mas que [...] outro não é o engenho que penetrar os objetivos altamente ocultos sob diversas categorias; e de compará-los entre si; por essa razão, agradecimentos infinitos

sejam dados ao nosso Autor, primeiro a abrir essa porta secreta a todas as Ciências. [...] (TESAURO, 1670, p. 107, tradução de Silva, 2018).

A condição do tempo-espaço intersubjetivo do Barroco e do Seiscentos se coloca: a (re)volução constante dos “astros” que não são divinos e também “naturalmente” o são. O sagrado e o profano não possuem separação, mas a ciência ganha sua independência. Estranha figuração à primeira vista. Não muito distante, no entanto, quando a hermenêutica pós-estruturalista nos revela o mundo estratificado e suas confluências, a transgramaticalização do ser a partir da cultura. Estariam, de fato, em 2018, apartadas as relações entre religião, ciência e estado? Não só a sociologia da cultura e a antropologia cultural o fizeram, como os estudos críticos em OC já desmentiu o fato algumas dezenas de vezes, por exemplo, em análises estruturais dos fundamentos arbitrários de classificações totalizantes, como Classificação Decimal de Dewey e Classificação Decimal Universal.

Desdobra-se aqui o papel das “Categorias”, diálogo pontual com Aristóteles. Tesouro (1670) recorrerá, posteriormente, no Índice (não coincidentemente Categórico) para identificar a capacidade de compreender e de (inovação barroca, de natureza retórica, porém com sede na própria Retórica, a obra, aristotélica) revelar e de construir o conhecimento a partir da linguagem (ou seja, das metáforas e seus conceitos – esta, estrutura fundamental de ser compreendida, ou seja, conceitos derivados de metáforas, e não o contrário). O movimento de aproximação e, enfim, de comparação, estratégia científica moderna fundamental para toda a *episteme* nascente (incluindo a futura teoria da evolução das espécies), funda-se como central no trato da linguagem.

[...] outro não sendo o filosofar, que voar com a mente por todas as Categorias: a pesquisar as Notícias, ou seja, *Circunstâncias*; para desenhar Argumentos: e quem deles compreende, melhor filósofa. (TESAURO, 1670, p. 107, tradução de Silva, 2018).

Do e no comparar podemos tocar o real através da dimensão relevada pelas “Categorias”, da “informação”, ou seja, das “Notícias” das coisas elas mesmas (preceito científico no plano empírico) às “Circunstâncias” de seu aparecimento, é possível estabelecer “Argumentos”. Tesouro (1670) “brinca” aqui com os tubos de ensaio do *Trivium*: da lógica à retórica, passando pela gramática, é através da, não fora, da linguagem que se pode conhecer o mundo (típica posição pós-Gutenbergiana que consegue o mundo, no Seiscentos, como um emaranhando natural já espelhado no livro – o livro da natureza já está em franca atividade tipográfica, relevando-se na natureza do discurso científico “provado” bibliograficamente.

5 Considerações em curso: Tesouro aquém e além dos Trópicos

O Barroco obedece ao movimento dos extremos que se reaproximam. O seu tempo-espaço é o da dinâmica das coisas naturais que se emaranham no simbólico e dele se (re)concebem. Tesouro está nesse tempo. E talvez por isso, para além das hipóteses de seu apagamento nos estudos em curso sobre o seu pensamento para a OC, seu acesso ainda seja raro e raso.

O hermetismo da linguagem tesauriana pode, nesse sentido, em um primeiro momento, não só afugentar a curiosidade do domínio, como equivocadamente identificar uma distância inalcançável. O espelhamento nome e instrumento, sujeito e linguagem documentária, perder-se-ia, pois, sem diálogos com (justamente) linguística e semiótica, searas afins à OC e dedicadas ao estudo de Tesouro. Poderíamos, pois, identificar o corpo de estudos tesaurianos atuais como parte de um enfoque histórico pontual nos trópicos, oriundo do Brasil.

Entretanto, com lentes epistemológico-históricas, focadas na instância constante da recontextualização, via o lento e necessário processo de tradução, o pensamento de Tesouro se abre, firmando-se não só aderente, como co-fundador (dentre tantas margens de construção teórica do domínio) da OC. Para, além disso, usando a própria metáfora tesauriana, se das lentes passamos à luneta, podemos enxergar-perceber que Tesouro adentra o cerne dos mais difíceis e instigantes dilemas, atuais e históricos, da própria OC.

Referências

- Almeida, Marco A. de; Crippa, Giulia (2009). De Bacon à internet: considerações sobre a organização do conhecimento e a constituição da ciência da informação. *Ponto de Acesso*, Salvador, 22(3): 109-131.
- Almeida, T.; Saldanha, G. S. (2017). Entre a abordagem analítica e os loci epistêmicos: um debate metametodológico para a organização do conhecimento. In: Encontro Nacional de Pesquisa em Ciência da Informação, 2017, Marília. Anais do XVIII Encontro Nacional de Pesquisa em Ciência da Informação. Marília: PPGCI UNESP. p. 1-20.
- Menezes, Vinícios Souza de. (2018). *Rasum tabulae*: um limiar metafórico-escritural dos estudos da informação,. Tese, 2018. Doutorado em Ciência da Informação. Rio de Janeiro: Programa de Pós-Graduação em Ciência da Informação – PPGCI IBICT UFRJ.
- Monteiro, Silvana D.; Giraldez, Maria Júlia C. (2008) Aspectos lógico-filosóficos da organização do conhecimento na esfera da ciência da informação. *Inf. & Soc.: Est.*, João Pessoa, v. 18, n. 3, set./dez., p. 13-27.
- Saldanha, G. S. (2017). Sobre a O²S²O, de Tesouro à Bourdieu: linguagem simbólica e a organização ordinária dos saberes socialmente oprimidos. In.: Lucas, Elaine Rosângela de Oliveira; Silveira, Murilo Artur Araújo da. (Org.). *A Ciência da informação encontra Pierre Bourdieu*. 1ed.Recife: UFPE. p. 173-196.

- Saldanha, G. S.; Souza, R. F. (2017). Teoria barroca da organização do conhecimento: Emanuele Tesauro e o espelho turvo das tensões entre epistemologia, metodologia e sociedade. *Informação & Informação* (Online), v. 22, p. 11-32.
- Saldanha, Gustavo S.; Silveira, Naira C. (2016). The treasure of Tesauro: knowledge organization, rhetoric and language. In.: Conference of the International Society for Knowledge Organization (ISKO). *Proceedings....* Berlim: Springer. Pp. 301-321.
- Saldanha, G. S.; Silveira, N. C. ; Crippa, G. ; Almeida, T. (2018). Who is Tesauro? The man, the words and the things. In: 15th International ISKO Conference, 2018, Porto. *Proceedings...* Porto: Universidade do Porto. v. 1. p. 1-15.
- Tesauro, Emanuele. (1670). *Cagioni Instrumentali delle argutezze oratorie simboliche et lapidarie: índice categórico*, p.107-122. In: Tesauro, Emanuele. (1670). *Il cannocchiale Aristotelico*. Turim. Tradução de Maria Helena Teixeira da Silva (2018).

Análise dos tipos de relações do modelo SKOS: perspectivas de representação de recursos audiovisuais

Rogério Aparecido Sá Ramalho¹ y Brígida Maria Nogueira Cervantes²

¹ ORCID [0000-0002-8491-3514](https://orcid.org/0000-0002-8491-3514). Professor Associado Universidade Federal de São Carlos ,
UFSCar

ramalho@ufscar.br

² ORCID [0000-0001-7356-1798](https://orcid.org/0000-0001-7356-1798). Professora Associada Universidade Estadual de Londrina,
UEL.

brigidacervantes@gmail.com

Resumo: Com o aumento exponencial da quantidade de bases de dados para o armazenamento de recursos audiovisuais, um novo desafio tem sido imposto para a área de Organização do Conhecimento, o de representar recursos audiovisuais. Entre os novos instrumentos de representação, destaca-se o *Simple Knowledge Organization System* – SKOS, um modelo de dados que tem como objetivo representar a estrutura básica e conteúdo de *Knowledge Organization System* – KOS em ambientes computacionais, favorecendo a incorporação de recursos audiovisuais nos processos de representação. Este trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelo CNPq, e objetiva analisar os tipos de relações previstas no modelo de dados SKOS e as perspectivas de utilização deste modelo para a representação de recursos audiovisuais. Apresenta caráter teórico-metodológico e abordagem qualitativa, de natureza aplicada. Reverbera-se que no modelo SKOS as relações constituem a ‘espinha dorsal’ para a modelagem de KOS, podendo ser classificadas em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização. Assim, o modelo SKOS apresenta novas perspectivas para a representação de recursos audiovisuais, a partir do fornecimento de uma estrutura simplificada, porém flexível, que contribui para uma maior interoperabilidade e alcance das relações estabelecidas, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

Palavras chave: Knowledge Organization Systems; Vocabulários Controlados; Modelagem de Dados; Tecnologias Semânticas; Ontologias.

Abstract: With the exponential increase in the number of databases for the storage of audiovisual resources, a new challenge has been imposed for the area of Organization of Knowledge, that is, to represent audiovisual resources. Among the new representational instruments, we highlight the Simple Knowledge Organization System (SKOS), a data model that aims to represent

the basic structure and content of Knowledge Organization System (KOS) in computational environments, favoring the incorporation of audiovisual resources in representation processes. This work is linked to an ongoing research project, funded by CNPq, and aims to analyze the types of relationships planned in the SKOS data model and the perspectives of using this model for the representation of audiovisual resources. It is an applied-natured work with a theoretical-methodological character and a qualitative approach. It is pointed out that in the SKOS model the relationships constitute the 'backbone' for the KOS modeling, being classified in three macro categories: Structural, Semantic and Compatibility Relationships. Thus, the SKOS model presents new perspectives for the representation of audiovisual resources, from the provision of a simplified but flexible structure that contributes to greater interoperability and reach of established relationships, providing a semantic dimension practically unexplored previously.

Keywords: Knowledge Organization Systems; Controlled Vocabularies; Data Modeling; Semantic Technologies; Ontology.

1 Introdução

A partir da incorporação de tecnologias computacionais nos processos de representação e organização tornou-se possível a compatibilização de vocabulários controlados de maneira mais eficiente, favorecendo a integração de tesouros em Sistemas de Recuperação de Informação e consequente novas possibilidades de representação (Alvite Díez, Pérez León, Martínez González & Vicente Blanco, 2010).

Ao longo das últimas décadas, com o aumento exponencial da quantidade de bases de dados para o armazenamento de recursos audiovisuais, um novo desafio tem sido imposto para a Organização do Conhecimento o de representar recursos audiovisuais. Entre os novos instrumentos de representação, destaca-se o *Simple Knowledge Organization System* – SKOS, um modelo de dados que tem como objetivo representar a estrutura básica e conteúdo de *Knowledge Organization System* – KOS em ambientes computacionais, apresentando novas possibilidades aos processos de representação.

O presente trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelo CNPq, e possui como objetivo analisar os tipos de relações previstas no modelo de dados SKOS, de modo a identificar novas perspectivas de utilização deste modelo para a representação de recursos audiovisuais.

Fundamentando-se em uma pesquisa de caráter teórico-metodológico, pretende oferecer respostas aos problemas de pesquisa apresentados. Possui abordagem descritiva, pois busca apresentar características voltadas à descrição, compreensão e explicação de relações de um determinado fenômeno de modo a considerar o caráter interativo entre os objetivos definidos e as orientações teóricas da pesquisa. Caracteriza-se, também, como pesquisa de natureza aplicada, uma vez que “objetiva

gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos” (Silveira & Córdova, 2009, p. 35).

2 Evolução dos Sistemas de Representação

Uma discussão recorrente alusiva aos modelos e sistemas de representação é o consenso sobre o que representar e como fazê-lo. Ao longo dos anos, diferentes áreas (Ciência da Informação, Ciência da Computação, Linguística, Ciência Cognitivas, dentre outras) têm apresentado diversos modelos de representação, considerando propósitos e características distintas.

No âmbito da área de Ciência da Informação para compreender a real importância da Representação torna-se necessário considerar seu próprio significado e abrangência, a partir dos estudos realizados que resultaram em grande número de padrões e diretrizes, bem como em desenvolvimentos teóricos como os princípios de Cutter, análise facetada e avanços da tecnologia da informação (Hjørland, 2003).

As linguagens documentárias podem ser consideradas como um dos instrumentos de representação mais difundidos no domínio da Ciência da Informação, caracterizando-se como sistemas de signos que visam à uniformização da linguagem utilizada em um Sistema de Informação, proporcionando uma representação padronizada do conteúdo informacional (Bocato, 2008).

Segundo Hodge (2000) o termo *Knowledge Organization Systems (KOS)* foi proposto em 1998 pelo *Networked Knowledge Organization Systems Working Group* para englobar sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, redes semânticas e ontologias. Na taxonomia de tipos de KOS proposta por Hodge (2000) também são incluídos, entre outros, dicionários, glossários, taxonomias e tesausos.

Lara (2015, p. 92) argumenta que “no Brasil, não há consenso sobre a utilização de um termo que abranja o conjunto de instrumentos de organização da informação e do conhecimento”. Apesar das variações de denominação identificadas na literatura nacional para designar tais instrumentos, observa-se um crescente aumento no número de pesquisas relacionadas a esta temática.

A partir da popularização do ambiente Web, novos instrumentos de representação organização e representação têm sido propostos, provocando convergências entre aportes teóricos e metodológicos de diversos campos científicos, no intuito de atender as demandas informacionais de uma sociedade cada vez mais baseada em interações virtuais.

Na última década foi constatado um crescente interesse de pesquisadores no estudo e desenvolvimento de ontologias, envolvendo pesquisas empenhadas, principalmente, na consolidação das definições de tal instrumento de representação e análise de possibilidades efetivas de uso (Santos, Correa & Silveira, 2013).

García-Marco (2007) ressalta que o crescente aumento do número de pesquisas relacionadas a temática de ontologias na área de Ciência da Informação constitui a

“ponta do iceberg” de um processo de reconfiguração disciplinar, como resultado da integração de diversas Ciências.

Para García Jiménez (2004), apesar das similaridades existentes entre tesauros e ontologias, ambos os instrumentos possuem vinculações operativas e teóricas distintas, sendo desenvolvidos a partir de diferentes trajetórias históricas. Além disso, é importante ressaltar que tradicionalmente os modelos clássicos de representação eram desenvolvidos para serem utilizados diretamente por usuários humanos, enquanto os novos instrumentos de representação desenvolvidos para serem utilizados em ambientes digitais, como as ontologias, tem como foco possibilitar a realização de inferências automáticas a partir de sistemas computacionais (Ramalho, 2015a).

Apesar das potencialidades oferecidas pelas ontologias, sua utilização ainda tem sido limitada, devido principalmente a complexidade e custos inerentes ao processo de modelagem. Por outro lado, Pastor-Sanchez, Martinez-Rodriguez e Mendez-Muñoz (2012) destacam que a exploração dos tesauros na Web é muito limitada, pois as diferentes formas de recuperação na Internet exigem vocabulários interoperáveis e multilíngues, abandonando o paradigma lexical em favor de um paradigma conceitual, de modo que o próprio conceito de tesouro precisa ser adaptado às novas necessidades informacionais emergentes. Segundo Sanchez-Jimenez e Gil-Urdiciain (2007) esta adaptação permite superar as deficiências das relações fortemente ancoradas ao nível léxico que caracterizam os tesauros tradicionais.

Clarke e Zeng (2012) lembram que os princípios para a construção de tesauros mudaram radicalmente da norma ISO 2788 para a ISO 25964, publicada em 2011, tais mudanças podem ser justificadas devido ao avanço das tecnologias digitais que demandam uma necessidade de interoperabilidade semântica, em que a distinção entre termo e conceito é fundamental.

Nesta perspectiva, o *Simple Knowledge Organization System* – SKOS tem ganhado destaque ao longo dos últimos anos como uma alternativa intermediária, que permite a transição de KOS tradicionais para o ambiente Web, fornecendo um elo entre o formalismo lógico rigoroso das linguagens de ontologias e o mundo caótico, informal e pouco estruturado de ferramentas de colaboração baseadas na Web.

3 O Modelo de Dados SKOS

No ano de 2009 o Word Wide Web Consortium (W3C) formalizou como Recomendação o padrão *Simple Knowledge Organization System* – SKOS (Sistema de Organização do Conhecimento Simples) como um modelo de dados para a representação de Sistemas de Organização do Conhecimento no ambiente Web, buscando atender às novas demandas informacionais relacionadas ao compartilhamento e interoperabilidade de vocabulários no ambiente Web (Miles & Bechhofer, 2009; Ramalho, 2015b).

O modelo SKOS proporciona um cânone para a representação formal da estrutura básica e conteúdo de Sistemas de Organização: como tesauros, esquemas de

classificação, listas de cabeçalho de assunto, taxonomias, entre outros tipos similares de vocabulários controlados (Ramalho, Vidotti & Fujita, 2007). Sua codificação é baseada no padrão RDF (*Resource Description Framework*), favorecendo a representação de informações de maneira estruturada, assim como a integração de diferentes esquemas conceituais.

Conforme os relatos de Pastor-Sanchez, Martínez-Mendez e Rodriguez-Muñoz (2009) o desenvolvimento do modelo SKOS começou em 2002, com as atividades do grupo de trabalho *Semantic Web Advanced Development for Europe* (SWAD-E) e foi difundido publicamente em novembro de 2005, com a publicação do documento *SKOS Core Guide* (Miles & Brickley, 2005). Contudo, apenas em agosto de 2009 com a publicação do *SKOS Simple Knowledge Organization System Reference* o SKOS passou a ser considerado como uma recomendação oficial do W3C, um modelo de dados para a disponibilização e compartilhamento de sistemas de organização do conhecimento na Web.

Tendo como principal objetivo favorecer a representação de vocabulários no ambiente Web, como base em um modelo simplificado, o SKOS não tem a função de substituir vocabulários controlados em seu contexto original de uso, e sim favorecer uma maior reutilização e interoperabilidade entre os vocabulários existentes (Ramalho, 2015a). Segundo Isaac e Summer (2009) o padrão SKOS possibilita que conceitos possam ser identificados utilizando *Uniform Resource Identifiers* (URI), rotulados com sequências textuais em uma ou mais línguas naturais, documentados por meio de diferentes tipos de notas, relacionados semanticamente entre si, a partir de hierarquias informais e redes associativas; e agregados em diferentes esquemas conceituais. Mustafa El Hadi (2015) argumenta que desde a padronização do modelo SKOS muitos vocabulários passaram a ser protegidos por direitos autorais para serem publicados como dados ligados na Web, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

Dextre Clarke e Zeng (2012, p. 23) destacam a importância do padrão SKOS

As Recomendações SKOS do W3C são projetadas para suportar publicações de vocabulários, como tesouros na Web. E na sua essência é um modelo de dados que distingue expressamente entre os conceitos e etiquetas para representar conceitos. [...] Além disso, que é veiculado pelo SKOS Core para as relações entre conceitos, a extensão fornece suporte adicional para identificar, descrever e linkar entidades lexicais.

De acordo com a norma ISO 25964-1 (International Organization for Standardization, 2011) tradicionalmente os KOS estão fundamentados em três tipos principais de relações: relações hierárquicas, as relações associativas e relações de equivalência. Contudo, apesar do crescente número de pesquisas com foco no desenvolvimento de vocabulários SKOS, verifica-se que os novos tipos de relações previstas por este modelo têm sido pouco explorados na literatura.

Como o modelo SKOS tem como objetivo representar formalmente a estrutura e conteúdo de KOS, sua implementação não se limita a representação de Conceitos, incorporando também a definição de Coleções Conceituais e Esquemas de Conceitos. As Coleções Conceituais (*skos:Collection*) são utilizadas para agrupar conceitos que compartilham determinadas características ou que necessitam de ordenação específica (*skos:OrderedCollection*). Em relação aos Esquemas de Conceitos (*skos:ConceptScheme*), estes são utilizadas para favorecer o estabelecimento de relações entre diferentes KOS em um mesmo vocabulário SKOS, possibilitando interoperabilidade semântica. As relações existentes entre Conceitos, Coleções e Esquemas Conceituais podem ser definidas como ‘Relações Estruturais’ e possibilitando a representação adequada dos diferentes elementos que compõem o modelo SKOS. A figura 01 apresenta graficamente as Relações Estruturais previstas no modelo SKOS.

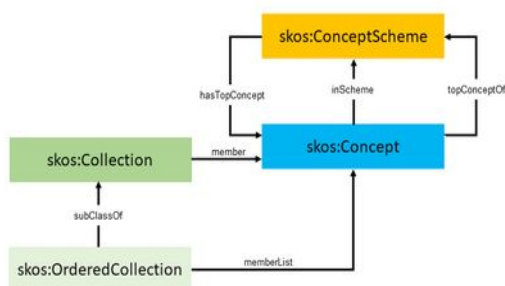


Figura 1. Relações Estruturais em SKOS.

Conforme pode ser observado na Figura 01 as Relações Estruturais previstas em SKOS são representadas a partir de diferentes propriedades previstas no modelo, à medida que os elementos Conceitos, Coleções e Esquemas Conceituais representam conjuntos disjuntos e não podem fazer parte de uma mesma hierarquia conceitual.

As Relações Semânticas são inegavelmente os tipos de relações mais explorados para a elaboração de KOS, do mesmo modo, em SKOS podem ser categorizadas como Relações Hierárquicas (*skos:broader* e *skos:narrower*), Relações Associativas (*skos:related*) e Relações de Mapeamento (*skos:mappingRelation*).

Relações de Mapeamento são tradicionalmente denominadas na literatura como Relações de Equivalência. Assim, traçando um paralelo das possibilidades de mapeamento entre conceitos oferecidas em SKOS com a literatura no campo de Organização do Conhecimento é possível estabelecer aproximações com a Matriz de Compatibilização Conceitual proposta por Dahlberg (1981, 1983).

Nesta perspectiva, considera-se adequado denominar como Relações de Compatibilização as relações previstas no modelo SKOS (*skos:exactMatch*, *closeMatch*, *skos:broadMatch*, *skos:narrowMatch* e *skos:relatedMatch*) que possuem como objetivo favorecer a compatibilização entre os conceitos existentes em

diferentes esquemas conceituais. A figura 02 apresenta graficamente uma síntese da classificação proposta para os diferentes tipos de Relações previstas em SKOS.

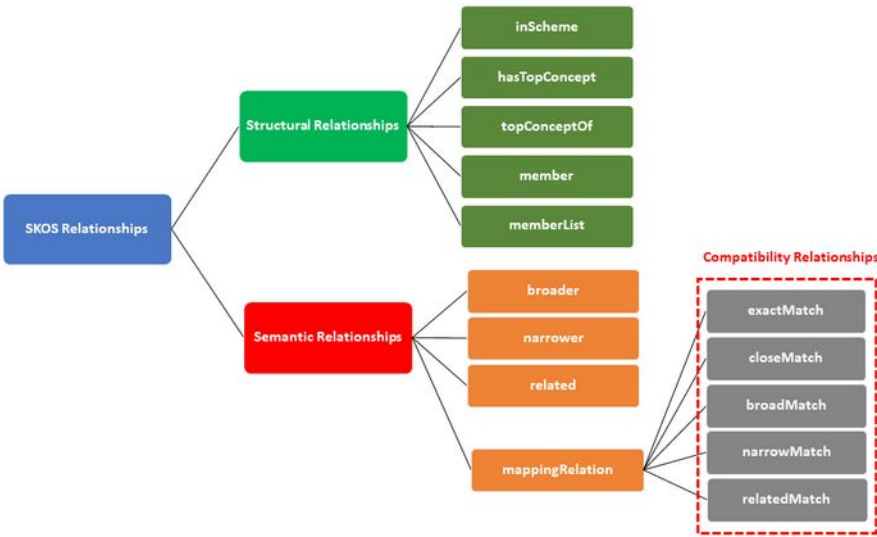


Figura 2. Categorização dos Tipos de Relações previstas em SKOS.

A figura 02 apresenta uma categorização das Relações previstas em SKOS em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização. Além das Relações Semânticas tradicionalmente exploradas em KOS convencionais, o modelo SKOS apresenta um novo subconjunto de Relações Semânticas que pode ser denominada como Relações de Compatibilização, pois favorece a compatibilização formal e o alinhamento de conceitos existentes em diferentes esquemas conceituais.

4 Considerações Finais

Este artigo teve como objetivo analisar os tipos de relações previstas no modelo de dados SKOS, a partir da análise realizada verificou-se que as relações constituem a ‘espinha dorsal’ para a modelagem de KOS, podendo ser classificadas em três macro categorias: Relações Estruturais, Relações Semânticas e Relações de Compatibilização.

A partir das Relações de Compatibilização é possível estabelecer aproximações com estudos relacionados à Matriz de Compatibilização Conceitual proposta por Dahlberg (1981, 1983). Destaca-se que tais relações favorecem ir além de representações puramente conceituais, apresentando novas perspectivas para a

representação a partir da compatibilização de recursos audiovisuais, proporcionando uma dimensão semântica praticamente inexplorada anteriormente.

As Relações Estruturais possibilitam o estabelecimento de uma estrutura simplificada, porém flexível, que contribui para uma maior interoperabilidade e alcance das relações estabelecidas. Assim o presente estudo corrobora a crescente necessidade de sistematização de fundamentos teóricos que possibilitem validar o desenvolvimento de novos instrumentos e métodos, à medida que avanços tecnológicos possam ser devidamente sedimentados no campo teórico, contribuindo para o fortalecimento de um arcabouço teórico que favoreça um melhor entendimento e utilização das tecnologias digitais contemporâneas nos processos de representação de recursos audiovisuais.

Referências

- Alvite Díez, M. L., Pérez León, B., Martínez González, M. M., & Vicente Blanco, D. J. (2010). Propuesta de representación del tesoro Eurovoc en SKOS para su integración en sistemas de información jurídica. *Scire: Representación y Organización del Conocimiento*, 16(2), 47-51.
- Boccato, V. (2008). A linguagem documentária como instrumento de organização e recuperação da informação. In Hoffmann, W. A. M., & Furnival, A. C. (Orgs.) *Olhar: ciência, tecnologia e sociedade* (p. 269-278). São Paulo: CECH-UFSCar.
- Clarke, S. G., & Zeng, M. L. (2012). From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. *Information Standards Quarterly (ISQ)*, 24(1).
- Dahlberg, I. (1981). Conceptual definitions for interconcept. *International Classification*, 8(1).
- Dahlberg, I. (1983). Terminological definitions: characteristics and demands. In: *Problèmes de la définition et de la synonymie en terminologie*. Québec, GIRSTERM, 13-51.
- Dextre Clarke, S. G., & Zeng, M. L. (2012). From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling. *Information Standards Quarterly (ISQ)*, 24(1).
- García Jiménez, A. (2004). Instrumentos de representación del conocimiento: tesauros versus ontologías. *Anales de documentación* (Vol. 7, pp. 79-95). Facultad de Comunicación y Documentación y Servicio de Publicaciones de la Universidad de Murcia. Recuperado de <http://www.um.es/fccd/anales/ad07/ad0706.pdf>
- García-Marco, F. J. (2007). Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. *El profesional de la información*, 16(6), 541-550.
- Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge organization*, 30(2), 87-111.
- Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington, D.C.: Digital Library Federation, Council on Library and Information Resources. Recuperado de <http://old.diglib.org/pubs/dlf090/dlf090.pdf>

- International Organization for Standardization. (2011). *ISO 25964-1:2011*: Information and documentation – thesauri and interoperability with other vocabularies – part 1: Thesauri for information retrieval. Genebra.
- Isaac, A., & Summers, E. (2009). *SKOS Simple Knowledge Organization System Primer*. World Wide Web Consortium (W3C) Working Group Note 18 August 2009. <http://www.w3.org/TR/skos-primer/>
- Lara, M. L. G. (2015). Propostas de tipologias de KOS: uma análise das referências de formas dominantes de organização do conhecimento. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 20 (n. esp.), 89-107. <http://dx.doi.org/10.5007/1518-2924.2015v20nesp1p89>
- Miles, A., & Bechhofer, S. (2009). *SKOS simple knowledge organization system reference*. World Wide Web Consortium (W3C) recommendation, 18. Recuperado de <http://www.w3.org/TR/skos-reference/>
- Miles, A., & Brickley, D. (2005). *SKOS Core Guide*. World Wide Web Consortium (W3C), November. <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/>
- Mustafa El Hadi, W. M. (2015). Cultural Interoperability and Knowledge Organization Systems. In Guimarães, J. A. C., & Dodebei, V. (Org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil; FUNDEPE, 575-606.
- Pastor-Sanchez, J. A., Martínez-Mendez, F. J., & Rodríguez-Muñoz, J. V. R. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14(4), 10.
- Pastor-Sánchez, J. A., Martínez-Méndez, F. J., & Rodríguez-Muñoz, J. V. (2012). Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. *El profesional de la información*, 21(3), 245-253.
- Ramalho, R. A. S., Vidotti, S. A. B. G., & Fujita, M. S. L. (2007). Web semântica: uma investigação sob o olhar da Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*, 8(6).
- Ramalho, R. A. S. (2015a). Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a web. *Informação & Tecnologia*, 2(1), 66-79.
- Ramalho, R. A. S. (2015b). Ontologias e Simple Knowledge Organization System (SKOS): aproximações e diferenças. Guimarães, J. A. C., & Dodebei, V. (Org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil; FUNDEPE, 100-107.
- Sánchez-Jiménez, R., & Gil-Urdiciain, B. (2007). Lenguajes documentales y ontologías. *El profesional de la información*, 16(6), 551-560.
- Santos, M. T., Correa, R. F., & Silveira, M. A. A. (2013). Estudos brasileiros sobre Ontologia na Ciência da Informação. *DataGramaZero - Revista de Ciência da Informação*, 14(1). Recuperado de http://www.dgz.org.br/fev13/Art_05.htm
- Silveira, D. T., & Córdova, F. P. (2009). A pesquisa científica. In: Gerhardt, T. E., & Silveira, D. T. (Orgs.). *Métodos de pesquisa*. Porto Alegre: Editora da UFRGS.

Metacategorias semânticas para aplicação em SKOS

Rita do Carmo Ferreira Laipelt¹, Luciana Monteiro Krebs² y Renê Faustino Gabriel Júnior³

¹ ORCID [0000-0002-7429-8490](https://orcid.org/0000-0002-7429-8490). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS), Rio Grande do Sul.

rita.laipelt@ufrgs.br

² ORCID [0000-0002-2882-6803](https://orcid.org/0000-0002-2882-6803). doutoranda do Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul. Rio Grande do Sul e do Doctoral Programme Social Sciences at Katholieke Universiteit Leuven (KU Leuven)

luciana.monteiro@ufrgs.br

³ ORCID [0000-0003-1021-3360](https://orcid.org/0000-0003-1021-3360). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul (PPGCIN/UFRGS), Rio Grande do Sul.

rene.gabriel@ufrgs.br

Resumo. Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em SKOS. Busca especificamente identificar tipologias de relações semânticas classificadas como relações associativas para aplicação no software Thesa - Tesouro Semântico Aplicado. No contexto atual, a interface de pesquisa de Sistemas de Recuperação da Informação (SRI), com suas diferentes possibilidades de busca, exercem um papel intermediário entre o usuário e o acervo documental de uma instituição. Assim, é necessário que esses sistemas sejam pensados de forma a facilitar a recuperação da informação pelos usuários. Uma das formas de fazer isso é investir no aperfeiçoamento das relações semânticas dos Sistemas de Organização do Conhecimento (SOCs). A inserção das metacategorias semânticas identificadas possibilitaram a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento como: Biologia vegetal, Eletromagnetismo, Comunicação Científica e Literatura. A consolidação dos resultados desta pesquisa possibilitará o desenvolvimento de tesouros semanticamente fortalecidos, numa transposição natural entre a teoria e a prática. Dos futuros resultados do uso do Thesa, está a possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas, bem como gerar inferências com bases em outros tesouros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

Palavras-chave: Organização do Conhecimento. Relações Semânticas. SKOS. Thesa.

Abstract. This work aims to develop a model of semantic metacategories for application in SKOS. It specifically seeks to identify typologies of semantic relations classified as associative relations for application in the Thesa - Thesaurus Applied Semantic software. In the current context, the Information Retrieval Systems (SRI) search interface, with its different search possibilities, plays an intermediate role between the user and the documentary collection of an institution. Thus, it is necessary that these systems be designed in order to facilitate the retrieval of information by users. One way to do this is to invest in improving the semantic relationships of Knowledge Organization Systems (SOCs). The insertion of the semantic metacategories identified allowed the observation and description of their functioning in different domains of knowledge such as: Plant Biology, Electromagnetism, Scientific Communication and Literature. The consolidation of the results of this research will allow the development of semantically strengthened thesauri, in a natural transposition between theory and practice. From the future results of the use of Thesa, there is the possibility of establishing thesaurus between different domains and themes, as well as generating inferences based on other thesauri, facilitating the operationalization of their managers with the use of Artificial Intelligence.

Keywords: Knowledge Organization. Semantic relations. SKOS. Thesa.

1 Introdução

Este trabalho tem por objetivo desenvolver um modelo de Metacategorias semânticas para aplicação em SKOS. Busca especificamente, Sistematizar as diferentes correntes teóricas sobre relações semânticas da área de Organização do conhecimento e seus respectivos métodos na Ciência da Informação. Entendemos que a partir desse mapeamento das relações semânticas, especialmente aquelas classificadas como relações associativas em SKOS será possível sistematizar Metacategorias de relações semânticas para aplicação no software Thesa - Tesouro Semântico Aplicado.

No contexto atual, a interface de pesquisa de Sistemas de Recuperação da Informação (SRI), com suas diferentes possibilidades de busca, exercem um papel intermediário entre o usuário e o acervo documental de uma instituição. Assim, é necessário que esses sistemas sejam pensados de forma a facilitar a recuperação da informação pelos usuários. Uma das formas de tornar os SRI mais acessíveis aos usuários é investir no aperfeiçoamento das relações semânticas dos Sistemas de Organização do Conhecimento (SOCs). Tendo em vista que na recuperação da informação, de acordo com Hjørland (2007, p. 392) “a função básica das relações semânticas é contribuir para a otimização da precisão e da revocação.” No entanto, de acordo com Green (2001), a magnitude e complexidade dos relacionamentos semânticos na área de organização do conhecimento dificultam o uso consistente dos mesmos tanto por profissionais da informação como por usuários finais. Logo, é necessário identificar e sistematizar esses relacionamentos semânticos para que os mesmos possam ser utilizados de maneira inequívoca em SOCs. Para Green (2001,

p.14) “embora a utilidade de relacionamentos explícitos seja mínima para usuários finais [...] a compreensão intuitiva que os humanos trazem para os relacionamentos não é compartilhada por dispositivos computacionais.” Por isso a explicitação desses relacionamentos “talvez sejam a nossa melhor esperança para infundir qualidade superior em nossos sistemas de recuperação.” (Green, 2001, p.14). Visto que, segundo Green (2001), a qualidade superior dos sistemas de recuperação da informação também pode vir a partir da possibilidade de identificação de fontes relevantes, as quais, sem o uso de relacionamentos semânticos, não seriam recuperadas.

Acreditamos que a explicitação de relações semânticas, sobretudo as relações associativas, podem contribuir tanto para o aperfeiçoamento de SKOS, no caso específico deste trabalho, do Thesa, como para a recuperação da informação e descoberta de novas fontes (nesse caso, quando tivermos um SKOS como o Thesa interligado com bases de dados como a Brapci, por exemplo, isso será possível.).

2 Relações Semânticas em Sistemas de Organização do Conhecimento

Os diferentes tipos de instrumentos de representação do conhecimento existentes atualmente são designados pelo termo “Sistemas de Organização do Conhecimento” (SOCs). Esta é uma nova denominação para o que conhecemos por linguagens documentárias, seu diferencial, no entanto está na incorporação de elementos de inovação tecnológica da era digital (Carlan, 2010). Logo, podemos considerar como um SOC diferentes tipos de instrumentos de representação da informação tais como: tesauros, ontologias, taxonomias, redes semânticas entre outros.

Os SOC produzem maior suporte semântico para sistemas de recuperação de informações (SRI). Considerando que as expressões utilizadas para representar documentos ou a pesquisa de usuários podem ter significados diferentes, sem atributos e links semânticos que contextualizem o domínio, é impossível para a máquina interpretar o seu significado. De modo que na ausência de propriedades semânticas, um SRI poderá apenas medir a semelhança entre as palavras usadas em um documento com as da consulta do usuário o que não é suficiente para a recuperação de resultados relevantes (Maculan, Lima, & Oliveira, 2017).

O modelo Simple Knowledge Organization System (SKOS), de acordo com Isaac e Summers (2009) é utilizado para expressar vários tipos de esquemas conceituais tais como tesauros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, Folksonomia, e outros tipos de vocabulários controlados Sua estrutura possibilita o compartilhamento e ligação dos Sistemas de Organização do Conhecimento (SOC) por meio da Web (Miles & Bechhofer, 2009). O SKOS compartilha muito dos princípios de elaboração de tesauros, taxonomias, esquemas de classificação e sistemas de cabeçalhos de assuntos, bem como possibilita capturar muito dessa semelhança e torna-la explícita, para que possa ser reutilizada por outras aplicações (Miles & Bechhofer, 2009). O modelo é composto por termos num vocabulário denominado SKOS Core Vocabulary (Miles & Brickley, 2005) que é um conjunto de

propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF (RDFSchema). Conforme Catarino (2014), o modelo SKOS baseia-se em classes e propriedades e para representar o conjunto de dados tem como elemento central o conceito.

Um tesauro, por exemplo, pode representar o conhecimento de um domínio através de um conjunto de conceitos que apresentam relações semânticas hierárquicas, associativas e de equivalência. Essas três tipologias de relações podem ser consideradas limitadas e em alguns casos causar ambiguidade entre conceitos e consequentemente problemas de recuperação da informação. Por isso, é fundamental aprimorar as conexões entre conceitos para que, através de sua explicitação, tenhamos uma semântica mais forte conforme ocorre nas ontologias. Um SOC seria capaz de contemplar, por meio de uma espécie de hospitalidade infinita, múltiplas perspectivas (pontos de vista) em uma estrutura mais flexível (Gnoli, 2008).

Porém, Café e Brascher (2011, p. 26) destacam que

A arbitrariedade da seleção dos agrupamentos e relacionamentos não pode ser vista, no entanto, como algo totalmente aleatório. Ela se deve aos aspectos que desejamos destacar numa determinada representação do conhecimento.”

Deve, portanto, haver correspondência entre o sistema e a realidade que este representa, tendo em vista o contexto onde as expressões ocorrem para a inferência dos significados. (Weiss & Brascher, 2013).

Verifica-se na literatura da área de organização do conhecimento o incentivo ao uso de outros tipos de relações semânticas. Assim, Hjørland (2007, p. 393) questiona “Como devemos explicar essa demanda por um conjunto muito mais rico de relações do que as normalmente usadas em tesouros?”, e explica que uma função adicional que as relações semânticas em um SOC poderiam desempenhar seria “fazer os diferentes interesses e paradigmas visíveis para que o usuário possa fazer uma escolha informado” (Hjørland, 2007, p. 389). Para tanto, o autor destaca a análise da literatura da área a ser representada pelo SOC como essencial para a identificação de diferentes pontos de vista.

A partir da análise da literatura dos domínios é possível identificar relações paradigmáticas e sintagmáticas. As relações paradigmáticas são constituídas por sinonímia, antonímia, meronímia, hiperonímia e hiponímia. Nesse caso as relações conceituais são por exemplo: oposição conceitual, parte de, é um tipo de. Já as relações sintagmáticas ocorrem entre entidades de diferente natureza, seus itens co-ocorrem. Essas relações incluem objetos e agentes, eventos e processos (Prevot, 2010).

As relações sintagmáticas podem ser tratadas nos tesouros a partir da explicitação das relações associativas. No entanto, de acordo com Maculan, Lima e Oliveira (2017) as relações associativas são consideradas as mais difíceis de definir e sobre as quais ainda não existe pesquisa suficiente para determinar suas bases teóricas. Portanto, de acordo com as autoras as relações associativas devem ser estabelecidas

principalmente a partir da análise da literatura, de modo que seja possível identificar os diferentes pontos de vista ali presentes.

Diante do exposto, é possível afirmar que em um tesauro teremos relações paradigmáticas nas relações hierárquicas, associativas e de equivalência. Já entre as relações sintagmáticas teremos apenas relações associativas.

Quadro 1. Tipos de Relações Paradigmáticas e Sintagmáticas

Relações paradigmáticas	As relações sintagmáticas
<p>Equivalência: sinonímia, variação denominativa (abreviatura, sigla, termo oculto, flexão verbal, etc)</p> <p>Hierárquica: hiperonímia (TG), hiponímia (TE)</p> <p>Associativa: meronímia (parte de), antonímia (oposição conceitual) coordenação (termos subordinados a um mesmo conceito - TG)</p>	<p>Associativa: ação/produto, causa/efeito, afinidade, características do produto.</p>

3 THESA – Tesauro Semântico Aplicado

O Tesauro Semântico Aplicado (THESA) foi desenvolvido com o objetivo de disponibilizar um instrumento de elaboração de tesauros para os estudantes de graduação em Biblioteconomia da UFRGS utilizarem na disciplina de Linguagens Documentárias III. Deseja-se com isso reduzir o trabalho operacional dos alunos e dar maior atenção ao trabalho de desenvolvimento cognitivo e conceitual referente a modelagem do domínio. Seu desenvolvimento baseou-se nas normas ISO e NISO vigentes, de forma a compatibilizar suas diretrizes com os requisitos semânticos prementes nas novas demandas dos SOCs. Com base na literatura disponível, nas normas de construção de tesauros da ISO e NISO foram identificados os elementos necessários para o desenvolvimento do protótipo, principalmente no que tange ao levantamento das propriedades de ligação entre os conceitos. O software funciona em ambiente Web e pode ser utilizado gratuitamente, para fins didáticos em disciplinas dos cursos de graduação e pós-graduação ou para uso profissional (Gabriel Junior & Laipelt, 2017).

A versão beta (de teste) do Thesa pode ser acessada no endereço <http://www.ufrgs.br/tesauros>.

Destacamos que com o Thesa é possível descrever relações específicas, viabilizando desta forma o compartilhando de estrutura e dados em sistemas complexos, como o Linked data, e ainda manter a compatibilidade com modelos simplificados como da ISO e NISO. Sua estrutura é baseada na concepção das relações entre os conceitos, partindo do pressuposto que um conceito pode ser representado por um termo, uma imagem, um som, um link ou qualquer outra forma que possa ser explicitada. Desta maneira, o conceito é perene, enquanto a sua

representação pode variar conforme o contexto histórico ou social, sendo definida uma forma preferencial, e inúmeras formas alternativas e ocultas.

4 Metodologia

Em consonância com nosso objetivo de desenvolver um modelo de metacategorias semânticas para aplicação em SKOS, este trabalho parte de pressupostos teóricos visando uma aplicação. Assim, a partir da sistematização e descrição de tipologias de relações semânticas disponíveis na literatura da área de organização do conhecimento realizamos a implementação das mesmas no software Thesa.

O Thesa já é utilizado por 277 usuários e faz a curadoria de 194 tesauros. No entanto, muitos deles não estão visíveis no site do software, visto que a publicação ou não dos tesauros prontos é uma opção dos seus autores. A inserção das metacategorias semânticas identificadas até o momento possibilitou a observação e descrição do funcionamento das mesmas em diferentes domínios do conhecimento. Para demonstração das relações semânticas inseridas no software utilizamos três tesauros elaborados por alunos do Curso de Biblioteconomia da UFRGS, são eles:

Comunicação Científica – elaborado por Priscilla Rodrigues Pereira

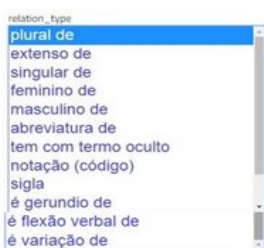
Eletromagnetismo – Elaborado por Viviane Marques

The Witcher – Elaborado por Osmar Wey

5 Resultados

O Thesa foi desenvolvido inicialmente para fins acadêmicos, principalmente para utilização dos estudantes do curso de Biblioteconomia da UFRGS. Porém, tendo em vista o potencial do sistema para a gestão de vocabulários percebeu-se a importância de aperfeiçoá-lo para utilização tanto com fins acadêmicos como da comunidade profissional de bibliotecários. A consolidação dos resultados desta pesquisa possibilitará o desenvolvimento de tesauros semanticamente fortalecidos, numa transposição natural entre a teoria e a prática. Na figura 1 apresentamos as relações associativas e de equivalência inseridas no Thesa até o momento.

Relações de Equivalência



Relações Associativas

```
... select_the_relation
Coordenação
ação/produto
causa/efeito
oposição
afinidade
parte de
características do produto
```

Figura 1. Relações Associativas e de Equivalência inseridas no Thesa

Nas figuras 2, 3 e 4 pode-se observar a aplicação dessas relações semânticas inseridas no Thesa em diferentes domínios do conhecimento.

Campo Magnético

<https://www.ufpa.br/tesauro/index.php/thesa/c/12334>

TG: Magnetismo

TR (Coordenação), Corrente induzida
 TR (causa/efeito), Energia Magnética
 TR (ação/produto), Força de Lorentz
 TR (ação/produto), Força Magnética
 TR (Coordenação), Lei de Ampère
 TR (ação/produto), Lei de Biot e Savart
 TR (causa/efeito), Lei de Lenz
 TR (ação/produto), Linha de Indução
 TR (Coordenação), Carga Elétrica
 TR (causa/efeito), Corrente Elétrica

Definição do conceito

Todo campo magnético se estabelece por cargas elétricas em movimento.

Definição do conceito

Cargas elétricas em movimento (corrente elétrica) criam, no espaço em torno delas, um campo magnético.

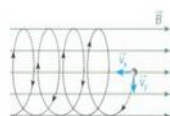


Figura 2. Tesauro Eletromagnetismo. (MARQUES, 2018).

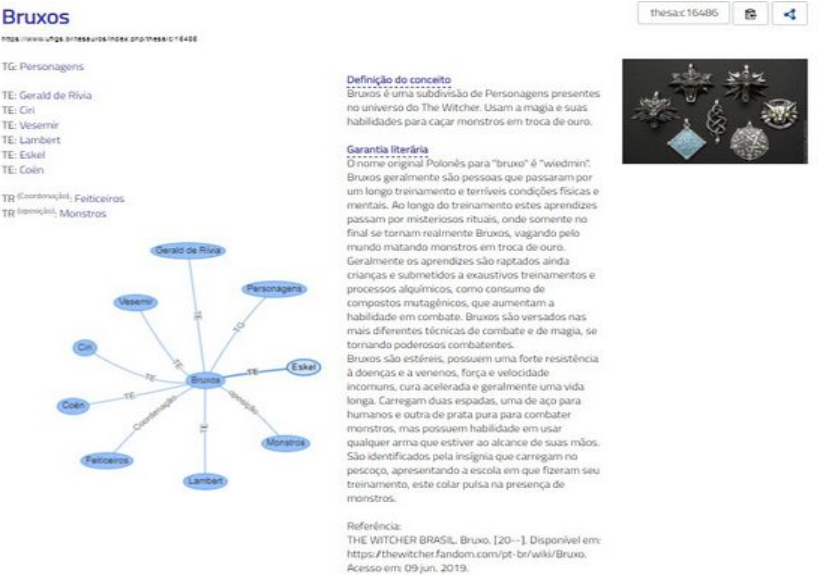


Figura 3. Tesauro The Witcher. (WEYH, 2019).



Figura 4. Tesauro de Comunicação Científica. (PEREIRA, 2018).

6 Considerações Finais

Espera-se que os resultados gerados com a atualização do software façam com que SRI incorporem a utilização do mesmo, e assim seja possível futuramente a contextualização da informação, tanto por parte do usuário como do profissional da

informação. Dos futuros resultados do uso do Thesa, está a possibilidade de estabelecer dicionários de sinônimos entre diversos domínios e temas, bem como gerar inferências com bases em outros tesauros, facilitando a operacionalização de seus gestores, com o uso de Inteligência Artificial.

Referencias

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1-22.
- Café, L. & Brascher, M. (2011). Organização do Conhecimento: Teorias Semânticas como base para estudo e representação de conceitos. *Informação e Informação*, 16(3), 25-51.
- Carlan, E. (2010). *Sistemas de organização do conhecimento: uma reflexão no contexto da Ciência da Informação*. 2010. 195 f. Dissertação. (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação. Brasília: Universidade de Brasília.
- Gabriel Junior, R. F. & Laipelt, R. C. F. (2017) Thesa: ferramenta para construção de tesouro semântico aplicado interoperável. *Revista P2P e Inovação*, 3(2), 124-145.
- Gnoli, C. (2008). Tem long-term research questions in knowledge organization. *Knowledge Organization*, 35(2/3), 137-149.
- Green, R. (2001). Relationships in the organization of knowledge: an overview. In Bean, A & Green, R. (Orgs.), *Relationships in the Organization of Knowledge* (pp. 3-18). Boston/Dordrecht/London: Kluwer Academic Publishers.
- HJØRLAND, B. (2007). Semantic and Knowledge organization. *Arist*, 41, 367-405.
- HJØRLAND, Birger. (2015). Theories are Knowledge Organizing Systems (KOS). *Knowledge Organization*, 42(2), 113-128.
- HODGE, Gail. (2000). *Systems of knowledge organization for digital libraries : beyond traditional authority files*. Washington, D.C.: The Digital Library Federation Council on Library Information Resources.
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). *Functional Requirements for Bibliographic Records (FRBR)*, final report. Retrieved from <http://www.ifla.org/VII/s13/frbr/>
- ISKO UK AGM. (2015). *The Great Debate: This House believes that the traditional thesaurus has no place in modern information retrieval*. 19 February 2015 London.
- ISKO, UKChapter.(2015) Annual General Meeting. Retrieved from <http://www.iskouk.org/content/great-debate#EventProgramme>

- ISO. International Organization For Standardization. (2011) ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 1.
- ISO. International Organization For Standardization. (2013) ISO 25.964-1. Information and documentation – Thesauri and interoperability with other vocabularies: Part 2.
- ISO. International Organization For Standardization. (1986). ISO 2788:1986. Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri.
- Kless, D., Simon, M., Kazmierczak, E., & Jutta, L. (2015). Thesauri and Ontology Structure: Formal and Pragmatic Differences and Similarities. *Journal of the Association for Information Science and Technology*, 66(7), 1348-1366.
- Lima, G. A. & Maculan, B. C. M. S. (2017). Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. *Ciência da Informação*, 46(1), 60-72, 2017.
- Marques, V. (2018). *Thesa: Eletromagnetismo*. Porto Alegre : UFRGS. Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa/terms/75>.
- Miles, A. & Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Reference. *W3C Recommendation*. Retrieved from <http://www.w3.org/TR/skos-reference/>
- Miles, A. & Brickley, D. (2005). *SKOS Core. Guide W3C*: 2005. Retrieved from <https://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/>
- NISO National Information Standards Organization. (2010) ANSI/NISO Z39.19-2005 (R2010) - Guidelines for the construction, format, and management of monolingual controlled vocabularies.
- ORBST. (2011). The need for ontologies: Bridging the barriers of terminology and data structure. *Special Paper of the Geological Society of America*, 482, 99-123.
- Pereira, P. R. (2018). *Thesa: Comunicação Científica*. . Porto Alegre : UFRGS. Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa/terms/104>.
- Prevot, L. et al. (2010). Ontology and the lexicon: a multidisciplinary perspective. In Huang, C. et al. *Ontology and the lexicon: a natural language processing perspective*. New York: Cambridge University Press.
- Ramalho, R. R. A. S. (2015). Análise do modelo de dados skos: sistema de organização do conhecimento simples para a web. *Informação & Tecnologia*, 2(1), 66-79.
- Weiss, L. C. & Brascher, M. (2013). Princípios teóricos para o estabelecimento de relações semânticas em tesauros. In *Encontro Nacional De Pesquisa Em Ciência Da Informação*, 14. Florianópolis, UFSC.
- Weyh, O. (2019). *Thesa: The Witcher*. Porto Alegre: UFRGS. Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa/terms/167>

Una terminología de teatro para organizar el patrimonio cultural universitario¹

Catalina Naumis Peña¹, Ariel Alejandro Rodríguez García² y Hugo Alberto Guadarrama Sánchez³

¹ ORCID [0000-0003-3152-3958](https://orcid.org/0000-0003-3152-3958). Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México, México.
naumis@unam.mx

² ORCID [0000-0001-5779-0371](https://orcid.org/0000-0001-5779-0371). Instituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México, México.
ariel@iibi.unam.mx

³ ORCID [0000-0002-1917-8015](https://orcid.org/0000-0002-1917-8015). Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, México.
bibliotekallica@comunidad.unam.mx

Resumen: El presente estudio tiene como objetivo identificar los términos funcionales en la comunicación sobre el tema del teatro para un sistema de información de espacios culturales universitarios en idioma español. En cuanto a la designación terminológica existe un conjunto de zonas, áreas, medios, instrumentos y herramientas para intercambiar entre tramoyistas, técnicos de sonido e iluminación, músicos, actores y artistas que deben ser manejadas por los responsables de la operación de los espacios y de las autoridades que programan actividades. La metodología se fundamentó en la búsqueda de los términos referentes al espacio teatral y la confrontación con el Art & Architecture Thesaurus (Getty Research Institute) Esta categorización permitió delimitar el universo del teatro de acuerdo con sus conjuntos y sus respectivos elementos, para establecer una estructura que jerarquiza el espacio teatral, con la finalidad de sustentar la organización del sistema de información en el que se insertará. Finalmente, se consultaron a los especialistas y autoridades para validar la propuesta.

Palabras clave: sistemas de información, terminología del teatro, patrimonio cultural universitario, recuperación de información, difusión cultural universitaria

Abstract: The objective of this study is to identify the functional terms in the communication on the theme of theater for an information system of university cultural spaces in the Spanish language. Regarding the terminological designation, there is a set of zones, areas, means, instruments and tools to exchange between stagehand, sound and lighting technicians, musicians, actors and artists that should be managed by those responsible for the operation of the spaces and the authorities that schedule

¹ Esta comunicación forma parte del proyecto apoyado con recursos PAPIIT IT400318

Naumis Peña, C., Rodríguez García, A.A., Guadarrama Sánchez, H.A. (2020). Una terminología de teatro para organizar el patrimonio cultural universitari. En J. Tramullas, P. Garrido-Picazo y G. Marco-Cuenca (eds.) *Actas del IV Congreso ISKO España y Portugal 2019* (pp. 419-429). <https://doi.org/10.5281/zenodo.3741974>

activities. The methodology was based on the search of the terms referring to the theatrical space and the confrontation with the Art & Architecture Thesaurus (Getty Research Institute) This categorization allowed to delimit the universe of the theater according to its sets and their respective elements, to establish a structure that hierarchizes the theatrical space, in order to support the organization of the information system in which it will be inserted. Finally, the specialists and authorities were consulted to validate the proposal.

Keywords: information systems, theater terminology, university cultural heritage, information retrieval, university cultural dissemination

1 Introducción

Las Industrias Culturales y Creativas (ICC) significan la manifestación intrínseca de la capacidad de la humanidad para pensar, crear, innovar y diseñar productos, experiencias y servicios, generando un valor económico dentro de un contexto comercial (Peris-Ortiz, Cabrera-Flores y Serrano-Santoyo, 2019, p. 3), que es algo opuesto a la concepción donde, el patrimonio cultural complementa el modelo de formación de los estudiantes, comunidad universitaria y público en general.

La Universidad Nacional Autónoma de México (UNAM) al disponer de varios campus en muchas ciudades de México y del extranjero plantea entre otros retos, manejar la información concerniente a los espacios destinados a la cultura.

Por tanto, la Coordinación de Difusión Cultural de la UNAM ha desarrollado un sistema de información sobre los espacios y activos culturales: Sistema de Información para el Registro Universitario de Espacios y Activos Culturales (SI-RUEyAC), con la asesoría del Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI). El rol de la información y la comunicación sobre el particular tiene como objetivo prioritario expandir el uso y acceso cultural de la comunidad universitaria hacia el desarrollo del individuo para concretar su proyecto de aprendizaje, a través de la programación de actividades. Entre los problemas detectados que se tratan de resolver mediante este proyecto es el lenguaje utilizado en el sistema de información.

De acuerdo con el SI-RUEyAC a la cabeza del árbol de dominio que abarca su sistema están los principales términos que agrupan los espacios universitarios existentes, que es bastante más amplio que sólo el teatro especializado: teatro multiuso, recintos especializados, explanadas y foros al aire libre. Las razones para encabezar los temas a través de esta propuesta son las necesidades de este sistema de información, ya que, en los planteles de educación media, de educación superior e institutos de investigación abundan los teatros multiusos o auditorios. Entre los recintos especializados existen teatros en los que se presentan obras para público en general y deben cumplir con los requisitos que se exigen en ellos. Una de las ideas que subyacen con el sistema es que sirva para evaluar la capacidad de los teatros multiusos o auditorios para ser usados en representaciones profesionales.

En el caso del presente trabajo se analizarán únicamente las relaciones jerárquicas de los términos del teatro bajo los recintos especializados teatrales y no en el sentido amplio de teatro. En este trabajo se presenta sólo una parte del tesoro que se elabora para sustentar el SI-RUEyAC y se explica cómo se identificaron los términos cuyos nodos conceptuales resuelven la situación comunicativa en el tema del espacio teatral.

Las obras teatrales, los ciclos, las ferias del arte y los festivales requieren de una plataforma, escenario o estructura para desplegarse y desarrollarse, es decir de un espacio actoral (Pavis, Patrice, 1998, p. 169).

Los espacios en donde se presentan las actividades y expresiones culturales poseen diversos componentes mecánicos, eléctricos y electrónicos orientados a la creación de atmósferas y dinámicas en las puestas en escena, dado que cada proyecto u obra teatral tiene características particulares por su naturaleza temática y público. Al nombrar las zonas, áreas, medios, instrumentos y herramientas para intercambiar información con los responsables de los espacios existe la posibilidad de conflictos a partir de la confusión y omisión de elementos, como resultado de esa misma diversidad terminológica ya sea de oficio, disciplinaria o profesional en el ámbito teatral (Dinulović, 2015, p. 4).

En el teatro cada especialista usa las denominaciones comunes a su área, que a veces no coinciden con las otorgadas por quienes programan las actividades. En este complejo plano de interacción concurren además las autoridades universitarias y los responsables de los espacios en general, que deben tomar las decisiones referentes a los usos espaciales y los activos correspondientes para cumplir con la actividad teatral.

El lenguaje coadyuva a las diversas aristas disciplinarias y administrativas que se relacionan en dicho contexto, para obtener un mejor entendimiento y comprensión acerca de cada elemento nombrado en el teatro, en donde interactúan músicos, cantantes, actores, tramoyistas o maquinistas, carpinteros, escenógrafos, utileros, vestuaristas y demás, cada uno de ellos comunicándose entre sí, a veces con confusiones e imprecisiones (Theatre projects consultants, 2016, p. 2).

El estudio se centra en la terminología del espacio teatral y en las estructuras en donde se lleva a cabo la actividad histriónica y áreas destinadas al público, no se incluyen términos relativos a las estructuras internas de las obras dramáticas, como por ejemplo las entradas y salidas de los actores frente al público o tras bambalinas, en forma física o a través únicamente de la voz, diferencias entre tipos de personajes o los movimientos de los actores en el escenario (Meierkhold, 1994, p. 152).

Los términos que se incluyen están relacionados con los objetivos del sistema de información desarrollado, la terminología de un dominio se conforma con el vocabulario integrado de, por lo menos, dos anillos: uno que podría denominarse como nuclear, y otro afin (Barité, 2017, p. 470).

2 Metodología

La investigación terminológica se realiza en un ámbito de conocimiento, buscando la representación de los fenómenos que ocurren al interior. Por ello cada término es investigado en su vehiculación de significado y en la garantía de uso.

Una vez estudiado el espacio físico a representar se investigaron las designaciones en las obras especializadas como los glosarios, los diccionarios, las enciclopedias, los vocabularios controlados existentes e inclusive en trabajos académicos universitarios en idioma español, confrontados con los usos en la práctica. Los términos obtenidos son contrastados en las entrevistas realizadas a los responsables de los teatros que han participado en el levantamiento de información llevado a cabo en los espacios.

En el caso del teatro el vocabulario controlado más difundido y usado es el Art & Architecture Thesaurus (Getty Research Institute). Este tesoro (Tesauro del P. G.) será continuamente consensado en relación a los términos a incorporar a la base de datos que se construye. Se consulta tanto la versión en inglés como la traducción al español.

El primer acercamiento a los términos de teatro se realizó a partir del mencionado tesoro, como se habrá de explicar a través de los ejemplos donde se detecta incompatibilidad con el sistema que se construye. Las equivalencias en el español del Getty con las usadas por la comunidad de hablantes del ámbito de difusión cultural universitario para referirse a los aspectos y características del teatro en México no son comprensibles en todos los casos, basta mencionar que la designación que en México se conoce como “sala del teatro” en el Getty se la designa como “espacio de auditorio”.

La terminología que se estudia estará integrada en un tesoro que comprende los espacios culturales universitarios en general. En los planteles de enseñanza se generaliza como espacio cultural al auditorio. Si el término utilizado en los recintos profesionales de teatro es “espacio de auditorio” se confunde la comunicación. De esta manera, no quedará especificada la conformación de la sala de teatro al interior del teatro y se presentaría en la misma categoría del espacio de auditorio que es mucho más general. Por ejemplo, un espacio de auditorio puede comprender el lugar al aire libre donde colocar sillas para comodidad de los asistentes.

Finalmente, después del estudio comparativo se propuso la terminología que será discutida con las autoridades de difusión cultural, los técnicos de teatro y especialistas para explorar los términos validados por la comunidad y establecer las relaciones funcionales en la comunicación, que asegure la comprensión por los diferentes tipos de usuarios a los que se destina el sistema de información en construcción.

3. Relación de términos entre el Art & Architecture Thesaurus (Getty Research Institute) y las necesidades del SI-RUEyAC

El propósito del Tesauro de Arte y Arquitectura del Paul Getty (Tesauro del P. G.) es constituir un recurso estructurado con el fin de mejorar el acceso a la información referente al arte y a los recintos culturales. Sin embargo, la organización de sus

metadatos y enlaces tiende a combinar distintos conjuntos temáticos al momento de realizar la recuperación de la información, provocando cierta confusión en su consulta a no especialistas en información. El propósito del SI-RUEyAC es programar las actividades culturales en una universidad.

En el análisis de los usos terminológicos utilizados en el SI-RUEyAC y luego de revisar la traducción de los términos al español del Tesaurus del P. G., así como los términos en inglés, a pesar de que sus términos cumplen con los estándares de valor para catalogar, no podrían ser utilizados en el contexto a resolver, dado que sus transliteraciones al español no corresponden del todo a los significados de los espacios y activos culturales contemplados. Del mismo modo la representación esquemática de la terminología teatral carece de algunos elementos y excede de otros, como por ejemplo los vinculados al teatro griego, de manera que, hay mayor presencia en el Tesaurus del P. G. de los términos que representan a los teatros al aire libre.

La UNAM carece de teatros al aire libre (se ha detectado uno) y si bien hay teatros especializados, en general son teatros multiusos o auditorios que se deben adaptar a diferentes funcionalidades. Por esta razón las características de los teatros especializados deben quedar bien definidas a través del vocabulario controlado que se construye, para hacer las comparaciones necesarias a la hora de la programación.

En un primer acercamiento al Tesaurus del P. G. se observa que las traducciones al español siguen el razonamiento utilizado en inglés, esto se percibe por el uso de la palabra teatro que en inglés es más amplia que en el español y sin embargo se usa en la traducción, según la tradición en inglés.

La primera jerarquía donde se asienta el teatro como espacio o edificio es precedido por la clasificación obtenida de la copia del Tesaurus del P. G. y que parte de Built Environment, con las siguientes meronimias, miniature theaters, puppet theaters y por último theater elements:

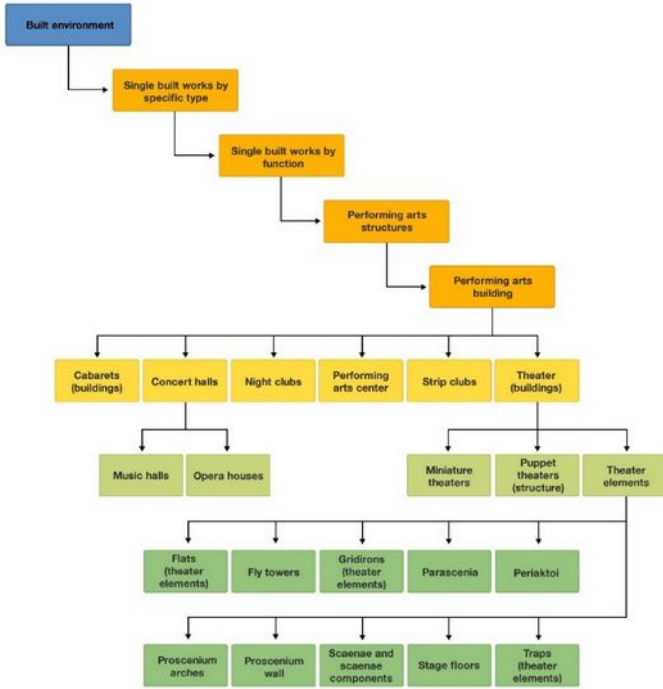


Figura 1: Ubicación jerárquica del Teatro como espacio en el Tesauro del P.G. Fuente: Adaptación de Luis Enrique Sánchez Rodríguez, a partir de la extracción del tesauro del P. G. con fecha del 18 de mayo del 2019.

Como se puede observar la relación jerárquica para llegar a Teatro como espacio es demasiado extendida. En la traducción al español se presenta el mismo esquema terminológico. En el Tesauro del P. G. <performing space building>, es decir el espacio escénico del teatro se comparte con los cabarets o las discotecas y los clubs de striptease. Tampoco interesan las meronimias establecidas para los teatros como espacios, porque los elementos del teatro comparten relaciones con los teatros en miniatura y los teatros de marionetas. En el SiRUEyAC los elementos del teatro como espacio (delimitado como tal) deben ser separados en áreas: pública, técnica, escénica y administrativa. Se parte de una taxonomía de la UNAM, donde cada sitio ingresado está relacionado con la dependencia donde se encuentra el espacio escénico, porque el objetivo prioritario es el aprovechamiento de los espacios disponibles en la universidad y conocer si son utilizables en la actividad que se programa.

De esta manera, el Tesauro del P.G. como herramienta de organización no es idónea al estar constituido por términos en inglés refiriéndose a un vocabulario de muchísima amplitud, mucha más de la necesaria para estructurar el sistema de información que se construye.

De acuerdo con el Getty Research Institute que sustenta el tesauro, éste crece constantemente mediante las aportaciones voluntarias, y esta sería una ocasión de

colaborar con el mismo. El universo de los espacios y activos culturales de la UNAM se delimita conceptualmente a partir de la creación de conjuntos o categorías y subcategorías con los términos en el español usado en México, basados en la garantía oral y literaria.

4 Análisis terminológico del teatro como espacio cultural

El teatro es la representación de ideas (texto), en un determinado espacio (escena), a través de la acción del cuerpo (expresión corporal), la cara (gesto) y la integración de la palabra (fonética), con la utilización de los códigos socialmente aceptados (o no), para estimular los sentidos y obtener reacciones (comunicación), hacia el medio o entorno (Psicosocial), estas reacciones van desde la reflexión hasta la pura y simple diversión, entretenimiento, recreación y esparcimiento (Saavedra, 2018).

Al transcurrir en un espacio en el que concurre la representación teatral con el público que lo observa, es necesario contar con las características adecuadas para el impacto del trabajo actoral frente al público. Ello significa que la programación de actividades debe considerar que los espacios propuestos deben de contar con los elementos que permitan la interacción y el lenguaje de unos y otros para el transcurso ordenado del proceso de presentación del espectáculo.

De manera que, se parte de un listado de términos obtenidos de programas de trabajo, textos sobre la difusión teatral, bitácoras de presentación de obras y textos sobre la actividad teatral. Estos elementos constituyen la garantía literaria que opera en los términos usados sobre el particular. De acuerdo con Barité (2011), “la garantía literaria proporciona los términos de referencia y el control del vocabulario les otorga la configuración definitiva y las relaciones” (p. 141).

Cada uno de los términos es confrontado en diccionarios de la especialidad y cuando las definiciones son amplias, se incluyen las acepciones que interesan al proyecto. Cada término es acompañado de una definición y cuando éstas no son claras para el sistema de información se incluyen notas de alcance, que explican el significado otorgado (Bails, 1802), (Plazola, 1996, p. 137), (Pavis, 1998, p. 169), (Romero y Pardo, 2003), (Zacarias, 2003, p. 263 y 264), (Pérez y Gardey, 2008), (Ferrera, 2009, p. 856, 1341 y 1456), (Especialidad de escenografía: Escuela Superior de Arte Dramático de Sevilla, 2011), (Consejo Nacional de la Cultura y las Artes, 2013), (Laiz, 2013), (Zertuche y Borrego, 2015), (Gaitán, Melina, 2016), (Glosarios servidor, 2016), (Theater projects, 2016), (Quizlet, 2019), (Departamento de composición ETSAB-UPC. Cuarto de control de iluminación, 2019) y (Real Academia Española, 2019).

Una vez que se recopilaban la generalidad de los términos y sus definiciones fueron agrupados bajo teatro, en las categorías mencionadas anteriormente: área pública externa, área pública interna, área escénica, área administrativa y área técnica. Ello con el fin de poder delimitar el universo del teatro de acuerdo a sus conjuntos y sus respectivos elementos, para establecer una estructura que jerarquice el espacio teatral, relaciones de equivalencia y asociativas entre descriptores que expresan cierta afinidad semántica entre conceptos distintos. Finalmente, el punto de vista más

influyente y fundamental será el de los especialistas en el tema para validar y obtener el producto intelectual, capaz de responder a las necesidades de comunidades interesadas en el tema del teatro en el idioma español, a quienes se someterá la siguiente propuesta.

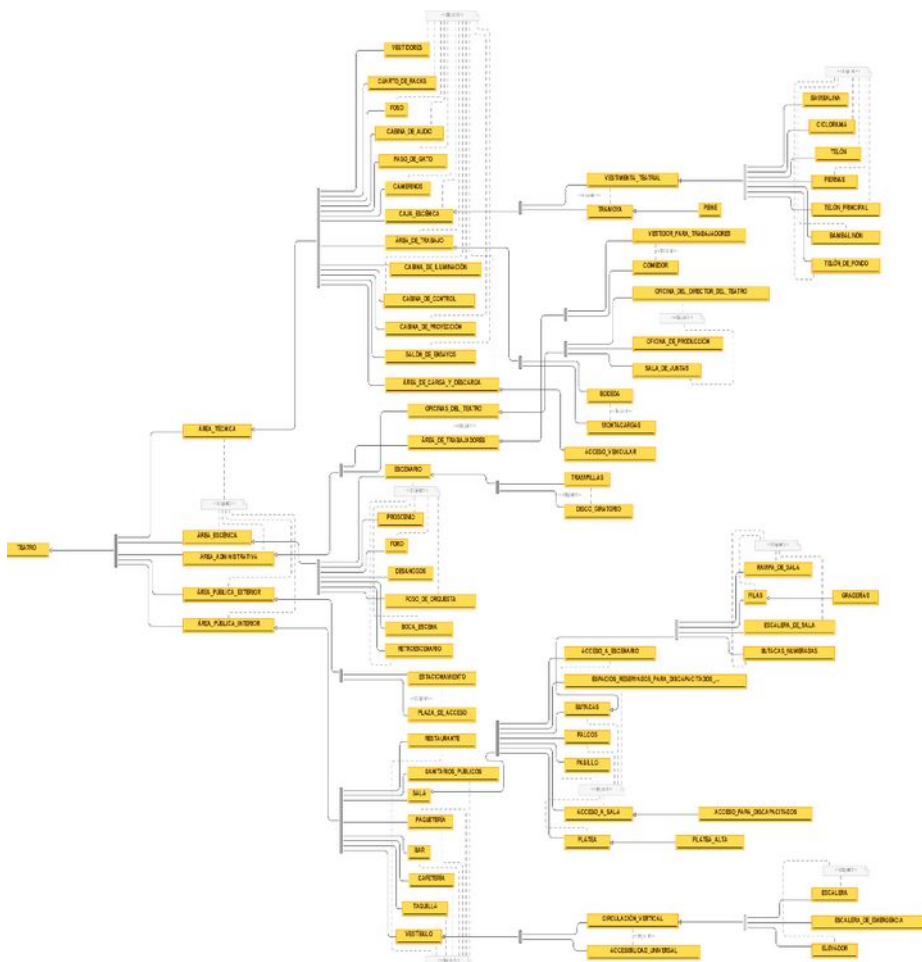


Figura 2. Propuesta del árbol de dominio sobre teatro para la base de datos del SI-RUEyAC

5 Discusión de los resultados

La agrupación en áreas del teatro ayudará posteriormente a la organización de los otros grupos de recintos que deberán utilizarse para complementar los espacios culturales que posee la universidad. Sin duda, los auditorios no contarán con un

desarrollo tan amplio de las áreas en que se distribuyen los términos del teatro, pero esa generalidad permitirá considerar aspectos menores que también se presentan en los auditorios. Por ejemplo, la programación exige conocer la estructura del área administrativa que gestiona los espacios en cada dependencia o las posibilidades del área técnica para acceder desde el exterior directamente al escenario, o tener conocimiento de la existencia de un vestíbulo o plaza de acceso donde se pueda programar un brindis.

El área pública interna del dominio (figura 2) está dividida en Sala y Vestíbulo. Sin embargo, el acceso para discapacitados está incluido en la sala, aunque contempla las facilidades que puedan existir para acceder al escenario o al vestíbulo que son elementos que pertenecen a otras categorías. Las decisiones se adoptan de acuerdo con un área, pero en el tesoro se relacionarán con las otras que intervienen en el uso de los espacios.

Otros detalles que se observan es que no todos los teatros disponen de vestidores, cuyo sentido en este sistema de información es el lugar donde se agrupan las vestimentas utilizadas en las diferentes presentaciones llevadas a cabo, a diferencia de los camerinos que son más privados o de los vestidores de los trabajadores que se incluyen bajo la categoría del área de trabajadores. Otro término que sobresale es la vestimenta teatral que no está relacionada con la vestimenta que usan los actores, sino que se relaciona con el escenario y los telones.

Si bien existe una categoría general que es el área externa, también en el área técnica se incluye un área externa, pero exclusivamente relacionada con el área técnica porque es el área de carga y descarga.

Cómo se observa la estructura básica jerarquizada de los términos del teatro es sencilla y directa para facilitar la organización de la información incorporada al sistema, a diferencia del Tesoro del P. G. cuya estructura involucra términos innecesarios en el SI-RUEyAC.

6 Consideraciones finales

La terminología final estará insertada en un sistema de información que incluye todo tipo de espacio cultural y los activos de infraestructura que incluyen, donde se puedan programar actividades universitarias, no únicamente de espectáculos como el teatro.

En el medio universitario mexicano se incluyen teatros (referidos a los teatros profesionales en el que se puede presentar una obra teatral), salas de conciertos y auditorios, además de espacios al aire libre como foros y explanadas, que se puedan adaptar a representaciones, como es el caso del carro de comedias.

El sistema de información en la etapa del proyecto que se desarrolla se alimentará usando el software MultiTes versión 2018.10.30u. Sin embargo, a futuro se piensa integrar la terminología del tesoro en una ontología que se construirá para incrementar el potencial del sistema y hacerlo más funcional, por lo que, desde ahora, se incluyen definiciones de cada término para los elementos que se requieran a futuro. (Barber, 2018, p. 14).

La propuesta que se presenta resuelve una situación comunicativa, pero es aplicable a sistemas de información que requieran de la terminología teatral en español.

Referencias

- Avita, A. (2011). *Teatro para principiantes*. México: Porrúa.
- Barber, E., Pisano, S., Romagnoli, S., de Pedro, G., Gregui, C., Blanco, N., & Mostaccio, M. (2018). Metodologías para el diseño de ontologías Web. *Información, cultura y sociedad*, 39, 13-36. Recuperado de <http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/5142>
- Barité, M. & Rauch, M. (2017). EL árbol de dominio como herramienta en organización del conocimiento: aplicación al área de educación superior. En *Perspetivas de investigação em representação e organização do conhecimento: Atualidade e tendencias*. Recuperado de http://sci.uc.pt/eventos/atas/comunicacoes/isko2017/isko2017_468_477.pdf
- Barité, M. (2011). *La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento: modelo y aplicación*. (Tesis doctoral). Universidad de Granada. Recuperado de <http://digibug.ugr.es/handle/10481/17583>
- Consejo nacional de la cultura y las artes (2013) El escenario: lugar de trabajo en equipo. [PDF]. Recuperado de http://www.adtres.cl/files/4814/2543/5438/el_escenario_vol3.pdf
- Dinulović, R. (2015). Space in the 20th Century Theatre: 1. Theatre and Architecture. *SEE J Archit Des*. 2015, 1-6. Recuperado de <http://www.id-press.eu/seejad/article/view/seejad.2015.10006/pdf>
- Escuela Superior de Arte Dramático de Sevilla. (2011). Glosario de términos. Sevilla, España: Especialidad de escenografía. Recuperado de <http://esadsevillaescenografia.blogspot.com/2011/01/glosario-de-terminos.html>
- Ferrera, J. (2009). *Glosario ilustrado de las artes escénicas*. España: Ediciones y Gráficas Solapas.
- Gaitán, M. (2016, mayo 23). *Historia de la arquitectura teatral*. [Video]. Recuperado de <https://www.youtube.com/watch?v=qSL5YWf6iGk>
- Glosarios de términos especializados. (2019). Página de inicio. En *Glosarios de términos especializados de las Ciencias, las Artes, las Técnicas y la Sociedad Alicante*. Recuperado de <https://glosarios.servidor-alicante.com/teatro/escenario>
- Laiz, M. (2013). El teatro. [Sitio web]. Recuperado de <https://es.slideshare.net/lululululululu/las-partes-del-teatro>
- Luneta. (1802). En *Diccionario de Arquitectura Civil*. Recuperado de [http://enciclopedia.us.es/index.php/Luneta_\(DACBB\)](http://enciclopedia.us.es/index.php/Luneta_(DACBB))
- Meierkhöld, V. (1994). *El actor sobre la escena: diccionario de practica teatral*. México: UAM: Grupo editorial gaceta.
- Pavis, P. (1998). *Diccionario del teatro*. Barcelona; México: Paidós.
- Peris-Ortiz, M., Cabrera-Flores, R. & Serrano-Santoyo, A. (2019). *Cultural and creative industries: A path to Entrepreneurship and Innovation*. Suiza: Springer. Recuperado de <https://doi.org/10.1007/978-3-319-99590-8>
- Plazola, A. (1996). *Enciclopedia de arquitectura Plazola*. Estado de México: Plazola Editores.
- Real Academia Española. (2019). Página de inicio. En *Real Academia Española*. Recuperado de <https://dle.rae.es/?w=diccionario>
- Romero, M. & Pardo, A. (2003). *Diseño de interior de teatros*. (Tesis de licenciatura). Universidad del Istmo. Recuperado de <http://glifos.unis.edu.gt/digital/tesis/2003/8515.pdf>

- Saavedra, R. (2018). Terminología básica del teatro. [Sitio web]. Recuperado de <http://eurekateatro.ve.tripod.com/grupoeurekateatro/id6.html>
- Sutherland, A. (2019). *Partes del teatro*. [Diapositivas]. Recuperado de <https://quizlet.com/85707091/partes-del-teatro-flash-cards/>
- Theatre projects (2016). Partes de un teatro. [PDF]. Recuperado de http://theatreprojects.com/files/pdf/resources_partsoftheatrebuilding_ES.pdf
- Universitat Politècnica de Catalunya Barcelonatech. (2010). Cuarto de control de iluminación. Barcelona, España: Departamento de composición ETSAB-UPC. Recuperado de <http://espaciosescenicos.org/Cabina-de-control-de-iluminacion>
- Zacarías, P. (2003). La arquitectura de los teatros veracruzanos durante el Porfiriato. México: Universidad Veracruzana.
- Zertuche, C. & Borrego, A. (2015, septiembre 8). *Partes del Teatro*. [Video]. Recuperado de <https://www.youtube.com/watch?v=UKoC1O-SjOQ>

Análises de Palavras-chave como Fonte de Dados para Obtenção de Conhecimento Sobre a Evolução da Ciência

Jether Oliveira Gomes¹, Thiago Magela Rodrigues Dias², Gray Farias Moita³ y Adilson Luiz Pinto⁴

¹ ORCID [0000-0002-4096-0837](https://orcid.org/0000-0002-4096-0837). Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil.
jethergoliveira@gmail.com

² ORCID [0000-0001-5057-9936](https://orcid.org/0000-0001-5057-9936). Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil.
thiagomagela@cefetmg.br

³ ORCID [0000-0002-6510-1019](https://orcid.org/0000-0002-6510-1019). Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG, Brasil.
gray@dppg.cefetmg.br

⁴ ORCID [0000-0002-4142-2061](https://orcid.org/0000-0002-4142-2061). Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil.
adilson@cin.ufsc.br

Resumo. Serviços como bibliotecas digitais e sítios para registro individual de produção científica são alguns exemplos de como a internet tem contribuído na quantidade de trabalhos publicados, permitindo que usuários não apenas acessem conteúdo disponível, mas também possam registrar a sua produção científica. Neste contexto, pesquisadores de todos os domínios têm dedicado esforços com intuito de analisar a produção científica sobre diferentes perspectivas, como: análises de citações, indicadores de produtividade, colaboração científica e análise de tópicos baseados em termos extraídos de resumos, títulos e palavras-chave de artigos científicos. Dentre os trabalhos encontrados na literatura, geralmente as análises realizadas utilizam repositórios internacionais, que são específicas de uma determinada área ou periódico. No entanto, por tratarem de análises específicas e utilizarem de repositórios internacionais, não podem representar de forma abrangente o que é produzido no Brasil. Com isso, analisar fonte de dados que englobe diversos tipos de publicações, principalmente em veículos nacionais de diversas áreas, passa a ser uma tarefa relevante para a compreensão da ciência brasileira. Assim sendo, é proposto neste trabalho uma análise temporal das principais palavras-chave existentes nos artigos científicos de todos os indivíduos com doutorado concluído que possuem os currículos cadastrados na Plataforma Lattes.

Palabras clave: Plataforma Lattes, Bibliometria, Recuperação de Informação.

Abstract. Services such as digital libraries and individual registration sites are some examples of how the internet has contributed to the amount of published work, allowing users not only to access available content, but also to record their scientific output. In this context, researchers from all domains have dedicated their efforts to analyzing scientific production from different perspectives, such as: citation analysis, productivity indicators, scientific collaboration and analysis of topics based on abstracts, titles and words scientific articles. Among the works found in the literature, the analyzes generally use international repositories, which are specific to a particular area or periodical. However, because they deal with specific analyzes and use of international repositories, they can not comprehensively represent what is produced in Brazil. Thus, analyzing a data source that encompasses several types of publications, especially in national vehicles of different areas, becomes a relevant task for the understanding of Brazilian science. Therefore, a temporal analysis of the main keywords in the scientific articles of all the individuals with a doctorate completed that have the curricula registered in the Lattes Platform is proposed in this work.

Keywords: Lattes Platform, Bibliometric, Information Retrieval.

1 Introdução

O grande número de informações disponibilizadas pela internet e a sociabilização da atividade científica por parte de redes de pesquisadores são os fatores essenciais para o atual desenvolvimento da ciência (BRITO et. al., 2016). Para Dias (2016), serviços como bibliotecas digitais e sítios para registro individual de produção científica são alguns exemplos de como a internet tem contribuído na quantidade de trabalhos publicados, permitindo que usuários não apenas acessem conteúdo disponível, mas também possam registrar a sua produção científica.

Neste contexto, pesquisadores de todos os domínios têm dedicado esforços com intuito de analisar a produção científica sobre diferentes perspectivas, como: análises de citações, indicadores de produtividade, colaboração científica e análise de tópicos baseados em termos extraídos de resumos, títulos e palavras-chave de artigos científicos. Nesse último caso, um tópico pode ser entendido como termo que representa um dos assuntos associados a um determinado documento (BORGES et. al., 2015).

Os trabalhos que analisam tópicos também funcionam como uma boa revisão da literatura, o que permite, por exemplo, a verificação por parte do setor industrial se o que está sendo desenvolvido pela ciência contempla as necessidades da indústria (KHAN e WOOD, 2015). Para tanto, tais trabalhos geralmente exploram repositórios de artigos científicos, analisando seus títulos, resumos ou até todo o texto para extrair

tópicos de pesquisas e analisá-los através de análises bibliométricas ou técnicas de análises de redes sociais.

Diversos outros trabalhos têm utilizado análises bibliométricas e técnicas de redes sociais, para extração de conhecimento acerca do que tem sido desenvolvido nas mais variadas áreas de pesquisas com propósitos distintos, a saber: taxas de carbono (ZHANG et al., 2016), linhas de produtos de software (HERADIO et. al. 2016), criatividade (ZHANG et. al. 2015); educação matemática (FADIGAS et. al., 2009); biomedicina (MADLOCK-BROWN,2014); epidemiologia na Alemanha (PETER et. al., 2016); e; saúde no Brasil (PEREIRA et. al., 2007).

Dentre os trabalhos encontrados na literatura, geralmente as análises realizadas utilizam repositórios internacionais, que são específicas de uma determinada área ou periódico. No entanto, por tratarem de análises específicas e utilizarem de repositórios internacionais, não podem representar de forma abrangente o que é produzido no Brasil. Com isso, analisar fonte de dados que englobe diversos tipos de publicações, principalmente em veículos nacionais de diversas áreas, passa a ser uma tarefa relevante para a compreensão da ciência brasileira.

De acordo com Brito et. al. (2016), estudos desta natureza são considerados urgentes no Brasil e podem retratar o que é desenvolvido e publicado em ciência, tecnologia e inovação, possibilitando gerar parâmetros que podem nortear esforços e investimentos para impulsionar resultados de pesquisa. Em Trucolo (2016), o autor destaca que muitas vezes os investimentos focam em áreas de pesquisas já consolidadas e populares, nas quais se acredita que haverá retorno, ou ainda, identificadas como tendências globais. Entretanto, num país com dimensões continentais como o Brasil, uma estratégia interessante seria investir nas áreas e tópicos de pesquisas com os maiores potenciais de crescimento, ampliando as chances do retorno da investigação científica e canalizando os recursos, que na maioria das vezes são reduzidos.

Para análises sobre o patamar científico brasileiro, o repositório de dados da Plataforma Lattes é tido como um diferencial (LANE, 2010). Esse repositório é composto por dados de grupos de pesquisas, instituições e currículos de mais de cinco milhões de indivíduos (DIAS, 2016). Esses currículos concentram dados sobre formação acadêmica, áreas de atuação, trabalhos em anais de congressos e em periódicos, entre outros. A Plataforma Lattes também é fonte de informação para órgãos que avaliam o Sistema Nacional de Pós-Graduação do Brasil, e agências de fomento que financiam pesquisas e ofertam bolsas de estudos. Contudo, mesmo estando disponível livremente na internet, esses dados ainda não foram amplamente analisados (DIGIAMPIETRI, 2015). Dos trabalhos que exploram estes dados, poucos são os que analisam as palavras-chave das publicações científicas. Geralmente utilizam termos extraídos dos títulos dos artigos de um conjunto restrito de currículos, na tentativa de destacar os principais assuntos abordados.

Assim sendo, é proposto neste trabalho uma análise temporal das principais palavras-chave existentes nos artigos científicos de todos os indivíduos com doutorado concluído que possuem os currículos cadastrados na Plataforma Lattes.

Para tanto, inicialmente, as palavras-chave dos artigos publicados em anais de congressos e periódicos são extraídas e processadas, para posteriormente serem analisadas através de análises bibliométricas e técnicas de análises de redes sociais, para destacar os principais tópicos de interesses dos pesquisadores brasileiros de todas as grandes áreas do conhecimento ao longo de 55 anos de pesquisas registrados na base curricular da Plataforma Lattes.

2 Metodologia

A motivação da escolha da Plataforma Lattes como fonte de informação está relacionada basicamente a quatro fatores: (1) aos dados estarem disponíveis na internet e não terem sido amplamente analisados (DIGIAMPIETRI, 2015); (2) tratar da integração de dados de produções científicas de todas as áreas de C&T existentes na ciência brasileira ao longo de toda trajetória; (3) por não negligenciar os artigos publicados em periódicos nacionais que muitas vezes não são indexados, e também os artigos de anais de congresso (DIAS, 2016); e, (4) por ser uma poderosa fonte para fornecimento de dados de alta qualidade para medir e avaliar o desempenho acadêmico nacional (LANE, 2010).

Neste ponto, vale ressaltar que, apesar dos dados disponibilizados, estes são apenas visualizados através de uma interface de consulta que apresenta apenas os dados de um único currículo. Logo, para uma análise mais detalhada de grupos, instituições ou até toda a nação de cientistas brasileiros, técnicas e ferramentas para análises dos dados se fazem necessárias.

A aquisição dos currículos dos doutores na versão XML (eXtensible Markup Language) foi realizada através da utilização do LattesDataXplorer, desenvolvido por Dias (2016) para coletar os dados científicos contido nos currículos cadastrados na Plataforma Lattes.

Logo, a Figura 1 apresenta uma visão geral do arcabouço de componentes desenvolvidos que suporta as análises desejadas. Nela, os componentes “filtragem dos dados” e “tratamento dos dados” são responsáveis por todo o processo de seleção, tratamento e modelagem das informações dos currículos que realmente necessitam ser processadas para atingir os objetivos propostos, e, concomitantemente, diminuir o tempo de processamento computacional.

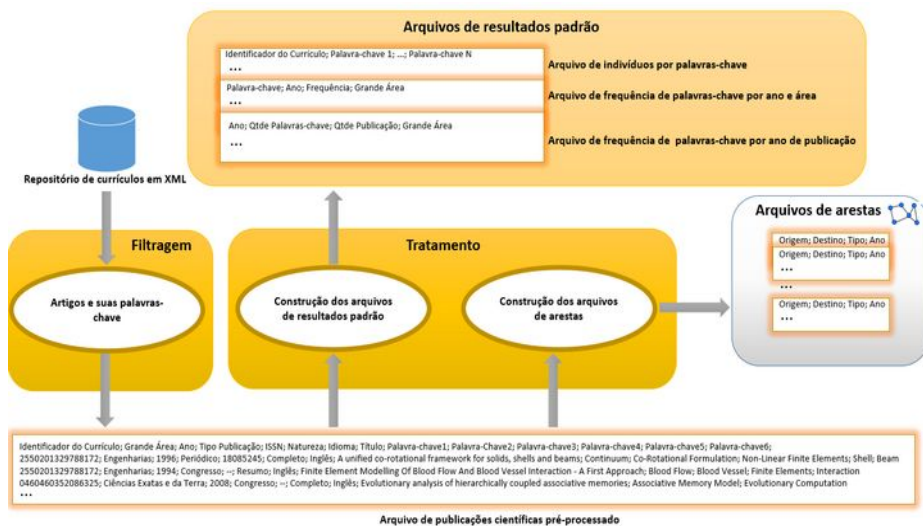


Figura 1. Processo de filtragem e tratamento dos dados.

O componente de “filtragem” realiza a etapa de mineração nos currículos para extrair as informações dos artigos, armazenando-as à parte num arquivo de publicações científicas, com isso, definindo o conjunto de dados centrais a serem estudados. As informações dos artigos incluem: grande área da publicação; ano de publicação; tipo de publicação; issn do periódico; idioma da publicação; título e palavras-chave.

Por outro lado, o componente de “tratamento”, processa os dados do arquivo de publicações científicas para tratá-los e caracterizá-los, e, a partir destes, constrói um conjunto dos arquivos de resultados padrão e arquivos de arestas para facilitar as análises. Esse componente realiza basicamente três etapas, a saber: limpeza dos dados, identificação de colaboração científica, e construção dos arquivos.

A limpeza dos dados serve para realizar o processamento das palavras-chave de tal forma a excluir os termos que não representam tópicos de estudos e agrupar as palavras que possuem mesmo valor semântico. Para isso, inicialmente, o método desenvolvido obtém as palavras-chave de cada artigo analisado. Em seguida, cada uma das palavras-chave são associadas ao idioma cadastrado para o artigo, servindo de referência no processo de radicalização. O processo de *lowercase* converte todas as palavras para minúsculo no intuito de padronizar o conjunto. No processo de *stopWords* são removidos os termos que não possuem valores semânticos. Posteriormente, no processo de normalização todos os acentos e pontuações são retirados das palavras-chave. Finalmente, o processo de radicalização é responsável pela redução da palavra-chave a seu radical. Contudo, em caso de palavras-chave compostas, este processo é executado em cada termo individualmente, e, concatenado formando uma única palavra.

3 Resultados

Os dados foram coletados em abril de 2017, totalizando 265.170 currículos de indivíduos com doutorado concluído. Para as análises, foram considerados os artigos únicos (colaboração) publicados em anais de congressos e em periódicos referentes ao período de 1962 até 2016, totalizando 10.040.664 artigos e 24.256.312 palavras-chave. Adicionalmente, vale destacar que a grande maioria dos currículos foram atualizados recentemente, onde 49,6% (131.660) possuem data de última atualização em 2017 e 73,3% (194.626) atualizados nos últimos dois anos. Neste ponto, vale ressaltar que o número de doutores representa cerca de 5% da quantidade total de indivíduos da Plataforma Lattes, porém, estes são responsáveis cerca de 70% de todos os artigos cadastrados nos currículos; que, aliado a diversidade e contemporaneidade de atualização destas informações, pondera a validade dos dados para auxiliar na compreensão sobre a evolução da produção científica brasileira (DIAS, 2016).

Inicialmente, todos os vértices (palavras-chave processadas) foram ranqueados de acordo com seu número de grau, e, posteriormente, todas as palavras-chave processadas tiveram suas frequências calculadas. Para facilitar a comparação entre as palavras-chave classificadas pelas duas medidas de importâncias utilizadas, a Tabela 1 apresenta as quinze principais palavras-chave ranqueadas pela medida de frequência e as respectivas ordem de ranqueamento considerando o grau do vértice referente ao conjunto histórico dos dados.

Tabela 1. Comparativo entre o ranqueamento das palavras-chave entre 1962 e 2016.

1962-2016	Frequência	Grau	R. Frequência	R. Grau
Educação	75.855	30.951	1	1
Formação do Professor	46.032	15.923	2	13
Enfermagem	40.228	13.710	3	25
Epidemiologia	40.158	25.398	4	3
Crianças	29.048	18.047	5	9
Políticas Públicas	28.100	15.929	6	12
Amazônia	27.626	23.800	7	4
Bovinos	27.236	16.883	8	10
Idoso	26.291	11.589	9	40
Ensino	25.760	14.268	10	18
Ratos	25.480	20.270	11	7
Brasil	24.981	28.527	12	2
Diagnóstico	24.520	22.172	13	5
Educação Ambiental	23.194	10.103	14	61
Cultura	22.123	15.026	15	15

Como pode ser observado, apesar de existir um alto percentual de palavras-chave iguais entre as principais ranqueadas pela frequência e grau, praticamente a ordem de destaque destas não são a mesma, exceto para “Educação” e “Cultura”. Diante disto, para melhor compreender a relação entre os resultados encontrados, uma análise de correlação foi realizada. Para tal análise, foi calculado o coeficiente de Spearman considerando todo o conjunto de palavras-chave entre frequência e grau. Os

resultados apresentam um coeficiente de correlação positivo de “0,9572” para frequência e grau.

Para conhecer os tópicos que se constituem como os principais assuntos de pesquisa em cada época, todas as 2.088.220 palavras-chave únicas foram ranqueadas de acordo com a medida de importância de popularidade baseada na frequência por quinquênios entre 1962 e 2016, para facilitar o entendimento sobre a evolução temporal dos principais temas abordados. Inicialmente, na tentativa de mapear a evolução dos principais interesses científicos por parte dos doutores brasileiros ao longo do tempo, a Tabela 2 apresenta o coeficiente de similaridade das 15 palavras-chave mais populares entre cada par de quinquênios do período analisado.

Tabela 2: Coeficiente de similaridade das palavras-chave mais frequentes entre os quinquênios.

	1962-1966	1967-1971	1972-1976	1977-1981	1982-1986	1987-1991	1992-1996	1997-2001	2002-2006	2007-2011	2012-2016
1962-1966	X	46,66%	26,66%	20,00%	20,00%	20,00%	20,00%	13,33%	13,33%	0%	6,66%
1967-1971		X	33,33%	20,00%	26,66%	20,00%	20,00%	13,33%	13,33%	0%	6,66%
1972-1976			X	73,33%	66,66%	53,33%	53,33%	40,00%	20,00%	0%	6,66%
1977-1981				X	73,33%	66,66%	66,66%	53,33%	33,33%	13,33%	20,00%
1982-1986					X	86,66%	80,00%	66,66%	46,66%	20,00%	26,66%
1987-1991						X	93,33%	73,33%	53,33%	26,66%	33,33%
1992-1996							X	80,00%	60,00%	33,33%	40,00%
1997-2001								X	80,00%	46,66%	53,33%
2002-2006									X	66,66%	73,33%
2007-2011										X	86,66%
2012-2016											X

Como pode ser notado, não houve ao longo do período analisado 100% de similaridade dos tópicos centrais estudados entre pares de quinquênios. Isso mostra que a cada 5 anos os principais interesses científicos tem mudado por algum motivo. No entanto, ao comparar os pares de quinquênios sequentes, nota-se que em 100% das comparações realizadas revelam que parte dos principais tópicos de pesquisas estudados no quinquênio atual foram considerados relevantes no quinquênio anterior, destacando os períodos de 1987-1991 e 1992-1996, onde o percentual de similaridade foi de 93,33%, divergindo apenas quanto as palavras-chave “Peixes” e “Enfermagem”. Nesse caso, a palavra-chave “Peixes” que foi frequentemente utilizada nos períodos de 1982-1986 e 1987-1991, deu lugar a palavra-chave “Enfermagem” no quinquênio de 1992-1996. Uma hipótese para tal acontecimento, é o processo de regulamentação do exercício profissional que reconheceu as categorias de enfermeiro durante a década de 1980.

4 Considerações Finais

Este estudo contribuiu para a identificação dos tópicos de pesquisas que obtiveram maior interesses em cada época por parte dos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes. Ao realizar uma análise temporal entre as principais palavras-chave de cada período, foi possível verificar que as preferências centrais dos doutores sobre os tópicos de pesquisas foram alterando gradualmente ao longo do tempo, possivelmente devido a uma demanda externa da sociedade ou por determinado assunto ter atingido maturidade. Adicionalmente, ao comparar as principais palavras-chave pela medida de frequência com a medida de grau, foi possível constatar alta correlação considerando todo o conjunto de análise. Diante disso, constata-se que a utilização do grau como medida de importância se mostra uma alternativa interessante à frequência.

Agradecimentos

Os autores agradecem ao CNPq e CEFET-MG pelo auxílio na pesquisa.

Referencias

- BORGES, V. A. et al. Uma análise exploratória de tópicos de pesquisa emergentes em Informática na Educação. *Revista Brasileira de Informática na Educação*, Porto Alegre, v. 23, n. 1, p. 1-13, mar. 2015.
- BRITO, A. G. C. et al. Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação*, Campinas, v. 28, n.1, p. 77-86, jan./abr. 2016.
- DIAS, T. M. R. Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes. 2016. Tese (Doutorado em Modelagem Matemática e Computacional) - Programa de Pós-Graduação em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Minas Gerais, 2016.
- DIAS, T. M. R.; MOITA, G. F. A method for the identification of collaboration in large scientific databases. *Em Questão*, Porto Alegre, v. 21, n. 2, p. 140-161, maio/ago. 2015.
- DIGIAMPIETRI, L. A. Análise da rede social acadêmica brasileira. 2015. Tese (Livre Docência) Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015.
- FADIGAS, I. et al. Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática. *Educação Matemática Pesquisa*, v.11, n.1, p. 67-193, 2009.

- HERADIO, R. et al. A bibliometric analysis of 20 years of research on software product lines. *Information and Software Technology*, Amsterdam, v. 72, p. 1- 15, Apr. 2016.
- LANE, J. Let's make science metrics more scientific. *Nature*, London, v. 464, n. 7288, p. 488-489, Mar. 2010.
- MADLOCK-BROWN, C. R. A framework for emerging topic detection in biomedicine. 2014. Tese (Doutorado em Filosofia em Informática na Saúde) - Graduate College, University of Iowa, Iowa City, 2014
- PEREIRA, J. C. R. et al. Who's who and what's what in Brazilian Public Health Sciences. *Scientometrics*, Dordrecht, v. 73, n. 1, p. 37-52, Oct. 2007.
- PETER, R. S. et al. Epidemiologic research topics in Germany: a keyword network analysis of 2014 DGEpi conference presentations. *European Journal of Epidemiology*, London, v. 31, n. 6, p. 635- 638, June 2016.
- TRUCOLO, C. C. Análise de tendências em redes sociais acadêmicas. 2016. 65 f. Dissertação (Mestrado em Ciências) - Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2016.
- ZHANG, K. et al. A bibliometric analysis of research on carbon tax from 1989 to 2014. *Renewable and Sustainable Energy Reviews*, Amsterdam, v. 58, p. 297-310, May 2016.

A música na Ciência da Informação: uma análise de domínio da temática no Brasil no período de 1972 a 2018

Emanuella Maria Barbosa Lourenço Bezerra¹, Andréa Carla Melo Marinho², Francisco Arrais Nascimento³, Daniel Martínez-Ávila⁴ y Fábio Assis Pinho⁵

¹ ORCID [0000-0002-4886-7557](https://orcid.org/0000-0002-4886-7557). Comissão Memorial na Universidade de Pernambuco - UPE/Pronto-Socorro Cardiológico de Pernambuco Prof. Luiz Tavares - PROCAPE/Biblioteca Prof. Ovidio Montenegro - BPOM, Recife, Pernambuco, Brasil.
emanuella.bezerra@upe.br

² ORCID [0000-0001-6957-6911](https://orcid.org/0000-0001-6957-6911). Programa de Pós-graduação em Comunicação da Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil.
andreacmmarinho@gmail.com

³ ORCID [0000-0003-4424-8844](https://orcid.org/0000-0003-4424-8844). Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, São Paulo, Brasil.
francisco.arrais.nascimento@gmail.com

⁴ ORCID [0000-0003-2236-553X](https://orcid.org/0000-0003-2236-553X). Departamento de Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, São Paulo, Brasil.
martinez.avila@unesp.br

⁵ ORCID [0000-0003-1346-3808](https://orcid.org/0000-0003-1346-3808). Universidade Federal de Pernambuco, Recife, Pernambuco.
fabiopinho@ufpe.br

Resumo. Objetivou-se analisar a produção científica brasileira sobre música no âmbito da Ciência da Informação, identificando como o objeto “música”, vem auferindo espaço em pesquisas interdisciplinares, permitindo a visualização de tal objeto enquanto domínio na Ciência da Informação (CI). Estudo de natureza descritivo, arraigada sob análise de domínio, apresenta como corpus 124 produções acadêmicas (artigos de periódicos, artigos provenientes de anais de eventos, dissertações e teses), publicados no período de 1974 a 2018. Observou-se que a categoria de análise predominante é a que trata a música sob um viés documental, voltado a estudos de organização de acervos musicais no que tange aos processos de catalogação, classificação e indexação desses recursos informacionais. Porém, observa-se a emergência de outras perspectivas de análise em trabalhos que têm direcionado suas pesquisas sobre aspectos socioculturais acerca do objeto, sobretudo no que se refere à música enquanto recurso informacional de memória e representação social.

Palavras-chave: Música; Ciência da Informação; Análise de Domínio; Organização da Informação.

Abstract. The objective was to analyze the Brazilian scientific production on music in the field of Information Science, identifying as the object "music", has been gaining space in interdisciplinary research, allowing the visualization of such object as a domain in Information Science (CI). A descriptive study, rooted in domain analysis, presents as a corpus 124 academic productions (journal articles, articles from the annals of events, dissertations and theses), published from 1974 to 2018. It was observed that the predominant category of analysis is that dealing with music under a documental bias, focused on studies of organization of musical collections with respect to the processes of cataloguing, classification and indexation of these information resources. However, it is observed the emergence of other perspectives of analysis in works that have directed their research on sociocultural aspects about the object, especially with regard to music as an informational resource of memory and social representation.

Keywords: Music; Information Science; Domain Analysis; Information Organization.

1 Introdução

O estudo identifica como o objeto “música”, vem auferindo espaço em pesquisas interdisciplinares, permitindo a visualização de tal objeto enquanto domínio na Ciência da Informação (CI). Nesse sentido, a CI sob o viés sociocultural tem ampliado seu escopo de pesquisas, trabalhando a informação a partir dos domínios diferenciados. Assim, objetos não convencionais como os artefatos culturais, configuram-se como fonte de informação e consequentemente objeto de estudo na área. Barros, Café e Almeida (2013), postulam que a “materialidade sonora da música”, não pode ser trabalhada apenas sob a perspectiva de sua sonoridade, mas sim, atentar para o seu conteúdo. Logo, a música fora eleita nesta pesquisa por sua universalidade e pela afinidade dos autores com o objeto e sob a égide da CI, dada sua natureza interdisciplinar, bem como por interseccionar arte e ciência.

Para compreender a atuação de tal objeto na CI, não ignorando sua relação e sentidos social e histórico, alocando a canção enquanto recurso informacional, indagou-se: Quais os trabalhos em CI que contemplam a música como objeto de pesquisa no Brasil? Quais as perspectivas de análise destas pesquisas? Quais instituições produziram tais trabalhos? Quais os meios de comunicação científica que os trabalhos foram divulgados?

Para tal empreendimento, foram mapeadas as pesquisas desenvolvidas nos programas de pós-graduação em Ciência da Informação, Gestão da Informação, Biblioteconomia, Comunicação e Informação e Memória Social, com dados atualizados do relatório de área de 2017 da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), além dos artigos de periódicos e de anais de

eventos publicados nas principais bases de dados da área que contemplassem a música enquanto objeto de pesquisa em Ciência da Informação no Brasil.

Assim, o trabalho tem por objetivo analisar a produção científica da CI no Brasil sob a perspectiva temática da música enquanto objeto de estudo da área.

2 Música e Ciência da Informação

A música, como parte da cultura ocidental está intimamente ligada à cultura medieval, visto que muito das memórias deste período foram pautadas por meio da tradição oral e com formas de consumo e produção fechados, circunscrita a pequenos grupos ou aldeias (IAZZETTA, 2001). Com o surgimento da imprensa, houve a possibilidade de novas formas de difusão do conhecimento e da informação, permitindo materializar a informação sonora da música, bem como sua disseminação, mas é só por meio do fonógrafo que foi possível alavancar sua difusão. Assim, o registro informacional da música (partitura e letra da canção) possibilitou seu registro como fonte documental de memória.

Le Coadic (2004, p. 5) afirma que a informação “[...] é um conhecimento inscrito” em qualquer suporte. Otlet (2018) faz alusão sobre a “documentação sonora”, onde a música e outros documentos sonoros estão aí contemplados, visto que para o autor tanto do ponto de vista da música quanto das letras, “o disco” é um tipo de documentação. Briet (2016), lança mão de um exemplo clássico quando pergunta se um “antílope é um documento?” e responde que não, mas tudo o que for produzido a partir dele é um documento. Da mesma forma, o som produzido por um instrumento não é um documento, mas a sua composição, as letras produzidas, os discos, os *Compact Disc* (CDs), entre outros são documentos.

Esse raciocínio evidencia não apenas a vertente informacional da música, mas as possibilidades de estudos utilizando-se deste objeto. Morigi e Bonotto (2004) apontam para o tratamento da música nos campos da Arquivologia, Biblioteconomia e Museologia, que tais objetos são tratados de forma a respeitar as várias facetas que se apresentam e não apenas às formas tradicionais no formato em papel. Nesse sentido, Toller e Bizello (2017, p. 272) colocam que o “[...] documento tem seu valor definido aos olhos do pesquisador” visto que tal registro é importante, independentemente da forma como ele se apresenta.

Saracevic (1996), evidencia que a CI apresenta três características que constituem a razão de sua existência e evolução, ou seja, sua natureza interdisciplinar, sua relação com a tecnologia da informação e sua contribuição na evolução da sociedade da informação, uma vez que apresenta uma forte dimensão social e humana.

Nesse sentido, é possível considerar que os estudos acerca de artefatos artístico-culturais enquanto objeto de pesquisa em CI apresentam as características mencionadas anteriormente, uma vez que há uma forte relação com outros campos de conhecimento (artes, ciências sociais, cultura e linguística), bem como essa produção

é armazenada em unidades de informação de instituições acadêmicas e culturais, ou mesmo em ambientes digitais.

Para além de tal compreensão, Von Simson (2006, p. 1), associa a música às questões memoriais, tratando-a não apenas como artefato informacional e cultural, mas sim, um documento imagético (impregnada de imagens, lembranças, memórias), além de meio de representações sociais, evidenciando a capacidade do ser humano de guardar fatos e experiências, bem como a necessidade não só da guarda, mas da difusão destes fatos para a posteridade, utilizando-se para isso, além da história oral, variados suportes de informação, tais como mecanismos de voz, imagem e texto. Em consonância com o autor, Morigi & Bonotto (2004, p. 148) falam sobre a narrativa musical que objetiva “expressar os sentimentos coletivos” e que tais documentos estão impregnados de sentido, que vai para a além da racionalidade, mas carrega em si muito do que retém os sentimentos.

Susana Sardo (2018, pp.9-10) afirma que “[...] o primeiro desafio que se nos depara quanto tentamos escrever ou falar sobre música é o de não saber como nomeá-la. [...] Por tanto, o exercício de objetificação, de fixação e de classificação da música é complexo e começa, desde logo, no momento de sua nomeação”. Estas reflexões tem relação com a tradição crítica da CI e a Organização do Conhecimento e as consequências éticas do “poder de nomear” (OLSON, 2002).

Percebe-se que a temática favorece uma discussão acerca do campo científico da CI, uma vez que possibilita uma diferente perspectiva de análise do seu objeto de estudo. De acordo com Santos & Kobashi (2007, p.106), quando falam daquilo que permeia e abaliza o campo da CI, visto que seus objetos de estudo estão ligados aos fatores “[...] epistemológicos, teóricos e metodológicos das disciplinas científicas”, tendo entre seus pares o entendimento necessário sobre os princípios, autenticidade, adequação das discussões propostas e de sua razoabilidade e pertinência das metodologias, procedimentos e ferramentas próprias para o desenvolvimento de pesquisas.

3 Aspectos Metodológicos

A pesquisa de natureza descritiva, arraigada sob análise de domínio, ancorando-se nas contribuições de Tennis (2002, 2003, 2012), Hjørland (1997, 2002, 2003, 2017) e Hjørland & Albrechtsen (1995). Apresenta como corpus 124 produções acadêmicas (artigos de periódicos, artigos provenientes de anais de eventos, dissertações e teses), que utilizam a música como objeto de pesquisa, publicados no período de 1974 a 2018, identificados nas seguintes fontes: Base de Dados Referencial de Artigos e Repositórios em Ciência da Informação (BRAPCI), Base de dados dos Anais dos Enancib (BENANCIB), Anais do Capítulo Brasil da International Society for Knowledge Organization (ISKO), Base de dados Rede PERI, da UFMG e na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) do IBICT, Banco de

Teses da CAPES e Repositórios Institucionais aos quais estão vinculados os 30 Programas de Pós-Graduação em Ciência da Informação no Brasil.

Os termos de busca utilizados foram “Música” e “Musical”, e após recuperação, os trabalhos foram organizados de acordo com uma categorização do conteúdo referente à temática da produção analisada. Assim, foram definidas 4 categorias de análise de acordo com a Figura 1.



Figura 1. Categorias de Análise

4 Análise e Discussão dos Resultados

Na BRAPCI foram selecionados 37 artigos produzidos no período de 1974 a 2018. Na base PERI foram recuperados 29 registros, publicados no período de 1980 a 2016. Na BENANCIB foram encontrados 33 artigos, publicados entre 1995 e 2018. Vale ressaltar que houve duplicidade na indexação de trabalhos nas bases PERI, BRAPCI e BENANCIB. Assim, foram encontrados 2 trabalhos iguais na PERI e BENANCIB e 11 entre a PERI e BRAPCI. Quanto aos eventos, nos Anais da ISKO Brasil foram encontrados 4 trabalhos nas edições do evento que foram realizadas nos anos 2013-2017.

A figura 2 demonstra a concentração e produtividade temática com o objeto música por meio da coleta dos artigos nas bases em CI no Brasil. Os resultados apontam que a pesquisa acerca do objeto tem crescido nos últimos anos, tendo a área trabalhado o objeto na maior parte dos trabalhos sob um viés documental, porém outras perspectivas de estudo com o objeto, a exemplo da música enquanto fonte de memória e representação social, o acesso e preservação de recursos audiovisuais em ambientes digitais também podem ser visualizadas.

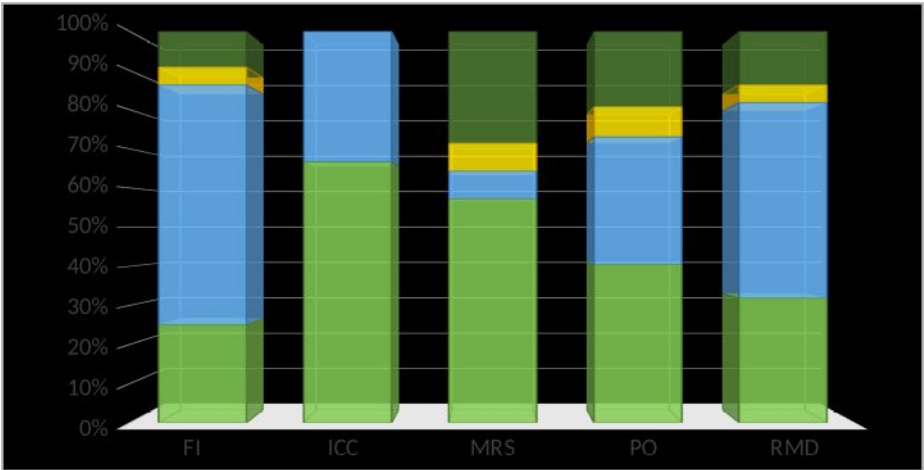


Figura 2. Produção na Área de Ciência da Informação segundo a Fonte e a Classificação Estabelecida pelos autores

Por fim, foi analisada a produção de dissertações e teses no Brasil, onde buscou-se na Banco de Teses da CAPES, BDTD, e nos repositórios das instituições (UNB, UFBA, UFPB, UFSC e USP) para fins de comprovação, totalizando 24 trabalhos (16 dissertações e 8 teses) (Figura 3). Sobre a temática da música em 8 programas de pós-graduação como é possível observar na Figura 4.

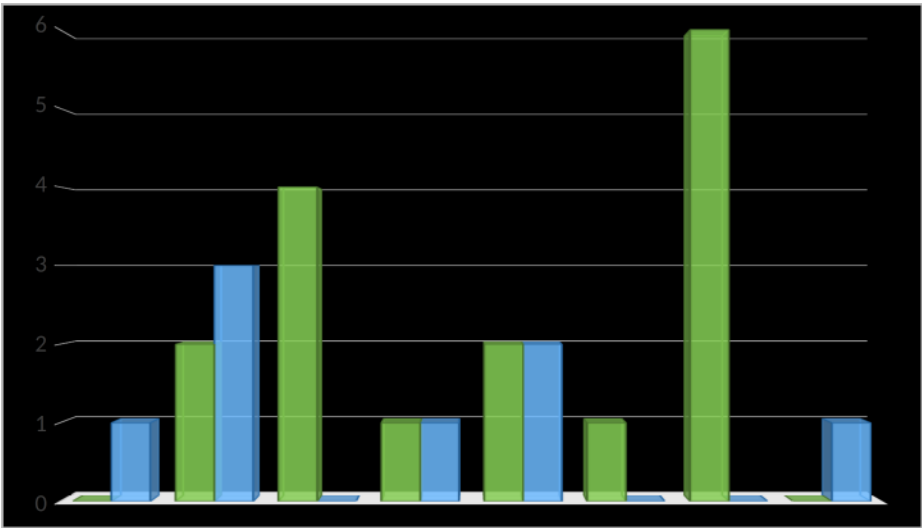


Figura 3. Produção científica segundo o nível de graduação

A figura 3 demonstra a distribuição de trabalhos acadêmicos produzidos nos 8 PPG's em CI no Brasil segundo a classificação estabelecida. Observou-se uma concentração de pesquisas desenvolvidas predominantemente no programa de Memória Social da UNIRIO com 6 dissertações. Contribuindo para o aumento da produtividade no domínio seguem o PPGCI-UFGM com 2 teses e 3 dissertações, o PPGCI-UNB com 2 teses e 2 dissertações, o PPGCI-UFPE com 4 dissertações e o PPGCI-UFSC com 1 tese e 1 dissertação. Os programas da UFBA, UNESP-Marília e USP produziram 1 dissertação.

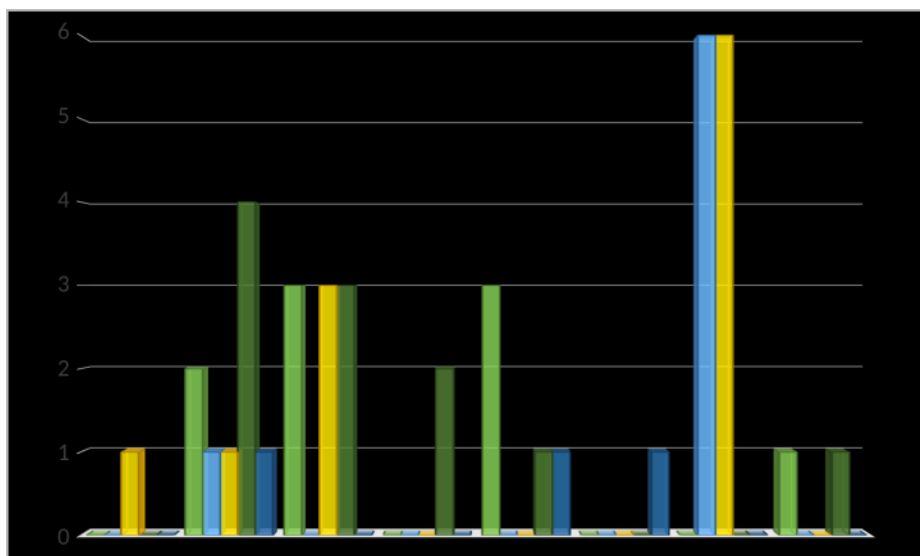


Figura 4. Produção científica segundo o PPG e a Classificação estabelecida

A Figura 4 mostra que o objeto é discutido proporcionalmente nas diferentes categorias de análise, o que difere dos dados referentes às bases em CI e aos anais do ENANCIB. Percebe-se que os estudos acerca da música, que se configura enquanto manifestação artístico-cultural são crescentes na CI e possibilitam uma discussão teórico-metodológica no âmbito da OI e estudos de memória, sobretudo a partir de produções do programa em Memória Social da Unirio, que concentra a maior produção de pesquisas sobre a temática, seguido dos PPGs da UFGM, UnB , UFPE e UFSC.

5 Considerações finais

Observou-se que a categoria de análise predominante é a que trata a música sob um viés documental, voltado a estudos de organização de acervos musicais no que tange aos processos de catalogação, classificação e indexação desses recursos informacionais. Porém, observa-se a emergência de outras perspectivas de análise em trabalhos que têm direcionado suas pesquisas sobre aspectos socioculturais acerca do objeto, sobretudo no que se refere à música enquanto recurso informacional de memória e representação social.

Referencias

- Barros, C. M., Café, L. M. A., & Almeida, C. C. de (2013, outubro). Informação musical e interpretação: contribuições semióticas para o campo da organização da informação. *Anais [do] Encontro Nacional de Pesquisa em Ciência da Informação*, Salvador, Bahia, 14. Recuperado de: <http://200.20.0.78/repositorios/bitstream/handle/123456789/2335/INFORMA%C3%87%C3%83O%20MUSICAL.pdf?sequence=1>.
- Briet, S. (2016). *O que é Documentação?* Brasília: Briquet de Lemos. Recuperado de: http://file:///C:/Users/AcerOficial/Downloads/Briet_Suzanne_o_que_eh-a_documento%C3%A7%C3%A3o.pdf.pdf
- García Gutiérrez, A. (2018). *Em Pedazos: el sentido de la desclasificación*. Madrid: ACCI.
- Hjørland, B. (1997). *Information seeking and subject representation: an activity theoretical approach to Information Science*. London: Greenwood Press.
- Hjørland, B. (2002). Domain analysis in information science. Eleven approaches — traditional as well as innovative. *Journal of Documentation*, 58, 4, 422-462.
- Hjørland, B. (2002). Domain analysis in information science. Eleven approaches — traditional as well as innovative. *Journal of Documentation*, 58, 4, 422-462.
- Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30, 2, 87-111.
- Hjørland, B. (2017). Domain analysis. *Knowledge Organization*, 44, 6, 436-464.
- Hjørland, B. & Hanne, A. (1995). Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science*, 46, 6, 400-425.
- Iazzetta, F. (2001, outubro). Reflexões sobre música e o meio. *Anais [da] Associação Nacional de Pesquisa e Pós-Graduação em Música*, Belo Horizonte, Minas Gerais, UFMG, 13. Recuperado de: http://www.anppom.com.br/anais/anppom_2001_1.pdf.

- Le Coadic, Y. (2004). *A ciência da informação*. (2a ed.). Brasília: Briquet de Lemos.
- Morigi, V., & Bonotto, M. (2004). A Narrativa Musical, Memória e Fonte de Informação Afetiva. *Em Questão*, 10 (1), 143-161
- Olson, H. A. (2002). *The power to name: locating the limits or subject representation in libraries*. Dordrecht: Kluwer Academic Publisher.
- Otlet, P. (2018). *Tratado de documentação: o livro sobre o livro: teoria e prática*. Brasília: Briquet de Lemos. Recuperado de: http://www.cfb.org.br/wp-content/uploads/2018/09/otlet_tratado_de_documenta%C3%A7%C3%A3o.pronto.pdf
- Santos, R. & Kobashi, N. (2007). Análise de teses e dissertações de Ciência da Informação: estudo de institucionalização de um campo científico. *Anais, 8 Encontro Nacional de Pesquisa em Ciência da Informação*. Salvador, Bahia, Universidade Federal da Bahia. Recuperado a partir de: <http://www.enancib.ppgci.ufba.br/artigos/GT7--109.pdf>.
- Saracevic, T. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, 1 (1), 41-62.
- Sardo, S. (2018). Presentación. O desconforto de escrever sobre música. In García Gutiérrez, A. (2018). *En Pedazos: el sentido de la desclasificación*. Madrid: ACCI.
- Tennis, J. (2012). A convenient verisimilitude or oppressive internalization? Characterizing the ethical arguments surrounding hierarchical structures in knowledge organization systems. *Knowledge Organization*, 39 (5), 394-397.
- Tennis, J. T. (2002). Subject ontogeny: Subject access through time and the dimensionality of classification. In: López-Huertas, M. J. (Ed.). *Challenges in knowledge representation and organization for the 21st Century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference, Granada (Spain), 2002*. Würzburg: Ergon. 8, 54-59.
- Tennis, J. T. (2003). Two axes of domains for domain analysis. *Knowledge Organization*, 30, nos. 3-4: 191-195.

Enriquecimiento de entidades de Wikidata mediante un modelo de descomposición y mapeado de categorías de Wikipedia

Tomás Saorín¹ y Juan-Antonio Pastor-Sánchez²

¹ ORCID [0000-0001-9448-0866](https://orcid.org/0000-0001-9448-0866). Departamento de Información y Documentación, Universidad de Murcia, España.
tsp@um.es

² ORCID [0000-0002-1677-1059](https://orcid.org/0000-0002-1677-1059). Departamento de Información y Documentación, Universidad de Murcia, España.
pastor@um.es

Resumen. El objetivo de este trabajo es explorar la relación entre las categorías asignadas a los artículos de Wikipedia con la descripción y metadatos generados en Wikidata. Se plantea utilizar la categorización de artículos de Wikipedia para enriquecer la descripción de entidades en Wikidata. Para ello se propone procesar los literales de las categorías mediante técnicas de procesamiento de lenguaje natural (PLN) estableciendo patrones que permitan identificar tanto propiedades como entidades o valores con los que construir declaraciones para una entidad. La secuencia de operaciones propuesta sería el siguiente: 1) Selección de un conjunto coherente de categorías, 2) Establecimiento de patrones de procesamiento de literales y asignación a propiedades y elementos de Wikidata, 3) Creación de declaraciones con cualificadores para cada categoría procesada y 4) Programación de bots para el procesamiento automático de categorías, enriquecimiento y validación de las descripciones de elementos de Wikidata. La propuesta recogida en este trabajo se centra en el uso de diferentes propiedades y entidades de Wikidata para desarrollar el punto 3. La automatización de un proceso para enriquecer y validar las declaraciones de cada elemento, permitiría aprovechar las dinámicas de edición existentes. Además ayudaría a elaborar un esquema de conceptos de más calidad, al especificarse el significado de las categorías que suponen una composición de varios términos y que en realidad resuelven necesidades descriptivas por otros medios.

Palabras clave: Wikipedia, Wikidata, Categories, Named-entity recognition, Knowledge organization

Abstract. This paper explores the relationship between the categories assigned to the Wikipedia articles with the description and metadata generated in Wikidata. It is proposed to use the categorization of Wikipedia articles to enrich

the description of entities in Wikidata. For this, the literals of the categories could be processed using natural language processing techniques (NLP), establishing patterns that allow the identification of properties as well as entities or values with which to construct statements for an entity. The sequence of operations would be the following: 1) Selection of a coherent set of categories, 2) Establishment of patterns of processing of literals and assignment to properties and elements of Wikidata, 3) Creation of declarations with qualifiers for each category processed and 4) Programming of bots for the automatic processing of categories, enrichment and validation of Wikidata element descriptions. The technique shown in this paper focuses on the use of different properties and entities of Wikidata to develop point 3. The automation of a process to enrich and validate the declarations of each element, would allow to reuse existing edition dynamics. It would also help to elaborate a scheme of concepts of higher quality, by specifying the meaning of the categories that suppose a composition of several terms and that actually solve descriptive needs by other means.

Keywords: Wikipedia, Wikidata, Categorías, Reconocimiento de entidades nombradas, Organización del Conocimiento

1 Introducción

Wikidata, el grafo de conocimiento vinculado a los proyectos Wikimedia, está suscitando un intenso interés en las comunidades sectoriales y en las investigaciones académicas, al igual que sucediera con Wikipedia (Saorín; Pastor-Sánchez, 2018). Uno de los aspectos que interesan a la comunidad GLAM, es el de la interconexión de vocabularios controlados, en donde Wikidata actuaría como un “hub of hubs” para los sistemas de control de autoridades y de terminología (Association of Research Libraries, 2019).

Una de las primeras fases de Wikidata fue la de integrar en un único elemento todos los enlaces a las diferentes versiones en cada idioma de los artículos. De este modo se construye un único elemento, con un identificador único y una URI, que representa el concepto, objeto o hecho, independientemente del idioma: cada artículo de Wikipedia se vincula con un concepto que actúa como enlace interlingüístico. La “Guerra de los 30 años” se concreta en el Q2487, accesible desde la URI <https://www.wikidata.org/wiki/Q2487>, y no ya como 113 artículos en diferentes Wikipedias.

Complementariamente se realizaron procesos automatizados (bots) para extraer del texto de los artículos declaraciones válidas, de forma análoga al procesado de Dbpedia, aprovechando sobre todo la información semiestructurada en infoboxes. Cada entidad es descrita como una instancia de algo, y se le añaden datos factuales como “fecha de nacimiento”, “coordenadas”, “autor”, “premios recibidos”, etc. según la naturaleza del contenido. Estas declaraciones se realizan a través de propiedades (P) que constituyen el modelo de descripción que emerge de Wikidata.

En este trabajo queremos explorar la relación entre las categorías asignadas a los artículos de Wikipedia con la descripción y metadatos generados en Wikidata. Paralelamente a los artículos – que son el contenido principal – existe en la enciclopedia un amplio entramado de categorías, con una función secundaria de navegación y agrupación de contenidos. Las categorías suponen un tipo especial de contenido dentro de Wikipedia: son meramente términos o conceptos, y son incorporadas a Wikidata, con ciertas peculiaridades que abordaremos en este trabajo. Sobre las categorías se están desarrollando líneas de investigación, con especial énfasis en su uso para mejorar para el procesamiento de corpus documentales externos (Tramullas, 2018).

Las categorías de Wikipedia representan un esfuerzo singular en la estructuración del conocimiento histórico, cultural, científico y factual, realizado de forma colaborativa y masiva. La comunidad de editores, al mismo tiempo que elabora contenidos (artículos), trabaja en su organización conforme a un esquema de conceptos (categorías), el cual evoluciona orgánicamente junto al contenido. Es decir, no existe una planificación previa sobre la estructura de categorías, sino que se crea conforme se elaboran nuevos artículos, y están a cargo de lo que se denomina “accidental taxonomist” (Heder, 2016); personas no expertas en organización del conocimiento y cuya dedicación a la construcción de la taxonomía es parcial e irregular.

La funcionalidad de categorías en el software MediaWiki de gestión de contenidos se incorporó a en 2004 a la Wikipedia en inglés, a los tres años de rodaje del proyecto. Proporciona un índice automático de contenidos y adopta una forma esencialmente pseudojeárquica, permitiendo establecer una forma laxa de taxonomía o, más frecuentemente, un grafo. La cantidad de categorías usadas en Wikipedia es excepcional: en mayo-junio de 2018, en Wikipedia en español (eswiki) hay 403.093 páginas de categorías para 1.421.034 de artículos, y en inglés son 1.711.155 para 5.669.291 artículos. A diferencia de otros sistemas de clasificación usados en colecciones documentales (bases de datos bibliográficas, por ejemplo), donde existe un conjunto de conceptos restringido que se usan para indizar documentos primarios, y que tiende a mantenerse estable, y que crecen lentamente conforme se detectan nuevos temas a los que prestar atención, en Wikipedia las categorías están en continuo crecimiento: los artículos (el contenido) y las categorías (el etiquetado) crece al mismo tiempo: hay una ratio de unas 0,58 categorías por artículo, con variaciones según idioma. Esta circunstancia se debe en buena parte a que los contenidos de Wikipedia son, hasta cierto punto, un propio esquema de conocimiento, al ser una enciclopedia. En una base de datos puede haber miles de artículos sobre el hidrógeno, pero en la enciclopedia hay uno solo.

Las categorías de Wikipedia no son en realidad un esquema de clasificación o un vocabulario controlado para describir el contenido, sino para facilitar la navegación entre elementos relacionados. Si las entendemos como una forma de taxonomía, hemos de tener en cuenta que en el medio digital, el uso de las taxonomías para organizar conocimiento supera la limitada tarea de indexar colecciones documentales,

sirviendo para hacer funcionar bancos de datos de imágenes, organizar la navegación web, potenciar el descubrimiento de contenidos a través de la búsqueda facetada en tiendas en línea, definir flujos de trabajo en un sistema de información corporativo o para describir las relaciones e interacciones entre usuarios en los medios sociales (Lambe, 2007).

Por otra parte, cuando se asignan categorías a una página de categoría, se va construyendo sobre la marcha una jerarquía de conceptos. Las páginas wiki de categorías nos informan tanto de los artículos o páginas categorizadas (P) con ellas, como de sus subcategorías (C).



Figura 1. Ejemplo de página de la categoría “Ríos de Eslovaquia” de la edición en Español de Wikipedia donde pueden observarse tanto las subcategorías como las páginas clasificadas bajo dicha categoría. Fuente: https://es.wikipedia.org/wiki/Categoría:Ríos_de_Eslovaquia

El sistema de categorías de Wikipedia puede ser entendido como un "organic thesaurus", conforme a la matriz de "infraestructuras de conocimiento" de Lambe (Lambe, 2007, p. 245). Desde el punto de vista de los lenguajes documentales, tomando como referencia la norma UNE-ISO 25964 sobre Tesauros, resulta difícil aceptar que sea estrictamente un vocabulario controlado, puesto que los usuarios pueden crear sobre la marcha las categorías que necesiten (como en una folksonomía). En este sentido las categorías de Wikipedia podrían considerarse un esquema de clasificación enumerativo, pero no finito. Sin embargo, el etiquetado no es libre y la asignación de categorías preexistentes parece predominar frente a la creación de nuevas, dándose numerosas actuaciones de reorganización y mantenimiento del sistema de categorías. No obstante, es un vocabulario estructurado y puede ser usado para la recuperación de información. Su estructura es claramente polijerárquica. Por otro lado, se trata de un vocabulario fuertemente precoordinado.

El sistema de categorías de Wikipedia no puede considerarse un tesauro, puesto que el criterio para la partición jerárquica no mantiene el rigor en todos los casos. Su forma impura se aproxima más a lo que la norma UNE-ISO considera como

taxonomía polijerárquica, que resuelven una necesidad principal de facilitar la navegación en un sitio web, son realizadas por no expertos y presentan asimetrías en su multilingüismo. Adaptando la propuesta anteriormente citada de Lambe, podríamos acuñar el término “Organic taxonomy”.

Las categorías son registradas, al igual que los artículos, como entidades en Wikidata, asignándoles una tipología específica que las distingue del resto de elementos¹. Sin embargo, no se recogen en Wikidata ni la indización ni las relaciones jerárquicas². La misma categoría en diferentes idiomas se unifica en un único elemento Wikidata, dando como resultado (a enero de 2019) 4.169.547 categorías diferentes en Wikidata, de las cuales sobre un 90% correspondían a las cargadas desde las más de 250 ediciones diferentes de Wikipedia, y el resto a los otros proyectos Wikimedia (Wikisource, Wikivíajes, Wikiquote, Wikinoticias, etc.)³. Nuestro foco de atención son las categorías de Wikipedia, que es de donde procede el grueso de las categorías registradas en Wikidata.

Para la incorporación de las categorías a Wikidata, la comunidad del proyecto tomó la decisión de no incluir la indización, es decir, no reflejar que un artículo ha sido categorizado con determinadas categorías. Tampoco se recogen las relaciones jerárquicas entre categorías. Por el contrario, en el caso de DBpedia, sí que se optó por formalizar la relación artículos-categorías y categorías-categorías, usando, para las relaciones jerárquicas, propiedades SKOS.

Una de las razones que justifican esta decisión es que la comunidad no las considera conocimiento factual, sino operativo para construir el sitio web de Wikipedia, y que la mejor forma de caracterizar los aspectos factuales de un artículo de Wikipedia son las propiedades específicas, tales como autor, fecha de publicación, ubicación, premios recibidos, tipo/instancia, etc. Por este motivo, cualquier dato que se incluya dentro de la etiqueta de una categoría, y que también esté contemplado en una declaración a través de una propiedad, generaría una redundancia innecesaria.

Se advierte, en los debates sobre categorías en Wikipedia y Wikidata, una palpable insatisfacción con ellas, que apuntaría a una posible sustitución futura por mecanismos mejor adaptados a la función que pretenden realizar: listado y agrupación⁴.

1 Se trata de instancias (P31) de Q4167836 (Wikimedia category) o de alguna de sus 20 subclases.

2 En anteriores trabajos hemos abordado la skosificación de la integración en Wikidata de las categorías (Pastor-Sánchez y Saorín, 2018).

3 Consulta SPARQL para categorías:

```
SELECT (COUNT(?item) AS ?total) WHERE {?item wdt:P31 wd:Q416783 }
```

Incluyendo las categorías tipificadas en alguna de las subclases de “Wikimedia category”, la consulta sería:

```
SELECT (COUNT(distinct ?item) AS ?total) WHERE {
  ?item wdt:P31 ?tipo. ?tipo wdt:P279* wd:Q416783 }
```

4 Los puntos de vista y las citas que extraemos proceden de dos debates sobre las propuestas de “Parent category” y “Category”, archivados en:

https://www.wikidata.org/wiki/Wikidata:Property_proposal/Archive/30#Parent_category

Es posible reproducir gran parte de las funciones de listado y agrupación de artículos que se realiza ahora mismo través de categorías, mediante consultas dinámicas sencillas usando “regular properties”. Esto es denominado en los debates como “category assignments” y debe entenderse en términos de “tagging”. También se argumenta que un statement de indización de un artículo no puede ser apoyado en una fuente externa, y por lo tanto no es conocimiento factual independiente, sino resultado de la actividad de la comunidad Wikimedia. También se señala el fuerte incremento de los statements para cada ítem, si se incorporaran sus categorías.

“If the categorisation can be expressed via a property:value statement then we should add that statement, not this vague category statement. If the categorisation can not be expressed as a statement then we almost certainly don't need it. The categorisation systems on english and other wikipedias certainly does contain a lot of information which can be harvested and converted into useful statements and we should certainly do this conversion even if it means we need to create more properties. For instance we need a "preparation method" property so we can include the statement "Chipotle:preparation method:smoked". The **raw data** however should stay in the wikipedias.”

Advertimos que las categorías de la Wikipeda no son aceptadas como algo “factual” (un objeto de la realidad, que deba ser objetivado) sino como una creación adhoc dentro de la Wikipedia. No es aceptado como una clasificación de conocimiento, sino como una estructura de apoyo:

“WP categories are a kind of classification without global overview. So instead of using an **empirical classification**, just start the work correctly.”

“**Category is no knowledge** but an arbitrary way to sort items that depends on the wikimedia project. It would not be possible to source that kind of statement and it would generate a lot of repetitions with the other statements.”

“WP categories are outdated: categories are a specific way to group articles which **is not common knowledge but particular point of view**.”

“Wikipedia use categories for classification. Wikidata use **something else**”

En cuanto a las relaciones jerárquicas, dado que en cada idioma están establecidas unas relaciones diferentes, y que el contenido declarado en los ítems (tipo, autor, , lugar, estilo, etc.) ya supone una forma de organización de conocimiento, no se refleja la categorización de categorías, que, en un sentido amplio, se corresponderían con la relación Broader Term de un Tesauro.

“The category system is pointless because there is no unique system, but different ways to classify articles.”

“Transferring Wikipedia's categories' content as it is is not an appropriate to exploit the data. This would generate repetitions and errors and force users to develop a second ontology to define the relation between categories.”

“Not to mention the fact that this is a central point and that there is as many category hierarchy as there is Wikipedia, this would be a mess to centralize everything here”

Además, en ambos casos, se plantea como barrera la ineficiencia de “replicar” o “mimic” estos datos, que son decisiones editoriales de usuarios, en Wikidata.

“Why do we need to reflect categories on Wikipedias? They will continue to be there and to be maintained, queryable and extractable. Why do we need to mimic them here, **would'nt they be messy and double** the work of maintaining them, a change of some wikipedia would have to be reflected here.”

Este es la situación actual implementada en Wikidata, pero creemos interesante recoger las razones de las posturas que opinan en sentido contrario: Los esquemas de categorías no son accidentes o transacciones automáticas en las wiki, sino el resultado de un esfuerzo constante y consciente por construir una herramienta de organización de ciertos aspectos de conocimiento que se consideran necesarios.

“Categories are most **assuredly knowledge. A bit messy, lacking in organization (true) but very comprehensive**. Wikipedians are very serious about their categorization [...] We also don't source Labels and Descriptions, but do you think you could live without them?”

La explotación de las categorías para generar nuevas declaraciones y nuevas propiedades puede ser una vía de futuro, pero elude la situación actual de Wikipedia en la que las categorías cumplen una función:

“It could/might be transferred into properly structured properties in Wikidata...maybe in 10 years time. But I'm not talking what we could have: I'm talking about what is already in Wikipedia but not reflected in Wikidata.”

También la variabilidad entre idiomas se enfoca como un valor y no como una limitación: Al igual que versión de artículo de un idioma diferente refleja una visión específica de una comunidad, las categorías asignadas y las relaciones entre categorías manifiestan las diversidades que sobre el conocimiento existen entre grupos humanos, que es uno de los puntos fuertes de Wikipedia, compatible con la política del punto de vista neutral.

“We'll merge the category trees of 11 languages: that's mostly a union, since category interlanguage links are only 0.35x compared to 2.2x for articles. I

don't view this as "discrepancies" but as useful local views that complement each other.”

En el momento actual hay varios aspectos operativos de las categorías de Wikipedia que se ha optado por no integrar en Wikidata, por lo que podemos afirmar que están infraformalizadas: parte del esfuerzo editor no fluye hacia la base de conocimientos central. Parece evidente que naturaleza de la gestión de contenidos en Wikipedia, una vez que se ha abierto el camino a la conexión con una “central de datos”, implicará abordar más tarde o más temprano que la asignación de categorías a los artículos se realice de forma integrada con Wikidata (tagging), así como la asignación de categorías a las categorías (concept schema). En la actualidad está integración podría simularse mediante el procedimiento habitual de bots que sincronizan datos de forma continua. La fuerza bruta de los bots puede suplir a la lógica de diseño de un sistema robusto de gestión de contenidos multilíngues y multiversión.

Sin embargo, nuestro trabajo parte del supuesto de que, dado que las categorías forman parte del conocimiento codificado presente ya en los artículos de Wikipedia, suponen otra vía complementaria para la extracción de conocimiento valioso. Los editores las están usando para expresar ciertas relaciones o necesidades de organización del conjunto de artículos, y por lo tanto podemos explotar su significado en relación con un artículo.

2 Metodología

El trabajo que planteamos busca aprovechar la categorización de artículos de Wikipedia para enriquecer la descripción de entidades en Wikidata. A partir del uso que se está haciendo de las categorías (listados y navegación) y por su propia naturaleza precoordinada, parece entreverse la posibilidad de extraer de ellas elementos descriptivos que pueden incorporarse como declaraciones. Es sencillo de entender con un ejemplo: si el artículo sobre Albert Einstein usa las categorías “Científicos judíos” y “Ganadores del premio nobel de física en 1921”, podemos inferir y formalizar como hechos factuales que Einstein era un científico, que era judío y que ganó el premio nobel de física en un determinado año. Conocimiento de este tipo se manifiesta en las categorías usadas en los artículos, y por lo tanto es aprovechable para el enriquecimiento y validación de declaraciones en Wikidata.

Tabla 1. Comparativa de contenido entre el artículo sobre Albert Einteins en la edición en Español de Wikidata y la entidad correspondiente en Wikidata.

Categorías asignadas al artículo sobre Albert Einstein en la edición en Español de Wikipedia https://es.wikipedia.org/wiki/Albert_Einstein	Declaraciones sobre la entidad sobre Albert Einstein en Wikidata https://www.wikidata.org/wiki/Q937
<ul style="list-style-type: none"> • Hombres, Nacidos en 1879 • Fallecidos en 1955 • Albert Einstein • Físicos teóricos • Físicos relativistas • Físicos cuánticos • Cosmólogos • Científicos exiliados del nazismo • Nacionalizados de Suiza • Nacionalizados de Estados Unidos • Alumnado de la Escuela Politécnica Federal de Zúrich, Profesores de la Universidad Carolina • Profesores de la Universidad Humboldt de Berlín • etc. 	<ul style="list-style-type: none"> • Instancia de (P31) Humano • Sexo (P21) Masculino • Nacionalidad (P27) Imperio Alemán; Suiza; República de Weimar; USA; Austro-Hungría • Fecha de nacimiento (P569) 14-3-1879 • Ocupación (P106) Físico teórico; Escritor científico; Inventor; Físico; Profesor universitario... • Campo de trabajo (P101) Física teórica • Miembro de (P463) Roya Society; merican Philosophical Society, ... • Premios recibidos (P166) Gold Medal of the Royal Astronomical Society (1926) ... • etc.

El modelo de representación de elementos en Wikidata se basa en declaraciones (pares de propiedades y vínculos a otras entidades de Wikidata). Al igual que el modelo RDF permite incluir reificaciones. De esta forma, cada declaración puede ser matizada por otra nueva declaración (cualificadores), apoyada en una fuente externa (referencia) y ponderada (Rank). Esta forma de expandirse o anotarse o contextualizarse permite diferenciar entre el Snak (Propiedad-Valor) y el Claim (Propiedad-Valor-Cualificador).

Tabla 2. Detalle sobre una declaración de la entidad Q937 (Albert Einstein) de Wikidata.
Fuente: <https://www.wikidata.org/wiki/Q937>

Item sujeto	Propiedad	Item objeto	Propiedad cualificador	Dato
Albert Eintein (Q937)	Premios recibidos (P166)	Gold Medal of the Royal Astronomical Society (Q753072)	Punto en el tiempo (P585)	1926

Para un determinado subconjunto de categorías en un idioma determinado, podemos procesar los literales mediante técnicas de procesamiento de lenguaje natural (PLN). Para ello, se establecerían patrones que permitan identificar propiedades, entidades o valores con los que construir declaraciones para una entidad. Muchas ramas de categorías presentan patrones reconocibles en los que es viable aplicar técnicas de

reconocimiento de entidades (Named Entities Recognition) presentes en la base de conocimiento de Wikidata, así como reconocer propiedades con las que usarlas. La secuencia de operaciones que se plantea sería la siguiente:

1. Selección de un conjunto de categorías asociadas a un dominio de conocimiento.
2. Establecimiento de patrones de procesamiento de literales y asignación a propiedades y elementos de Wikidata.
3. Creación de declaraciones con cualificadores para cada categoría procesada, que contengan las entidades reconocidas y las propiedades sobre las que aplican.
4. Programación de bots para el procesamiento automático de categorías, enriquecimiento y validación de las descripciones de elementos de Wikidata.

Este trabajo se centra en las operaciones del tercer bloque (3) y, por lo tanto solo comentaremos superficialmente lo relativo al establecimiento del subconjunto de categorías (1) o las técnicas de PNL para establecer patrones (2). El objetivo específico es presentar los mecanismos de formalización en Wikidata de los resultados obtenidos del análisis de componentes de cualquier categoría (3) y en la descripción esquemática del procesamiento por lotes que posteriormente podría realizarse mediante bots y otros procedimientos automatizados (4).

Se utilizarían al máximo los propios recursos de Wikidata como base de conocimiento, para mejorar la propia representación de las categorías y describir los componentes semánticos contenidos en la misma. Lo anterior permitiría formalizar un mecanismo para crear declaraciones en Wikidata a partir de la indización con categorías en Wikipedia.

Sobre el primer bloque (1), la selección de un conjunto de categorías asociadas a un dominio de conocimiento, existen un sinfín de posibilidades. Mediante técnicas de clustering se pueden obtener conjuntos de artículos de temática afín (Minguillón, 2017), desde los que extraer las categorías usadas y construir un conjunto para su procesamiento para la detección de patrones (2). Para el dominio de las obras literarias, una sencilla consulta a las categorías que contienen la palabra “Novelas” nos permite reconocer patrones de los que se pueden derivar con relativa facilidad información descriptiva: autor, fecha, país de publicación, género u otras características. El desarrollo de patrones de lenguaje natural es un campo muy maduro, y su aplicación sobre conjuntos de términos como las categorías es mucho más operativa que sobre textos completos. La concisión y relativa homogeneidad de las expresiones usadas hacen de esta tarea una labor que podemos considerar una “commodity”, abordable mediante un conjunto conocido y variable de técnicas, librerías y algoritmos que ofrecerán resultados contrastados.

Tabla 3. Muestra de categorías que contienen el término “Novelas”⁵

Item	Literal
wd:Q6185886	Categoría:Novelas de Daniel Chavarría
wd:Q5882155	Categoría:Novelas de 1851
wd:Q3919902	Categoría:Novelas
wd:Q4064950	Categoría:Novelas de 2001
wd:Q6214588	Categoría:Novelas cortas de Japón
wd:Q6267331	Categoría:Novelas de Dominique Lapierre
wd:Q6285624	Categoría:Novelas de 1786
wd:Q6285939	Categoría:Novelas de Líbano
wd:Q6285951	Categoría:Novelas ligeras adaptadas de anime o manga
wd:Q6285955	Categoría:Novelas polémicas
wd:Q6285962	Categoría:Novelas tragicómicas
wd:Q6285960	Categoría:Novelas premiadas con el Premio Goncourt
wd:Q6285944	Categoría:Novelas ejemplares
wd:Q6285949	Categoría:Novelas fantásticas de Lois McMaster Bujold
wd:Q6285953	Categoría:Novelas pastoriles

3 Resultados

Una vez identificado un patrón de extracción de conceptos de una categoría, lo que queremos destacar en este trabajo que la propia estructura de datos de Wikidata permite codificar los resultados obtenidos. El análisis lingüístico de las categorías puede realizarse con cualquier algoritmo o librerías diseñado ad hoc, pero sus resultados podrían incorporarse al propio registro de Wikidata. Este punto es esencial en nuestro planteamiento, dado que los resultados del trabajo de extracción realizado externamente a la plataforma, a menudo como proyecto de investigación, queda incorporado usando el propio modelo de datos de Wikidata: de este forma, el Item correspondiente a la categoría procesada se enriquecería con las declaraciones de la descomposición de los conceptos extraídos, quedando este conocimiento accesible en el grafo de conocimiento para ser explotado de cualquier forma que posteriormente se encuentre oportuna.

Analizaremos el encaje en nuestra prueba de concepto de las propiedades actualmente disponibles en Wikidata (más de 9000). Para nuestro propósito hemos

⁵ Estos datos del Wikidata Query Service (<https://query.wikidata.org/>) con la siguiente consulta SPARQL:

```
SELECT distinct ?item ?literal WHERE {
  ?item wdt:P31 wd:Q4167836 .
  ?item rdfs:label ?literal
  FILTER contains(?literal,"Novelas") }
```

seleccionado las propiedades P971, P4224 y P1687, que se definen del siguiente modo:

- A. La propiedad P971 “Category combine topics” está prevista para describir los conceptos combinados en una categoría específica y actualmente se aplica sobre 648.478 categorías (16% del total de categorías de Wikidata).
- B. La propiedad P4224 “category contains”, que indica que la categoría contiene elementos que son instancias de un cierto elemento y se usa sobre 583.296 categorías (14% del total).
- C. La propiedad P1687 “propiedad Wikidata” se usa para relacionar un ítem con la principal propiedad con la que está relacionado y se usa en 4.440 ítems⁶.

(A) El uso de la propiedad P971 nos permite saber que la categoría “Novelas de Dominique Lapierre” (Q6267331) combina los Items “Novela” (Q8261) y “Dominique Lapierre” (Q1238249).

Tabla 4. Ejemplo de identificación de “topics” que definen de forma combinada la entidad de la categoría “Novelas de Dominique Lapierre” de Wikidata. Fuente: <https://www.wikidata.org/wiki/Q6267331>

Item sujeto	Propiedad	Item objeto
Novelas de Dominique Lapierre (Q6267331)	Category combine topics (P971)	Novela (Q8261) Dominique Lapierre (Q1238249)

De este modo, para todos los Items correspondientes a los artículos de Wikipedia indizados mediante dicha categoría se podría comprobar si se han incluido como objeto de alguna declaración los ítems Q8261 y Q1238249. En caso de no existir, podrían añadirse dichas declaraciones previa identificación de la propiedad más adecuada para vincular los ítems de los artículos con los ítems de autor y género. Sin embargo, no conocemos la propiedad a utilizar, puesto que, siguiendo el ejemplo anterior, en la combinación de temas del ítem Wikidata correspondiente a la categoría Q6267331 no se recoge la vinculación de cada uno de ellos con P50 (autor) o P136 (género literario).

De esta forma simple solo se recogen actualmente los objetos, y no las propiedades con las que pueden ser usados. Como la estructura de Wikidata permite aumentar la expresividad de las declaraciones mediante cualificadores, podríamos usarlos para expresar “este objeto puede usarse con esta propiedad”. Es necesario identificar una propiedad existente actualmente en Wikidata, que permita usarse como cualificador y que tenga un significado compatible con este propósito. Existen unas 180 específicas para ser usadas como cualificadores⁷. De todas ellas se ha optado por la propiedad P642 “Of”, entendida como “Scope of” y que puede usarse para

6 Consulta SPARQL (13-enero-2019): SELECT distinct ?item WHERE { ?item wdt:P971 ?topic } ;
SELECT distinct ?item WHERE { ?item wdt:P4224 ?topic } ; Conjuntamente son usadas en 63.283 categorías.

cualificar el ámbito de aplicación una declaración. Esta reificación contempla los elementos necesarios para usarla en la construcción de nuevas declaraciones en los ítem de Wikidata correspondientes a cualquier artículo de Wikipedia indizado con esa categoría. Idealmente nos encontraríamos con unos datos que descomponen la categoría usado de ejemplo de la siguiente manera:

Tabla 6. Propuesta de descomposición de la entidad Q6267331. Fuente: elaboración propia.

Ítem sujeto	Propiedad	Ítem objeto	Cualificador	Cualificación
Novelas de Dominique Lapierre (Q6267331)	Category combine topics (P971)	Novela (Q8261)	Scope of (P642)	Géneros literarios (Q223394)
Propiedad género literario (P136)	subject item of this property (P1629)	Género literario (Q483394)	-	-
Géneros literarios (Q223393)	Wikidata property (P1687)	Propiedad género literario (P136)	-	-

Esta cadena de relaciones podría expresarse como $Q6267331 > [P971 > Q8261] P642 > Q223393 > P1629 > Q483394 > P136$. Este mecanismo, pese a obtener los resultados finales deseados, implica la combinación de demasiados elementos, y una lógica susceptible de inconsistencias en caso de que no se establezcan relaciones unívocas entre ítems y propiedades, o no existan tales relaciones para una propiedad determinada.

(B) El uso de la propiedad P4224 nos permite saber que la categoría “Novelas de Dominique Lapierre” (Q6267331) se usa en ítems que del tipo - son instancias de (P31) - de “Obra literaria”. Como en la propia definición de la propiedad se incluye que la relación de instanciación, no hay duda sobre el uso potencial para construir statements. Puede usarse para validar o añadir la declaración inicial que todo ítem debe tener.

(C) Por último, la propiedad P1687 nos permite saber que la categoría “Novelas de Dominique Lapierre” (Q6267331) contiene aspectos de autoría y género, al servir para guardar relaciones con propiedades, en este caso P50 y P136. Al igual que sucedía con las declaraciones de P971, nos falta un elemento para abarcar el significado completo que buscamos, es decir, quién es el autor y cuál es el género. En este caso, el uso de los cualificadores sí es eficiente, puesto que la relación a través de la propiedad P642 “of” se establecerá con ítems (Q).

7 Se trata de aquellas definidas con el tipo “Wikidata qualifier” (Q15720608) y que extraemos del WDQS con la siguiente consulta SPARQL: `SELECT ?item WHERE { ?item wdt:P31 ?tipo. ?tipo wdt:P279* wd:Q15720608. }`

Tabla 7. Propuesta de descomposición de la entidad Q6267331 en propiedades de Wikidata.
Fuente: elaboración propia.

Item sujeto	Propiedad	Item objeto	Cualificador	Cualificación
Novelas de Dominique Lapierre (Q6267331)	Propiedad Wikidata (P1687)	Autor (P50)	Scope of (P642)	Dominique Lapierre (Q1238249)
	Propiedad Wikidata (P1687)	Propiedad género literario (P136)	Scope of (P642)	Novela (Q8261)



Figura 2. Modificación en Wikidata de la entidad Q6267331 aplicando la propuesta. Fuente: <https://www.wikidata.org/wiki/Q6267331>

En este ejemplo, para la categoría Q6267331 obtendríamos, a través de una consulta SPARQL, los siguientes pares de elementos y propiedades apropiados para construir declaraciones:

```
SELECT ?property ?propertyLabel ?object ?objectLabel WHERE
{
  wd:Q6267331 p:P1687 ?statement .
  ?statement ps:P1687 ?property .
  ?statement pq:P642 ?object .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en"
}
```

property	propertyLabel	object	objectLabel
wd:P50	author	wd:Q1238249	Dominique Lapierre
wd:P136	genre	wd:Q8261	novel

Figura 3. Ejemplo de consulta SPARQL y resultado de su ejecución en Wikidata Query Service. Fuente: elaboración propia.

4 Conclusiones

En este marco de convergencia entre metadata y taxonomías (Hedden, 2014) encontramos en Wikidata un caso de lo que Morville denominó como "hybrid

metadata ecology" (Morville, 2005). El trabajo de los editores al categorizar artículos se está aportando una forma de descripción de aspectos del contenido (metadatos descriptivos), aunque sin usar un modelo de datos adecuado: las categorías actuales cumplen más un papel descriptivo que de auténtica categorización.

El valor potencial de Wikidata como fuente de datos depende de la calidad de las descripciones de sus entidades, y, por lo tanto, de la cantidad de trabajo que la comunidad de editores sea capaz de movilizar eficientemente. Dado que el apoyo de bots es uno de los factores que complementan la participación de editores humanos en Wikipedia, en el caso de Wikidata se ofrecen aún más posibilidades de enriquecimiento y validación de contenidos, al tratarse de datos formalizados. El trabajo existente en la categorización de artículos en Wikipedia permite ser explotado en dos sentidos diferentes: para añadir nuevas declaraciones a los elementos, y para validar las ya existentes.

Aunque el sistema de categorías de Wikipedia está sometido a discusión y responde a necesidades coyunturales de organización y navegación de contenidos, no puede olvidarse que se trata de un trabajo directo de los editores de la enciclopedia. Este trabajo materializa conocimiento en relación con los artículos, y puede aprovecharse para mejorar y contrastar los datos almacenados en Wikidata. Esto responde al interés primario de la comunidad y respeta la construcción colaborativa de conocimiento por editores responsables.

Las categorías de Wikipedia, en su estado actual, no son el mecanismo óptimo de organización del conocimiento. Sin embargo, su utilidad para aumentar el alcance y capacidad del conjunto de datos semántico de Wikidata ofrece una nueva visión de ellas, como nexo para una transición hacia un grafo de conocimiento exhaustivo de todo lo editado hasta el momento en Wikipedia. De esta manera puede plantearse su desaparición o transformación sin pérdida de todo el trabajo realizado por miles de editores en todos los idiomas. El trabajo en un idioma se generaliza como conocimiento univocal al volcar conocimiento en un grafo común.

Hemos descrito un mecanismo operativo para codificar los conceptos y propiedades implícitos en una categoría, que usa elementos ya disponibles en la infraestructura de Wikidata. Se valida funcionalmente el uso de propiedad P4224 (B) para vincular con ítems a través de la propiedad P31 y de la propiedad P1687 (A) para especificar las propiedades inferidas al procesar los literales de una categoría y, a través de cualificadores vía la propiedad P642, indicar el elemento objeto de la relación. Se descarta, por el contrario, el uso de la propiedad P971 (A) dado que no especificarse las declaraciones con construirse cualificadores que apunten a propiedades. No obstante, para una mayor rigurosidad, se considera recomendable la creación de una nueva propiedad específica denominada "Category combines properties" cuyo dominio sean ítems instancia de "Wikimedia category" y su rango "Properties".

Estamos ante un proceso automatizable que, a partir de las categorías asignadas a cada artículo en cualquier idioma, enriquecería y validaría las declaraciones de cada elemento. Sea cual sea nuestra opinión de las categorías, la realidad es que se miles de

editores las vienen usando en los artículos de Wikipedia, y que por lo tanto se puede aprovechar una dinámica de edición que ya existe para extraer un valor secundario de ella de forma automatizada. Al mismo tiempo, serviría para obtener un esquema de conceptos de más calidad, al especificarse las unidades de significado de las categorías que suponen una composición de varios conceptos. En caso de una futura sustitución de las categorías por otros mecanismos más óptimos de navegación y agrupación de artículos, el conocimiento generado el usarlas para etiquetar artículos se habría reutilizado para ampliar y validar el grado de conocimiento de Wikidata.

Referencias

- Association of Research Libraries (2019) *ARL White Paper on Wikidata: Opportunities and Recommendations*. abril de 2019
- Hedden, H. (2016). *The accidental taxonomist* (2a). Information Today.
- Lambe, P. (2007). *Organising knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford: Chandos Publishing.
- Minguillón, Julià; Lerga, Maura; Aibar, Eduard; Lladós-Masllorens, Josep; Meseguer-Artola, Antoni (2017). Semi-automatic generation of a corpus of Wikipedia articles on science and technology. *El profesional de la información*, v. 26, n. 5, pp. 995-1004. Recuperado de: <https://doi.org/10.3145/epi.2017.sep.20>.
- Morville, Peter (2005) *Ambient findability: What We Find Changes Who We Become*. Sebastopol: O'Reilly.
- Pastor-Sánchez, J.A. & Saorín, T. (2018) Proposal for the integration of the semantic structure of Wikipedia categories using SKOS. *15th International ISKO Conference* (OPorto, 2018). Recuperado de: <http://hdl.handle.net/10760/38627>
- Saorín, T. & Pastor-Sánchez, J.A (2018). “Wikidata y DBpedia: viaje al centro de la web de datos”. *Anuario ThinkEPI*, v. 12, pp. 207-214. <https://doi.org/10.3145/thinkepi.2018.31>
- Tramullas, Jesús; Sánchez Casabón, Ana-Isabel; Garrido, Piedad. Wikipedia categories in research: towards a qualitative review of uses and applications. En Fernanda Ribeiro, Maria Elisa Cerveira (Coords.) *Challenges and Opportunities for Knowledge Organization in the Digital Age*, pp. 490 – 498, Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal

A representação do domínio da Ética em Organização do Conhecimento nas bases de dados Scopus e KO Literature

Isadora Victorino Evangelista¹, José Augusto Chaves Guimarães² y Daniel Martinez-Avila³

¹ ORCID [0000-0002-4682-9813](https://orcid.org/0000-0002-4682-9813). Departamento de Ciência da Informação, Universidade Estadual Paulista, Brasil.

isadora.evangelista@unesp.br

² ORCID [0000-0002-0310-2331](https://orcid.org/0000-0002-0310-2331). Departamento de Ciência da Informação, Universidade Estadual Paulista, Brasil.

chaves.guimaraes@unesp.br

³ ORCID [0000-0003-2236-553X](https://orcid.org/0000-0003-2236-553X). Departamento de Ciência da Informação, Universidade Estadual Paulista, Brasil.

martinez.avila@unesp.br

Resumo. Os estudos éticos sobre Organização do Conhecimento estão em constante desenvolvimento e uma forma de verificar sua importância como temática de pesquisa, é examinando sua representação em bases de dados utilizadas na área. A presente pesquisa baseou-se em pesquisa anterior, desenvolvida por Martínez-Ávila et al. (2015) e utilizou duas bases de dados de uso na Ciência da Informação – KO Literature e Scopus – para analisar a recuperação de artigos por meio dos termos *ethics* e *ethical*. O corpus gerou um total de 54 artigos, em que foi possível evidenciar o periódico da ISKO Knowledge Organization como fonte que mais publica sobre o tema e um grupo de 11 autores que se destacam na produção da temática. Foi realizada ainda uma análise de conteúdo, de maneira a identificar os valores éticos mais trabalhados pelo corpus, as metodologias e teorias utilizadas e ainda, os principais processos e ferramentas de organização do conhecimento que foram objeto de pesquisa. Como conclusão, destacou-se que a base KO Literature possui um alto nível de recuperação de documentos nessa temática, mas que ainda apresenta algumas barreiras em sua utilização.

Palabras clave: Ética em Organização do Conhecimento. KO Literature. Scopus. Análise de Domínio. Análise de Conteúdo.

Abstract. The studies on ethics in Knowledge Organization are in constant development. An effective way of verifying their importance as a research subject is to examine their representation in databases. The present paper builds upon previous research developed by Martínez-Ávila et al. (2015) and analyzes the literature on the topic that is indexed by two databases widely used in

Library & Information Science: KO Literature and Scopus. Our search for the terms "ethics" and "ethical" retrieved a total of 54 articles. Among the result we highlight the ISKO journal Knowledge Organization journal as the most frequent source and a group of 11 authors that stand out as the most productive ones. We also conducted a content analysis to identify the ethical values that are present in the corpus, the methodologies and theories that are used in the papers, and the main knowledge organization processes and tools of knowledge that object of research. In conclusion, we point out the KO Literature database as an effective database for the retrieval of documents on this topic, although still presenting some shortcomings.

Keywords: Ethics in Knowledge Organization. KO Literature. Scopus.

1 Introdução e pressupostos teóricos

Os estudos sobre os aspectos éticos em organização do conhecimento remontam à década de 70, relativamente à incidência de preconceitos em sistemas de organização do conhecimento - SOC's, notadamente classificações (Berman, 1971, Foskett, 1971).

Desde então, estudos foram realizados sobre aspectos tendenciosidade tesouros e no momento da indexação (Judge, 1981, Bell, 1991), visões culturais em tesouros multilíngues (Hudon, 1997), a representação homoafetiva em SOC's (Campbell, 2000, Pinho, 2010, Guimarães, Nascimento & Pinho, 2017), os conceitos de garantia e hospitalidade cultural (Beghtol, 2002), o poder do indexador para nomear e para estabelecer uma ética transcultural de mediação (Olson, 2002, García-Gutiérrez, 2002), e os valores, problemas e compromissos éticos na organização do conhecimento (Guimarães, Fernandez-Molina, Pinho & Milani, 2008, Tennis, 2013, Guimarães, Milani & Evangelista, 2015), dentre outros.

A isso se aliam três congressos sobre ética em Organização do conhecimento, realizados na cidade de Milwaukee respectivamente nos anos de 2009, 2012 e 2015 abordando questões como o respeito às necessidades dos usuários e à ética na catalogação (Hoffman, 2009, Shoemaker, 2015), a classificação em temas de segurança nacional (Unsworth, 2009), a representação de etnias e grupos nacionais em SOC's (Beall, 2009), a representação de conteúdos sexuais (Keilty, 2012), as consequências éticas na digitalização de documentos de arquivo (Zhang, 2012), a representação homoafetiva em linguagens documentárias brasileiras (Pinho & Guimarães, 2012), o uso de social tagging como alternativa à utilização de SOC's, (Adler, 2012), os conceitos de disfemismo, eufemismo e ortofemismo na indexação (Milani & Pinho, 2012), o conceito de cultura na construção de SOC's (Lee, 2015), aspectos éticos da autoria (Oravec, 2015), a mediação de discursos na representação do conhecimento (Arboit & Guimarães, 2015), os valores éticos nos processos de descrição e classificação arquivística (Silva, Guimarães & Tognoli, 2015), as tendenciosidades nas recomendações de serviços da Netflix (Lawrence, 2015), dentre outros.

Desse modo, e considerando que os estudos éticos sobre Organização do Conhecimento estão em constante desenvolvimento podendo sua importância de pesquisa ser verificada a partir de sua representação em bases de dados e considerando pesquisa anterior de Martinez-Avila, Guimarães, Pinho & Adler (2015) sobre a representação da temática ética em organização e representação do conhecimento em duas importantes bases de dados utilizadas no universo da Ciência da Informação: Web of Science (WoS) e Library and Information Science and Technology Abstracts (LISTA), o presente trabalho busca identificar como se dá a presença de temas sobre ética em organização e representação do conhecimento nas bases Scopus, editada pela Elsevier e considerada a maior base de dados de textos científicos revisados por pares (Scopus, 2018) e, de forma mais específica, na base da International Society for Knowledge Organization, KO Literature que é constantemente alimentada por especialistas internacionais em organização do conhecimento, proporcionando, regular e periodicamente, referências sobre as diferentes temáticas que compõem o Classification System for Knowledge Organization Literature (DALHERG, 1993).

Para tanto, realizou-se uma análise quantitativa de incidência dessa literatura nas referidas bases, de modo a identificar, ainda, as fontes de publicação e os autores preponderantes sobre esse tema. Aliado a esses aspectos, realizou-se ainda uma análise de conteúdo de forma a identificar valores éticos, metodologias e teorias utilizadas e principais processos e ferramentas de organização do conhecimento que foram objeto de pesquisa.

2 Metodologia

Como metodologia de estudo, utilizou-se a análise de domínio na abordagem bibliométrica, como um “paradigma social” que busca entender o campo a partir do coletivo de indivíduos que se dedicam a uma mesma temática (Hjørland & Albrechtsen, 1995) aliada à análise de conteúdo (Bardin, 2009), pautada na análise dos textos dos documentos a partir de inferências controladas por meio de categorias pré-definidas, respaldadas pelo conhecimento prévio do pesquisador.

Cumprido ressaltar a importância das bases de dados analisadas para a Ciência da Informação. A base referencial Scopus é editada pela Elsevier, que afirmam ser a maior base de dados de textos científicos revisados por pares. Além de uma alta representatividade na área das ciências sociais, possuindo ainda diversas ferramentas para análise dos dados. Já a KO Literature é a base de dados da maior organização internacional em organização do conhecimento, a *International Society for Knowledge Organization* (ISKO). Essa base é alimentada por especialistas internacionais no âmbito da organização do conhecimento e ligados à ISKO que proporcionam regular e periodicamente referências sobre as diferentes temáticas que compõem a classificação desenvolvida por Ingetraut Dahlberg *Classification System for Knowledge Organization Literature* (DALHERG, 1993).

Embora no artigo de Martínez-Ávila et al. (2015) a base de dados Scopus tenha sido preterida pela WoS, por conta de a segunda indexar artigos que remontam ao ano 1900, a escolha da presente pesquisa pela Scopus deve-se ao fato de ser uma comparação justa com a segunda base utilizada, KO Literature, que possui documentos indexados desde o início do periódicos e congressos da sociedade ISKO, no ano de 1988.

Como mecanismos de busca, na base de dados KO Literature é possível pesquisas apenas por termos no título do documento, ano ou autor e para estabelecer equidade entre as bases, optou-se por pesquisar na Scopus apenas por termos nos títulos dos documentos, embora a base de dados ofereça mais opções de pesquisa.

Em um primeiro momento, tentou-se utilizar o truncamento de pesquisa *ethic**, mas a base KO Literature ainda não aceita recuperações por truncamento ou operadores booleanos. Dessa forma, nessa base de dados foi necessário realizar duas pesquisas, nos títulos dos documentos: a primeira pelo termo *ethics* e a segunda, pelo termo *ethical*, visto que são os dois termos mais abrangentes em relação à temática pesquisada.

Na base de dados Scopus, mesmo pesquisando pelo termo *ethics*, a base inferiu a importância do termo *ethical* para pesquisa e recuperou documentos que também possuíam esse segundo termo em seus títulos. Como essa não é uma base específica para a área de organização do conhecimento – diferente da anterior –, nessa busca foi necessário combinar o termo “*knowledge organization*” em todas as áreas dos documentos, para que a recuperação fosse estritamente de documentos da área.

A pesquisa foi realizada no dia 16 de agosto de 2018 e não foi deliberado nenhum filtro temporal, recuperando assim os artigos publicados com esses termos independentemente do ano de publicação.

3 Resultados

Foram recuperados 54 artigos na base KO Literature e 07 artigos na base Scopus, sendo que todos esses artigos já se encontravam recuperados na base KO. Nesse contexto, a revista Knowledge Organization, da ISKO, lidera fortemente com 17 artigos, seguida da série Advances in KO, também da ISKO, com 06 trabalhos. Em terceiro lugar figura a revista Cataloging and Classification Quarterly, com 04 artigos. Vale destacar que a revista KO publicou todos os trabalhos apresentados nos congressos de ética em KO realizado em Milwaukee (EUA) nos anos de 2012 e 2015 e a revista Cataloging and Classification Quarterly, publicou trabalhos apresentados no referido congresso realizado no ano de 2009. Ressalta-se ainda a presença do periódico alemão Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare e os proceedings do capítulo alemão da ISKO, totalizando 04 trabalhos, o que em muito se explica pelo fato de o congresso de ISKO-Alemanha de 2006 possuir como temática a ética em organização do conhecimento.

Outra presença destacar, nesse contexto, é a brasileira, com 06 trabalhos publicados nas revistas *Encontros Bibli*, *Informação & Sociedade* e *Brazilian Journal of Information Studies (BRAJIS)* bem como nos *proceedings* dos congressos da ISKO-Brasil.

Ainda que fora do universo das revistas e dos eventos especializados, destaca-se também a ocorrência do capítulo intitulado “Ethical decision-making for knowledge representation and organization systems for global use”, de autoria de Clare Beghtol, integrando o livro *Knowledge Organization, Information Systems and other Essays*,

Nesse contexto, observou-se a forte prevalência de textos em língua inglesa (36 textos), o que se explica pelo fato de ser esta a língua oficial da ISKO, mas foi possível ainda, obter documentos em alemão (05 documentos), português (06 documentos) e espanhol (07 documentos).

Essa produção científica encontra-se a cargo de 66 autores, com destaque para onze deles (20% do universo) que publicaram ao menos dois artigos cada e em forte colaboração, sinalizando para a existência de uma comunidade epistêmica (Meyer & Molineux-Hodgson, 2010) integrada por Guimarães, Pinho, Milani, Martínez-Ávila e Tognoli e Silva, do Brasil; Beghtol, Fox e Smiraglia, dos Estados Unidos, Fernández-Molina, da Espanha; e Van der Walt, da África do Sul. Em termos institucionais, destacam-se a Universidade Estadual Paulista (Brasil) e a University of Wisconsin-Milwaukee (Estados Unidos) como núcleos investigativos nessa temática.

Cumprir destacar que, por conta de uma limitação metodológica da pesquisa nas referidas bases de dados – necessária para a adoção de um mesmo critério para ambas – importantes pesquisas na temática estudada não se fizeram presentes, como é o caso de Olson (2002), Hudon (1997), Feinberg (2007) e Tennis (2013).

Em um segundo momento, foi realizada a análise de conteúdo dos artigos (Bardin, 2009), de maneira a identificar as principais temáticas de pesquisa estudadas por esses especialistas. Como categorias de análise teve-se os valores éticos abrangidos (Guimarães et al., 2008, Guimarães, Milani & Evangelista, 2015) e abordagens metodológicas utilizadas.

Identificaram-se 24 valores, em um total de 114 incidências, cuja prevalência recaiu nos valores relativos à responsabilidade (profissional e jurídica) sobre o desenvolvimento de processos, produtos e instrumentos de OC (presente em 20% do corpus); acesso à informação e respeito à diversidade (cada um presente em 17% do corpus), adequação (precisão) na representação do conhecimento (13% do corpus), foco no usuário, garantia cultural e hospitalidade cultural (cada um presente em 11% do corpus). Como se pode observar, tem-se supervalores, como o acesso à informação, que, como já explicado por Guimarães et al. (2008) consiste em valores eminentemente voltados para a figura do usuário e valores que dizem respeito a aspectos técnicos da atuação profissional, como a precisão da representação.

Essa produção de conhecimento respaldou-se de um conjunto de teorias específicas, em que se destaca uma maior presença do pragmatismo, da teoria crítica, do pós-estruturalismo e do utilitarismo em trabalhos fortemente pautados na revisão de literatura e nas análises bibliométricas, bem como no recurso a metodologias como

a análise de domínio, a análise de conteúdo e a análise do discurso. Esse panorama revela, por um lado, uma preocupação com a configuração teórica dessa temática e, por outro, com a aplicabilidade dessas discussões no universo social, em que sujeitos, comunidades e discursos interagem constantemente.

4 Conclusões

A vista dos resultados obtidos pode-se concluir que a base KO Literature revelou-se mais representativa da produção científica na temática analisada, ainda que apresente a limitação de permitir busca tão somente por título, autor ou área de classificação do periódico Knowledge Organization.

Em suma, ficou demonstrada a liderança do universo científico da ISKO nas discussões dessa temática, uma vez que sua revista e os anais de seus congressos científicos lideram fortemente o corpus recuperado.

Esse universo se caracteriza por dois polos principais de produção científica, situados no Brasil e nos Estados Unidos, com uma comunidade epistêmica em forte interação.

Por fim, observa-se que essa temática vem sendo objeto de uma reflexão teórica para identificar suas bases conceituais, seus valores e seus traços distintivos, mas sem deixar de lado a importância de sua aplicabilidade em distintos contextos sociais. Essa reflexão, por sua vez, pauta-se em um conjunto de valores éticos que permeiam tanto a organização do conhecimento como um objeto em si mesmo quanto seu papel social, o que confirma o fato de a ética em organização do conhecimento haver se tornado efetivamente uma dimensão temática desse campo, com fortes perspectivas de crescimento, que se intensificam com as demandas de um mundo cada vez mais globalizado e permeado pela tecnologia.

Agradecimentos: a autora Isadora Victorino Evangelista agradece o financiamento por meio do processo nº 2017/02327-8, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). O autor José Augusto Chaves Guimarães agradece o financiamento por meio do processo nº 2017/00584-3, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- Adler, M. A (2012). Disciplining knowledge at the Library of Congress. *Knowledge Organization*, 39 (5), pp. 370-376.
- Arboit, Aline Elis & Guimarães, José Augusto Chaves (2015). The ethics of knowledge organization and representation from a Bakhtinian perspective. *Knowledge Organization*, 42 (5), pp. 324-331.
- Bardin, Laurence (2009). *Análise de conteúdo*. Edições 70: Lisboa.

- Beall, Julianne (2012). Racially mixed people, DDC Table 5 Ethnic and National Groups, and MARC 21 Bibliographic Format Field 083. *Cataloging and Classification Quarterly*, 47 (7), pp. 657-670.
- Beghtol, Clare (2002). A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, 58 (5), pp. 507-532.
- Beghtol, Clare (2006). Ethical decision-making for knowledge representation and organization systems for global use. In: Raghavan, K. S. & Prasad, K. N. (Ed.) *Knowledge Organization, Information Systems and other Essays: Professor A. Neelamegham Festschrift*. New Delhi: Ess Ess Publications, pp.41-52.
- Bell, Hazel (1991). *Bias in indexing and loaded language*. The Indexer, 17 (3), pp. 173-177.
- Berman, S. (1971). *Prejudices and antipathies: a tract of Library of Congress Subjects Headings concerning people*. Metuchen: Scarecrow Press.
- Campbell, Grant (2000). Queer theory and the creation of contextual subject access tools for gay and lesbian communities. *Knowledge Organization*, 27 (3), pp. 122-131.
- Dahlberg, Ingetraut (1993). Knowledge organization: its scope and possibilities. *Knowledge Organization*, 20 (4), pp. 211-222.
- Feinberg, Melanie (2007). Hidden bias to responsible bias: an approach to information systems based on Haraway's situated knowledges. *Proceedings of the sixth International Conference on Conceptions of Library and Information Science – "Featuring the future"*, 12 (4).
- Foskett, A. C. (1971). Misogynists all: a study in critical classification. *Library Resources & Technical Services*, 15 (2), pp. 117-121.
- García Gutiérrez, Antonio (2002). Knowledge organization from a "culture of the border": towards a transcultural ethics of mediation. In: López-Huertas, M.M. (Ed.). *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Würzburg: ERGON-Verlag, pp. 516-522.
- Guimarães, José Augusto Chaves, Fernandez-Molina, Juan Carlos, Pinho, Fábio Assis & Milani, Suellen de Oliveira (2008). Ethics in the knowledge organization environment: an overview of values and problems in the LIS literature. In: Arsenault, C. & Tennis, J. T. (Ed.). *Cultural and Identity in Knowledge Organization*. Würzburg: ERGON Verlag, pp. 340-346.
- Guimarães, José Augusto Chaves, Milani, Suellen de Oliveira & Evangelista, Isadora Victorino (2015). Valores éticos na organização e representação do conhecimento. *Encontros Bibli*, 20 (1), pp. 19-32.
- Guimarães, José Augusto Chaves, Nascimento, Francisco Arrais & Pinho, Fábio de Assis (2017). The metaphorical dimension of LGBTQ information: challenges for its subject representation. *Informação & Sociedade*, 27 (3), pp. 49-57.
- Hjørland, Birger & Albrechtsen, Hanne (1995). Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46 (6), pp. 400-425.
- Hoffman, Gretchen (2009). Meeting users' needs in cataloging: what is the right thing to do?. *Cataloging and Classification Quarterly*, 47 (7), pp. 631-641.

- Hudon, Michele (1997). Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. *Knowledge Organization*, 24 (2), pp. 84-91.
- Judge, A (1981). *Anti-developmental biases in thesaurus design*. Recuperado em 06 agosto, 2018, de <http://www.laetusinpraesens.org/docs/antidev.php>.
- Keilty, Patrick (2012). Tagging and Sexual Boundaries. *Knowledge Organization*, 39 (5), pp. 320-324.
- Lawrence, Emily (2015). Everything is a ecommendation: netflix, altgenres and the construction of taste. *Knowledge Organization*, 42 (5), pp. 358-364.
- Lee, Wan-Chen (2015). Culture and classification: an introduction to thinking about ethical issues of adopting global classification standards to local environments. *Knowledge Organization*, 42 (5), pp. 302-307.
- Martínez-Ávila, Daniel, Guimarães, José Augusto Chaves, Pinho, Fabio Assis & Fox, Melodie J. (2015). The representation of ethics and knowledge organization in the WoS and LISTA databases. *Knowledge Organization*, 42 (5), pp. 269-275.
- Milani, Suellen Oliveira & Pinho, Fabio Assis (2012). Knowledge Representation and Orthophemism: a reflection aiming to a concept. *Knowledge Organization*, 39 (5), pp. 377-393.
- Oravec, Jo Ann (2015). The moral imagination in an Era of “Gaming Academia”: implications of emerging reputational issues in scholarly activities for knowledge organization practices. *Knowledge Organization*, 42 (5), pp. 316-323.
- Pinho, Fabio Assis (2010). *Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina: uma análise da precisão em linguagens de indexação brasileiras*. Doutorado em Ciência da Informação. Universidade Estadual Paulista.
- Pinho, Fabio Assis & Guimarães, José Augusto Chaves (2012). Male homosexuality in brazilian indexing languages: some ethical questions. *Knowledge Organization*, 39 (5), pp. 363-369.
- Scopus (2018). *What is Scopus?*. Recuperado em 29 novembro, 2018, de <https://www.elsevier.com/solutions/scopus>.
- Shoemaker, Elizabeth (2015). No one can whistle a symphony: seeking a catalogers’ code of ethics. *Knowledge Organization*, 42 (5), pp. 353-357.
- Silva, Andrieli Pachú, Guimarães, José Augusto Chaves & Tognoli, Natália Bolfarini (2015). Ethical values in archival arrangement and description: an analysis of professional codes of ethics. *Knowledge Organization*, 42 (5), pp. 346-352.
- Tennis, Joseph (2013). Ethos and ideology of knowledge organization: toward precepts for an engaged knowledge organization. *Knowledge Organization*, 40 (1), pp. 42-49.
- Unsworth, Kristene. (2009) Ethical Concerns of Information Policy and Organization in National Security. *Cataloging and Classification Quarterly*, 47 (7), pp. 642-656.
- Zhang, Jane (2012). Archival context, digital content, and the ethics of digital archival representation. *Knowledge Organization*, 39 (5), pp. 332-339.

Garantía cultural: avance de elementos teóricos y metodológicos para su aplicación

Mario Barité

¹ ORCID [0000-0002-2992-6582](https://orcid.org/0000-0002-2992-6582). Universidad de la República, Uruguay.
mario.barite@fic.edu.uy.

Resumen. Las garantías son la justificación para seleccionar rigurosamente términos para la representación temática (clasificación, indización o etiquetado). La garantía cultural atiende los valores y las creencias particulares que una comunidad puede asumir, entender o interpelar al conocimiento establecido. Se presenta sumariamente el estado de situación de la garantía cultural en sus aspectos teóricos, metodológicos y aplicativos. Se reseñan cuatro elementos teóricos orientadores: noción de cultura como problema; dimensión local del conocimiento; representación del conocimiento para sectores sociales minoritarios o discriminados; factor ético. Además de cuatro tipos genéricos de análisis cualitativo, se indican otras tres destinadas a implantar una perspectiva cultural: creación de espacios paradójicos, adaptaciones locales de sistemas universales, implantación de sesgos positivos. Se concluye que la garantía cultural, como concepto en evolución, promueve el sentido social y democrático ya que postula la tolerancia entre concepciones diferentes, y el respeto por la integridad cultural de colectivos insertos en nuestras sociedades. .

Palabras clave: Garantía cultural; Sesgos; Sesgo positivo; Hospitalidad cultural; Metodologías cualitativas.

1 Las garantías en Organización del Conocimiento

La cuestión de las garantías guarda una nueva dimensión en Organización del Conocimiento (OC), a punto de partida de los trabajos analíticos y sistematizadores de Rodríguez (1984) y Beghtol (1986). Entendidas como la justificación para incorporar o excluir términos de referencia para la representación temática (clasificación, indización o etiquetado) y la recuperación de información, las garantías pretenden facilitar agrupaciones de documentos que tratan tópicos comunes y búsquedas temáticas en cualquier contexto informacional, así como aportar mayor rigor en la selección de la terminología de representación.

Beghtol estableció que la garantía es la autoridad que el clasificacionista invoca para seleccionar la terminología de un sistema de clasificación (Beghtol, 1986, p.

Barité, M. (2020). Garantía cultural: avance de elementos teóricos y metodológicos para su aplicación. En J. Tramullas, P. Garrido-Picazo y G. Marco-Cuenca (eds.) *Actas del IV Congreso ISKO España y Portugal 2019* (pp. 475-483). <https://doi.org/10.5281/zenodo.3743443>

110-111), idea que en una perspectiva más amplia puede extenderse sin dificultad a cualquier otro sistema de organización del conocimiento (SOC), y a cualquier contexto de información, convencional o digital (Bullard, 2017).

Un trabajo reciente (Barité, 2018, p. 528) ofrece un cuadro que registra veintiuna garantías, propuestas por distintos autores durante más de un siglo. Algunas parecen ser variaciones o especificaciones de las cuatro más mencionadas en la literatura: la garantía literaria que se fundamenta en la terminología suficientemente documentada; la garantía de usuario que pone énfasis en el lenguaje de las personas que buscan información y trabajan con ella; la garantía académica, que recalca en los consensos científico-técnicos y especializados denominativos de objetos y fenómenos; y la garantía cultural, orientada a atender los valores, las creencias y las miradas particulares con los que una comunidad puede asumir, entender o interpelar al conocimiento establecido, o puede crear su propio universo semántico al compartir las mismas coordenadas culturales.

En este trabajo se pretende presentar en grandes líneas, un estado de situación de la garantía cultural en sus aspectos teóricos, metodológicos y aplicativos, identificando los puntos críticos que requerirían mayor atención de los investigadores.

2 Noción de la garantía cultural

El concepto de garantía cultural (*cultural warrant*) fue acuñado por Lee (1976), a partir de una interpretación propia de los alcances de la garantía literaria, concepto introducido por Hulme en 1911. En efecto, para Lee, la garantía literaria se concibe como una aplicación práctica de los valores culturales que se expresan en la literatura producida dentro de una cultura. Argumentaba que si una clasificación está basada en la literatura existente de un área del conocimiento, ella reflejará las pautas culturales imperantes, lo que al fin y al cabo podría resultar más fidedigno que un orden ideal de las ciencias (Lee, 1976).

Cabe a Beghtol el mérito de rescatar la noción de garantía cultural y construir sus primeros cimientos (Beghtol, 1986). En él señala que la garantía cultural es “*an umbrella concept that covers and at least partially explains the developmental changes in the kinds of semantic warrant*”. Y agrega que “*changes in the conceptions and uses of literary warrant, scientific/philosophical warrant, and educational warrant can all, then, be viewed as detailed case studies of the more general concept of cultural warrant*” (Beghtol, 1986, p. 121).

Más adelante, es también Beghtol quien afina y extiende sus ideas señalando que “*cultural warrant posits that every classification system is based on the assumptions and preoccupations of a certain culture, whether the culture is that of a country, or of some smaller or larger social unit (e.g. ethnic group, academic discipline, arts domain, political party, religion and/or language)*” (Beghtol, 2002a, p. 45).

En esa afirmación despuntan cuatro elementos teóricos orientadores para la comprensión del concepto, que se reseñarán más adelante:

- i. la noción de cultura como problema, vista desde la Antropología y otras disciplinas ;
- ii. la atención preferente a la dimensión local del conocimiento frente a perspectivas universalistas;
- iii. la preocupación por favorecer la representación del conocimiento útil para sectores sociales minoritarios, relegados o discriminados, pero también para grupos empoderados que exigen imponer sus propios sistemas nocionales a través del lenguaje común y del discurso especializado;
- iv. el factor ético en relación con el respeto por la integridad del pensamiento cultural de una comunidad (Barité, 2011).

La definición de Beghtol también obliga, por vía tácita, al necesario desarrollo de metodologías para la aplicación adecuada del principio de garantía cultural en las prácticas de representación del conocimiento y de recuperación de información.

Los SOC, si bien intentan representar el mapa de las disciplinas y especialidades con objetividad y neutralidad, están condicionados histórica y culturalmente, ya que reflejan las mentalidades sociales, políticas y religiosas de sus respectivos tiempos (González Casanova, 1996). Por esa razón, y a su favor, se ha aventurado incluso que los sistemas de clasificación pueden constituirse en índices culturales fiables para los historiadores (Bertrand-Gastaldy, 1993).

Todos los recursos de información que tienen valor patrimonial, de identidad y memoria social, están impregnados de concepciones y referencias culturales lo que hace impacto en aspectos de gestión cultural y patrimonial a nivel de bibliotecas, archivos, repositorios y museos. Ello permite justipreciar la necesidad de prestar mayor atención a esas concepciones y referencias desde la perspectiva de la OC.

En los apartados siguientes se ensayarán en grandes trazos, formulaciones que comiencen a dar respuesta a estas cuestiones teóricas y metodológicas..

3 Cuestiones teóricas

Adentrarse en los dominios de la garantía cultural implica considerar -como punto de partida- una definición de cultura. Todo intento al respecto debe considerar los elementos que aportan homogeneidad a una comunidad de personas, y consolidan una manera particular de concebir la realidad. Esta visión puede integrarse y convivir con otras visiones, o puede excluir o enfrentar a otras dentro de una misma sociedad. En cualquier circunstancia, la noción de cultura que se tome como referencia se encontrará implicada al momento de establecer políticas y criterios de clasificación, indización y otras formas de representación del conocimiento.

Boccato y Biscalchin reseñan diferentes dimensiones de lo cultural, que incidirán en la forma de representar asuntos atingentes a las ciencias sociales y humanas. Estos autores mencionan a la culturalidad, el culturalismo, la multiculturalidad, el multiculturalismo, la interculturalidad y la transculturalidad (Boccato y Biscalchin, 2014). El tratamiento particularizado y comparado de esas dimensiones incide no solo en una mejor desagregación del concepto de cultura, sino además, en las precauciones de método que deben considerarse al momento de llevar a campo la aplicación de la

garantía cultural. Surge pues el imperativo de trabajar en una perspectiva particular de la cultura en la OC, que incorpore todas las dimensiones mencionadas a su construcción teórica y metodológica.

Reconocer las dimensiones locales del conocimiento equivale a enfrentar la concepción de la ciencia y la técnica como un conjunto de ideas, soluciones y aplicaciones válido en cualquier lugar y tiempo, con la de las aplicaciones particulares de una ciencia o una técnica general que, por atender a creencias, valores y contextos específicos asume una legitimidad más inmediata. Parece razonable considerar que personas pertenecientes a diferentes culturas, cuentan con distintas necesidades de información y formas diversas de interpretación del conocimiento establecido. En el fondo hay una pugna entre las soluciones universales propugnadas por los responsables de sistemas de clasificación, tesauros, listas y otros instrumentos, que parten de la noción de conocimiento compartido a escala internacional, y quienes prefieren herramientas de representación temática que valgan en determinado territorio o momento.

Gran parte de la investigación aplicada tiene referencias locales, y hay campos temáticos que son relevantes solo en un espacio geográfico (que es también una dimensión cultural) y no en otro: la música, el lenguaje, las costumbres, el Derecho, la Historia. Como expresan García Aguilar & Villén Rueda, “cada sociedad establece y determina los símbolos y significados de su cultura (...) que le son esenciales y que le distinguen de otras” (García Aguilar & Villén Rueda, 2000, p. 295). Así, la cultura del mate está ampliamente extendida en Paraguay, Argentina, Uruguay y el sur del Brasil, y cuenta con un cuerpo propio de literatura, es decir, con su propia garantía literaria. Por ello sería del todo pertinente contar con esquemas de clasificación que representen temáticamente la documentación específica de esa infusión y de todos sus componentes culturales (ritos, objetos, procedimientos, códigos de conducta) que la acompañan.

En la misma línea, una comunidad socialmente minoritaria pero con una firme cohesión religiosa, étnica, ideológica y/o filosófica, suele generar cultores, documentación, reglamentos o códigos, objetos, ritos y procedimientos que exigen ser contemplados en los sistemas de información o en los SOC destinados a la clasificación y la indización documental. Por contraste, las rémoras ideológicas, también apoyadas en la herencia cultural -que no siempre asume valores positivos-, llevan a que los SOC de uso más extendido dejen traslucir en sus esquemas percepciones socioculturales que pueden resultar sesgadas, tendenciosas y hasta ofensivas para comunidades locales (Pacey, 1989; Caro y San Segundo, 1999; Colombo, 2018).

Al involucrar en la representación temática la visibilidad de grupos particulares, exhibiendo sus propias pautas de convivencia y de comunicación, la garantía cultural introduce el factor ético en la OC (Beghtol, 2002b; Guimarães, 2006). Esto es especialmente importante cuando se enfoca la atención en las perspectivas de movimientos sociales que impulsan nuevas epistemologías, así como la

deconstrucción de perspectivas o mentalidades anteriores, como en las cuestiones de género (Guimarães et al., 2008; Guimarães & Pinho, 2007).

4 Las garantías en Organización del Conocimiento

La producción escrita sobre métodos de la garantía cultural es relativamente escasa, al punto que a la fecha solo es posible realizar una primera aproximación sistemática.

Lo primero a establecer es que las metodologías aplicables a este respecto son cualitativas, pues implican análisis valorativos a partir de la formación experta en información: i) análisis de contenido: clásica herramienta al servicio de la indización y el etiquetado de contenidos (Krippendorff, 2004; Hsieh & Shannon, 2005); ii) análisis terminológico: que permite identificar y seleccionar los términos y las relaciones entre términos más apropiados para una comunidad de usuarios (por ejemplo, Benyaich, 2014); iii) análisis del discurso (Van Dijk, 1999); y, iv) análisis de dominio: que contribuye en la configuración interna de un área del saber, al identificar sus fundamentos, teorías, métodos, prácticas, aplicaciones y productos (Hjørland, 2002, 2004).

Los cuatro métodos analíticos mencionados requieren de al menos cuatro formas de adecuación a un contexto cultural particular:

a) adaptación al conjunto de valores y creencias que una comunidad cultural acepta como verdaderos o plausibles;

b) adaptación terminológica, de modo de seleccionar los conceptos con sus términos y las relaciones entre términos que reflejen el sistema de valores y creencias de dicha comunidad;

c) adecuación ética, que neutralice el uso de expresiones discriminatorias, estigmatizantes o peyorativas;

d) adecuación a los equilibrios internos del dominio, que evite casos como el de la representación de la Religión en los sistemas decimales CDU y Dewey, en donde la amplia mayoría de las tablas se asigna a las religiones cristianas, en detrimento de otras ampliamente cultivadas en el mundo, como el Islam o el Judaísmo.

Existen además soluciones metodológicas que se han promovido para resolver problemas de representación temática vinculados al enfoque cultural. Se trata de técnicas y métodos que favorecen la hospitalidad cultural, entendida esta como el conjunto de previsiones para incluir aspectos y tópicos específicos de diferentes culturas en los SOC, de modo de volverlos permeables a distintas perspectivas culturales (Beghtol, 2002a). Podrían entenderse como formas de hospitalidad cultural la inserción premeditada de términos, facetas o esquemas completos con cierta orientación cultural, dentro de clasificaciones preexistentes.

En ese contexto, la creación de espacios paradójicos, herramienta propuesta por Olson y Ward (1998) para insertar terminología de género en las tablas de la Clasificación Decimal Dewey, es una manifestación de hospitalidad cultural. Implica subdividir un concepto general a efectos de incorporar conceptos particulares omitidos. A modo de ejemplo, Olson y Ward (1998) subdividen el término 'Base económica del trabajo' para incluir el tópico 'trabajo no remunerado', como sigue:

- 331.116 Base económica del trabajo (registrado en CDD).
- 331.116 2 Trabajo no remunerado (registrado en CDD)
- 331.116 3 Trabajo remunerado (incorporado por espacio paradójico)

Las adaptaciones locales de sistemas universales de clasificación, constituyen otra modalidad de expresión de la hospitalidad cultural, en la medida en que pretenden atender las peculiaridades de un país o una región. Estas adaptaciones pueden ser diseñadas por los responsables oficiales de los SOC (por ejemplo, Beall, 2003; Choi, 2018), o pueden tener un cariz local, generado en lo interno de un sistema de información, o de un sistema o red de bibliotecas, con las ventajas y desventajas que esto conlleva.

A partir de sus trabajos sobre sesgos, Colombo y Barité (2015) y Colombo (2018) sugieren que lo que denominan sesgo positivo, entendido como la voluntad premeditada de orientar la terminología de un SOC en una determinada dirección (el pensamiento budista, la concepción marxista, una aproximación evangélica, una postura feminista en la elección de los términos) es otra puerta de acceso a la atención, en este caso integral, de una perspectiva cultural.

5 Conclusiones

La garantía cultural es un concepto en evolución, que hasta el presente solo tiene asentadas de forma muy general sus premisas teóricas, y no ha contado con una aproximación sistemática a las modalidades metodológicas más apropiadas para su inclusión o implantación en SOC y en la terminología interna de los sistemas de información. De hecho, se requeriría establecer una perspectiva propia del concepto de cultura dentro de la OC y de las dimensiones de lo cultural que fueron mencionadas más arriba, lo que también es una tarea pendiente.

Este trabajo pretende contribuir a pavimentar el camino que lleve a una mayor atención a estas cuestiones. Ello implica -por añadidura- jerarquizar el análisis de las maneras en que los usuarios se relacionan con los términos de indización y con las formas de organización de los documentos, pues todas esas maneras involucran sus bagajes culturales individuales, y sus formas de entender la realidad y asimilar el conocimiento establecido.

La implantación de una perspectiva cultural para la OC exige la construcción de una terminología consistente, apegada al conjunto de creencias o de ideas que se pretende representar, y depurada de todos los elementos peyorativos, discriminatorios o inapropiados que puedan ponerla en cuestión. En la garantía cultural se trabaja, pues, con material sensible para comunidades, colectivos y movimientos sociales que actúan en actitud reivindicativa, y que además suelen contener corrientes disimiles en cuanto a estrategias y pensamiento político. Por ello, conviene que los trabajos de campo y los productos terminológicos creados con una perspectiva cultural puedan contar con la validación -al menos sumaria- de los colectivos atendidos.

La convicción de que las metodologías relacionadas con la garantía cultural son de índole cualitativa, obliga a explorar los beneficios que dichas metodologías pueden aportar para llevar a tierra los postulados teóricos que han prevalecido hasta el presente en la literatura.

Por último, cabe mencionar el sentido social y democrático que aporta la garantía cultural a nuestra área del saber, en la medida en que postula la tolerancia entre concepciones diferentes, y el respeto por la integridad cultural de colectivos insertos en nuestras sociedades, con derecho reconocido a mantener vivos los valores de sus culturas particulares.

Referencias

- Barité, Mario (2018). Literary warrant. *Knowledge Organization*, 45(6), 517-536.
- Barité, Mario (2011). La garantía cultural como justificación en sistemas de organización del conocimiento: aproximación crítica. *Palabra Clave*, 1(1), 2-11.
- Beall, Julianne (2003). Approaches to Expansions: Case Studies from the German and Vietnamese Translations. En *World Library and Information Congress: 69th IFLA General Conference and Council 1-9 August 2003, Berlin*. Recuperado de <http://www.ifla.org/IV/ifla69/papers/123e-Beall.pdf>
- Beghtol, Clare (1986). Semantic validity: concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, 30(2), 109-123.
- Beghtol, Clare (2002a). Universal concepts, cultural warrant and cultural hospitality. En López-Huertas, M.J., editor. *Proceedings of the 7th International ISKO Conference : 10-13 July 2002, Granada*. Würzburg: Ergon Verlag. p. 45-49.
- Beghtol, Clare (2002b). A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, 58(5), 507-532.
- Benyaich Benyaich, Sokaina (2014). *Estudio terminológico de la Mudawana: Código de Familia marroquí: Árabe-español*. Tesis de maestría dirigida por Mohamed El Madkouri. Alcalá de Henares: Universidad de Alcalá. Recuperado de <https://ebuah.uah.es/dspace/bitstream/handle/10017/23696/TFM%20SOKAINA%20BENYAICH.pdf?sequence=1>
- Bertrand-Gastaldy, S. (1993) Analyse documentaire et intertextualité. En *Les Sciences du texte juridique: Le droit saisi par l'ordinateur*. Sous la direction de Claude Thomasset, René Côté et Danièle Bourcier. Cowansville: Les Éditions Yvon Blais. p. 139-173. Recuperado de <http://www.ling.uqam.ca/sato/publications/bibliographie/Juri92.htm>
- Bocato, Vera Regina Casari y Biscalchin, Ricardo (2014). As dimensões culturais no contexto da construção de vocabulários controlados multilíngues. *Revista Interamericana de Bibliotecologia*, 37(3): 237-250.

- Bullard, Julia (2017). Warrant as a Means to Study Classification System Design. *Journal of Documentation*, 73(1): 75-90.
- Caro, C. & San Segundo, R (1999). Lenguajes documentales y exclusión social. En *La Representación y la Organización del Conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información (Actas del IV Congreso ISKO-España EOCONSID '99; Granada, 22 al 24 de abril de 1999)*. Granada: ISKO España. p. 101-108.
- Choi, Inkyung (2018). *Toward a model of intercultural warrant: a case of the korean decimal classification's cross-cultural adaptation of the Dewey Decimal Classification*. A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Information Studies. Milwaukee: University of Wisconsin.
- Colombo, Stephanie (2018). *El término bias en la Organización del Conocimiento: Un aporte histórico-conceptual* (Tesis de maestría). Montevideo: Facultad de Información y Comunicación.
- Colombo, Stephanie y Barité, Mario (2015). Tres enfoques de *bias* en Organización del Conocimiento: *Bias* neutro, *bias* negativo y *bias* positivo. *Brazilian Journal of Information Studies*, 10(2): 9-13.
- García Aguilar, I. & Villén Rueda, L. (2000). La construcción de espacios de diálogo multidisciplinario para el estudio y la salvaguarda del patrimonio documental en el entorno iberoamericano. En *V Encuentro de EDIBCIC. Actas*. Granada: Universidad de Granada. p. 292-305.
- González Casanova, Pablo (1996). Clasificaciones y definiciones: nota para un bibliotecario. *Investigación Bibliotecológica*, 10(20): 3-8.
- Guimarães, Jose Augusto Chaves (2006). Aspectos éticos em organização e representação do conhecimento (ORC): uma reflexão preliminar. En González de Gómez, M.N. & Orrico, E.G.D., organizadoras. *Políticas de memória e informação: reflexos na organização do conhecimento*. Natal: EDUFRRN. p. 237-264.
- Guimarães, Jose Augusto Chaves et al (2008). Ethics in the Knowledge Organization Environment: an overview of values and problems in the LIS literature. En Arsenault, C.; Tennis, J.T., editors. *Culture and Identity in Knowledge Organization*. Würzburg: Ergon. p. 340-346. (Advances in Knowledge Organization, 11).
- Guimarães, Jose Augusto Chaves & Pinho, Fabio A. (2007). Desafios da representação do conhecimento: abordagem ética. *Informação&Informação*, 12(1).
- Hjørland, Birger (2002). Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of documentation*, 58(4), 422–462.
- Hjørland, Birger (2004). Domain analysis: a socio-cognitive orientation for information science research. *Bulletin of the American Society for Information Science and Technology*, 30(3), 17-21
- Hsieh H.-F. & Shannon S. (2005) Three approaches to qualitative content analysis. *Qualitative Health Research* 15(9), 1277-1288.

- Krippendorff, Klaus. (2004) *Content Analysis: An Introduction to its Methodology*. 2nd ed. Thousand Oaks: Sage Publications.
- Lee, J.M. (1976). E. Wyndham Hulme: a reconsideration. En Rayward, W.B., editor. *The variety of Librarianship: essays in honour of John Wallace Metcalfe*. Sydney: LAA.
- Olson, Hope A. & Ward, D. (1998). Charting a journey across knowledge domains: feminism in the Decimal Dewey Classification. En *Advances in Knowledge Organization*, vol 6. Würzburg: Ergon Verlag. p. 238-244.
- Pacey, P. (1989). The Classification of Literature in the Dewey Decimal Classification: The Primacy of Language and the Taint of Colonialism. *Cataloging & Classification Quarterly*, 9(4), 101-107.
- Rodríguez, Robert D. (1984). Hulme's Concept of Literary Warrant. *Cataloging & Classification Quarterly*, 5(1): 17-26.
- Van Dijk, Teun A. (1999). El análisis crítico del discurso. *Anthropos* (186), p. 23-36 (septiembre-octubre 1999). Recuperado de <http://www.discursos.org/oldarticles/El%20an%20l%20cr%20del%20discurso.pdf>

A contribuição da organização do conhecimento no procedimento de classificação e indexação e nos processos crime com perspectiva de gênero: interpelações acerca dos feminicídios no Estado de São Paulo - Brasil

Denise Cristina Belan Fioravanti¹, Francisco Arrais Nascimento² y Deise Maria Antonio Sabbag³

¹ ORCID [0000-0002-0333-1601](https://orcid.org/0000-0002-0333-1601). Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, São Paulo, Brasil. denisebfioravanti@gmail.com

² ORCID [0000-0003-4424-8844](https://orcid.org/0000-0003-4424-8844). Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, São Paulo, Brasil. francisco.arrais.nascimento@gmail.com

³ ORCID [0000-0001-6392-4719](https://orcid.org/0000-0001-6392-4719). Universidade de São Paulo (USP/FFCLRP), Campus de Ribeirão Preto; Pós-Graduação em Ciência da Informação (UNESP/Campus de Marília); Vice Coordenadora ISKO/BRASIL. deisemarian@gmail.com

Resumo. Objetivou-se analisar a forma de indexação e classificação nos processos crime com Perspectiva de Gênero alocado enquanto Mortes Violentas de Mulheres – Feminicídios no Estado de São Paulo. Estudo descritivo ancora-se em uma Análise de Domínio (Tennis, 2002; 2003; Hjørland, 1995; 1997; 1998; 2002; 2003 e 2005), alicerçado em cartografia de documentos orientada sob as linhas de Rolnik (1989). Observou-se que existe uma demanda social que emerge dos movimentos sociais tanto Feministas, LGBTQI+ quando de Direitos Humanos por uma equidade de direitos que vem auferindo largo espaço nos debates acadêmicos e que necessitam de dados confiáveis para que possam mensurar com verossimilhança o cenário nacional para que assim se possa impulsionar, criar e acompanhar políticas públicas protetivas dos direitos das mulheres. Tal ação é freada pelas “imprecisões”, “distorções” e “inverossimilhanças” presentes no vocabulário dos processos crime, que dificultam a recuperação da informação.

Palavras-chave: Análise de Domínio; Classificação; Mortes Violentas de Mulheres; Feminicídio; São Paulo.

Resumen. El objetivo de este estudio fue analizar la indexación y clasificación de casos criminales con perspectiva de género asignados como muertes violentas de mujeres - Femicidios en el estado de São Paulo. Un estudio

descriptivo se basa en un análisis de dominio (Tennis, 2002; 2003; Hjørland, 1995; 1997; 1998; 2002; 2003 y 2005), basado en la cartografía de documentos orientada en la línea de Rolnik (1989). Se ha observado que existe una demanda social que surge de los movimientos sociales feministas, LGBTQI + y de derechos humanos por una equidad de derechos que ha ganado un amplio espacio en los debates académicos y que necesitan datos confiables para poder medir el escenario nacional con probabilidad. De esa manera, se puede promover, crear y monitorear la protección de las políticas públicas de derechos de las mujeres. Dicha acción se ve limitada por las "inexactitudes", "distorsiones" e "improbabilidad" presentes en el vocabulario del proceso penal, lo que dificulta la recuperación de información.

Palabras clave: Análisis de dominio; Clasificación; Muertes violentas de mujeres; Femicidio São Paulo.

Abstract. The objective of this study was to analyze the indexation and classification of criminal cases with Gender Perspective allocated as Violent Deaths of Women - Femicides in the State of São Paulo. A descriptive study is anchored in a Domain Analysis (Tennis, 2002; 2003; Hjørland, 1995; 1997; 1998; 2002; 2003 and 2005), based on document cartography oriented along the lines of Rolnik (1989). It has been noted that there is a social demand that emerges from both feminist, LGBTQI + and human rights social movements for an equity of rights that has been gaining ample space in academic debates and that they need reliable data so that they can measure the national scenario with likelihood. That way, it can be promoted, created and monitored the protection of women rights public policies. Such action is curbed by the "inaccuracies", "distortions" and "improbability" present in the criminal proceedings vocabulary, which make it difficult to retrieve information.

Keywords: Domain Analysis; Ranking; Violent Deaths of Women; Femicide; São Paulo

1 Introdução

Ao imergir no contexto dos Estudos de Gênero (EG), sob a égide da Violência de Gênero (VG), se pode vislumbrar toda uma construção social amparada em diversas formas de violência que compreendem agressões de caráter físico, psicológico, sexual e patrimonial que ocorrem em um continuum que pode culminar com a morte por homicídio (MENEGHEL; PORTELLA, 2017).

O Brasil ocupa posição de destaque no cenário mundial, sendo alocado na 5ª posição no ranking da violência de Gênero. Segundo dados do Mapa da Violência no Brasil as taxas apresentam-se em curva ascendente de mortes femininas ocasionadas por agressão no período de 1980-2010 e coeficientes de mortalidade que passaram de 2,3/100.000 para 4,8/100.000, representando um aumento de 111% no período (MENEGHEL; PORTELLA, 2017). Diante disso, alicerçando-se em inquietações diante dos crimes cometidos com perspectiva de Gênero alocado enquanto Mortes Violentas de Mulheres – Feminicídios no Estado de São Paulo. Ressalta-se que o Estado não foi eleito aleatoriamente, mas sim por apresentar os maiores índices de

em uma zona onde a humanidade é desconstruída, os direitos são cerceados e a própria materialização do sujeito torna-se impraticável.

No âmbito da Ciência da Informação (CI), sob a óptica da Organização do Conhecimento (OC), aloca-se o argumento na natureza interdisciplinar da CI (LE COADIC, 2004; SARACEVIC, 1996). O relevo dado à área da CI, é devido ao esforço para enfrentar os problemas de organização, crescimento e disseminação do conhecimento registrado, que teve acentuado crescimento desde a invenção da imprensa apresentando desde então um crescimento exponencial. Compreende-se que os objetos da CI, fazem parte de um processo dinâmico de desenvolvimento, constituído de relações com outras áreas do conhecimento, como as ciências Humanas, Sociais e Exatas, o que estimula uma relação de interação entre áreas que propicia uma abordagem, que perpassa as mais diversas áreas com o objetivo de compreender, organizar, preservar e disseminar a informação.

Nesse sentido, nas subáreas da CI, a preservação é um aspecto central no cenário contemporâneo e em que se volta para a representação, classificação e a própria preservação da memória nas suas mais variadas vertentes. Aqui se aloca a importância de tal estudo, uma vez que existe uma escassez de pesquisas no tocante a indexação e classificação nos processos crime com Perspectiva de Gênero. Esse recorte é norteador pelas contribuições de Hjørland & Albrechtsen (1995) e Hjørland (2002) que segundo Pinho (2010, p.4) “[...] versa sobre o entendimento do objeto de estudo da CI – a informação registrada e socializada – por meio da análise dos domínios do conhecimento como um todo ou de comunidades discursivas, sendo como uma das abordagens os estudos terminológicos”. De tal modo, compreende-se que as iniciativas de organizar a informação para o público em geral considerando a classificação adequada para que sejam devidamente recuperadas, de modo a refletir o universo idiossincrático da Violência de Gênero. Esses estudos terminológicos para o domínio Violência de Gênero consideram as garantias literária, cultural e de uso (BEGHTOL, 1986; 2002; 2005), respeitando uma ética transcultural de mediação (GARCÍA GUTIÉRREZ, 2002) e impondo limites a um “poder de nomear” do indexador (OLSON, 2002) em seus eventuais preconceitos e antipatias (BERMAN, 1993).

Deste modo, sob a óptica de sua conceituação e prática em Borko (1968), alinhado com o referencial da OC, sob as linhas de Hjørland (1995, 1997) que expõe em sua teoria, afirmando primeiramente, que a unidade de análise da CI é formada pelos campos coletivos de conhecimento ou domínios de conhecimento concernentes às suas comunidades discursivas. Estas comunidades não são entidades autônomas, mas distintas construções sociais compreendidas por indivíduos sincronizados em pensamento, linguagem e conhecimento, e constituintes da sociedade moderna. Naturalmente, concatenadas às dimensões culturais e sociais.

Assim, propôs-se adentrar ao campo da CI, com o auxílio da OC sob o viés da classificação e da Indexação, ancorando-se na terminologia e na terminografia, onde aqui segundo as linhas de Sonneveld (1993) conceitua de forma ampliada a terminologia como sendo qualquer atividade relacionada com a sistematização e representação de conceitos ou apresentação de termos baseados em princípios e

métodos estabelecidos. Para tanto, torna-se claro o direcionamento da problemática a qual norteia este estudo, a saber: Como são classificados e indexados os crimes nos processos crime contra a mulher no Estado de São Paulo? Diante disso, objetivou-se analisar a forma de indexação e classificação nos processos crime com Perspectiva de Gênero alocado enquanto Mortes Violentas de Mulheres – Feminicídios no Estado de São Paulo.

2 Metodologia

A pesquisa descritiva ancora-se em uma Análise de Domínio (Tennis, 2002; 2003; Hjørland, 1995; 1997; 1998; 2002; 2003 e 2005), alicerçada em cartografia de documentos orientada sob as linhas de Rolnik (1989). A pesquisa em questão seguiu o seguinte roteiro:

1. Identificação dos crimes que serviram de amostra para a pesquisa, sendo estes selecionados segundo:

I. O território geográfico marcadamente delimitado no Estado de São Paulo;

II. O recorte cronológico compreendido 2013-2018. Ressalta-se que nos anos de 2013 e 2014 anos que antecedem a Lei n.º 13.104/2015, não foram recuperados documentos com o termo “Feminicídio”.

2. Análise dos termos e designações criminais empregados nos processos;

3. Identificar as “distorções” e “inverossimilhanças” presentes no vocabulário dos processos crime;

4. Identificar e analisar os motivos pelos quais tais possíveis “imprecisões”, “distorções” e “inverossimilhanças” presentes no vocabulário dos processos crime surgem podendo assim contribuir para a sua minimização ou extinção.

Em uma busca realizada no Portal de Serviços do Tribunal de Justiça do Estado de São Paulo, tomando como recorte geológico a primeiro e o último processo indexação com termo Feminicídio, chegou-se ao corpus de 357 documentos.

3 Análises e Resultados

Para Diniz, Costa e Gumeri (2015, p. 225-239) “Nomear é um gesto político, uma forma inicial de apreensão de experiências ou fenômenos escondidos. Esta pode ser a história política do conceito de feminicídio – nomear para fazer existir a morte evitável de mulheres”, para os autores existe uma necessidade por parte das autoridades competentes de organização nas mortes violentas praticadas contra mulheres, para isso a necessidade de nomear devidamente e evitar distorções que prejudiquem a busca e recuperação da informação (Figura 3).

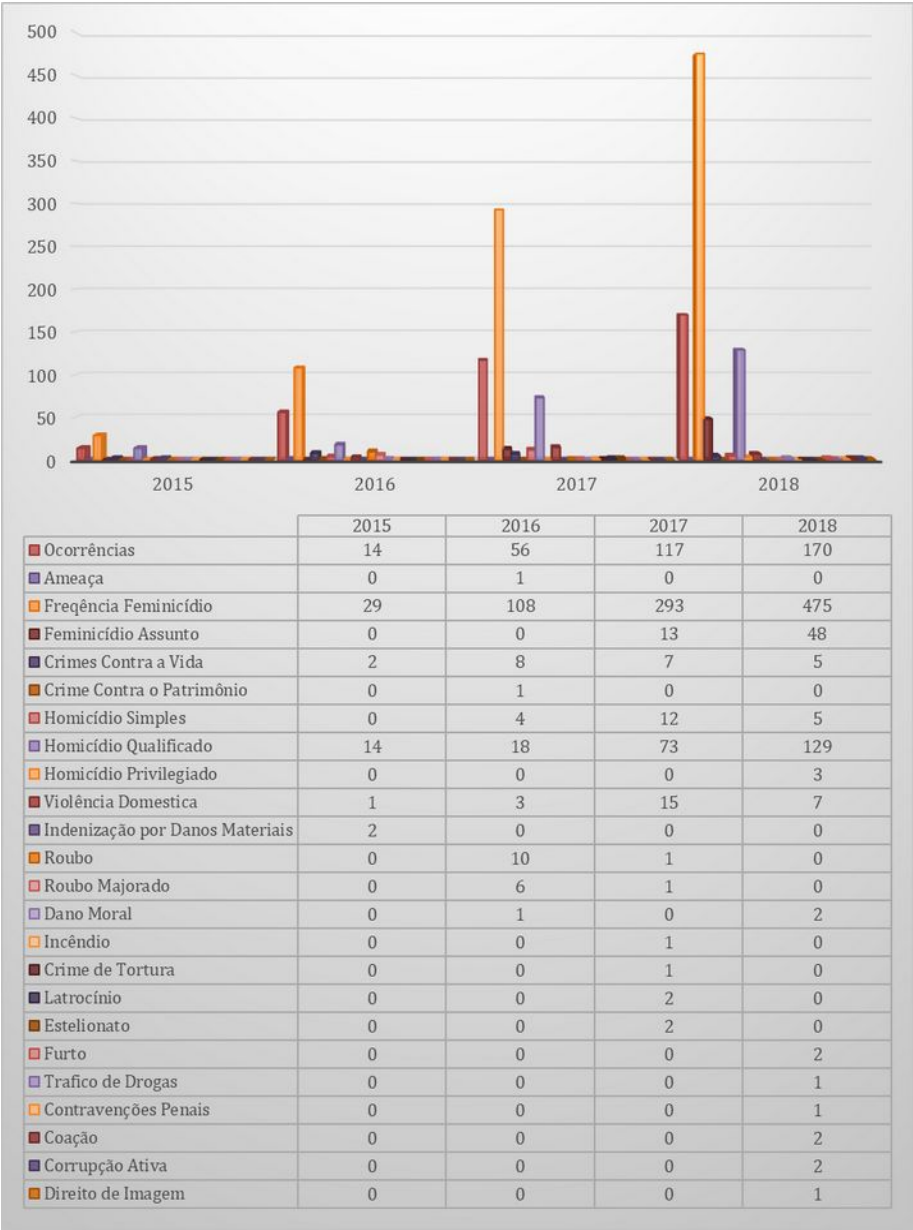


Figure 2. Representação do Corpus segundo a Classificação do Crime.

Para pesquisadores que buscam dados relacionados ao tema, a busca por resultados e estatísticas são de extrema dificuldade, isso ocorre pois existe uma grande difusão na classificação e indexação, atrapalhando a recuperação dos casos. Grande parte dos crimes que resultam em morte são classificados e indexados como Homicídio

Qualificado e não como Feminicídio, tirando a visibilidade que o processo deveria ter, dispersando dados e tornando a recuperação da informação difícil. Tal afirmação pode ser visualizada ao analisar os dados condensados na Figura 2. Em 357 documentos analisados, o termo feminicídio utilizado como buscador, apresenta uma frequência crescente, no entanto só passa a ser assunto no ano de 2017, dois anos após a criação da aplicação Lei n.º 13.104/2015, termos outros são utilizados com frequência para designar crimes que provavelmente podem ser enquadrados como Feminicídios, no entanto sofrem dispersões pela abertura legislativa como forma de diminuição de pena, preconceitos e antipatias geradas por uma sociedades marcadamente machista, além de “distorções” e “inverossimilhanças” presentes no vocabulário dos processos crime, que dificultam a recuperação da informação.

4 Considerações Finais

Observou-se que existe uma demanda social que emerge dos movimentos sociais tanto Feministas, LGBTQI+ quando de Direitos Humanos por uma equidade de direitos que vem auferindo largo espaço nos debates acadêmicos e que necessitam de dados confiáveis para que possam mensurar com verossimilhança o cenário nacional para que assim se possa impulsionar, criar e acompanhar políticas públicas protetivas dos direitos das mulheres. Tal ação é freada pelas “imprecisões”, “distorções” e “inverossimilhanças” presentes no vocabulário dos processos crime, que dificultam a recuperação da informação. Ressalta-se que, os dados apresentados são na verdade esforços individuais de instituições independentes e/ou órgãos internacionais que visam denunciar o que vem acontecendo, mas que devido as divergências de classificação, indexação e representação da informação, tais esforços em recuperar e representar a informação de forma a disseminar a mesma se toram impossíveis, inverossimilhantes e não exatos.

Referencias

- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113.
- Beghtol, C. (2005). Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology*, 56(9), 903-912.
- Beghtol, C. (2002). Universal concepts, cultural warrant, and cultural hospitality. In: LÓPEZ-HUERTAS, M. J. (Ed.). *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Würzburg: ERGON-Verlag, 45-49.
- Berman, S. (1993). *Prejudice and antipathies: a tract on the LC subject heads concerning people*. Jefferson: McFarland & Company Inc. Publishers.

- Borko, H. (1968). Information Science: What is it? *American Documentation*, 19(1), 3-5.
- Diniz, D.; Costa, B. S. & Gumieri, S. (2015). Nomear feminicídio: conhecer, simbolizar e punir. *Revista Brasileira de Ciências Criminais*, 114/2015, 225 – 239.
- García Gutiérrez, A. (2002). Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In: López-Huertas, M. J. (Ed.). *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries*. Würzburg: ERGON-Verlag, p. 516-522.
- Hjørland, B. (2002). Domain analysis in information science. Eleven approaches — traditional as well as innovative. *Journal of Documentation*, 58(4), 422-462.
- Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2), 87-111.
- Hjørland, B. (1997). *Information seeking and subject representation: an activity theoretical approach to Information Science*. London: Greenwood Press.
- Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Le Coadic, Y. (2004). *A ciência da informação*. 2. ed. Brasília: Briquet de Lemos.
- Meneghel, S. N.; Portella, A. P. (2017). Feminicídios: conceitos, tipos e cenários. *Ciênc. saúde coletiva*, Rio de Janeiro, 22(9), 3077-3086.
- Olson, H. A. (2002). *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic Publisher.
- Prins, B., & Meijer, I. C. (2002). Como os corpos se tornam matéria: entrevista com Judith Butler. *Revista Estudos Feministas*, 10(1), 155-167.
- Pinho, F. A. (2010). *Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina: uma análise da precisão em linguagens de indexação brasileiras*. 2010. 149f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília.
- Rolnik, S. (1989). *Cartografia Sentimental: Transformações contemporâneas do desejo*. Editora Estação Liberdade, São Paulo.
- Saracevic, T. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, 1 (1), 41-62.
- Sonneveld, H. B. et al. (1993). *Terminology: applications in interdisciplinary communication*. Amsterdam, Philadelphia: J. Benjamins.
- Tennis, J. T. (2002). Subject ontogeny: Subject access through time and the dimensionality of classification. In: López-Huertas, M. J. (Ed.). *Challenges in knowledge representation and organization for the 21st Century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference, Granada (Spain), 2002*. Würzburg: Ergon. 8, 54-59.
- Tennis, J. T. (2003). Two axes of domains for domain analysis. *Knowledge Organization*, 30, nos. 3-4: 191-195.

O assassinato de Marielle Franco e os algoritmos racistas: dimensões aplicadas da teoria crítica da organização do conhecimento

Gustavo Silva Saldanha¹, Franciéle Carneiro Garcês da Silva², Graziela Santos Lima³, Dirnéle Carneiro Garcês⁴ y Natália Lima Romeiro⁵

¹ ORCID [0000-0002-7679-8552](https://orcid.org/0000-0002-7679-8552). IBICT ; UNIRIO, Rio de Janeiro, Brasil.
gustavosalldanha@ibict.br

² ORCID [0000-0002-2828-416X](https://orcid.org/0000-0002-2828-416X). PPGCI UNESP, São Paulo, Brasil.
graziela.dsl@gmail.com

³ ORCID [0000-0003-3861-2937](https://orcid.org/0000-0003-3861-2937). PPGCI IBICT UFRJ, Rio de Janeiro, Brasil.
francigarces@yahoo.com.br

⁴ ORCID [0000-0002-3061-9352](https://orcid.org/0000-0002-3061-9352). UFSC, PPGCIN Florianópolis, Brasil.
dirnele.garcez@yahoo.com.br

⁵ ORCID [0000-0002-6274-4836](https://orcid.org/0000-0002-6274-4836). PPGCI IBICT UFRJ, Rio de Janeiro, Brasil.
ntromeiro91@gmail.com

Resumen. Com foco em uma teoria crítica da organização do conhecimento em construção, desdobrando os estudos como os de Martínez-Ávila, Daniel; Semidao, R.; Ferreira, M. (2016) e Olson (2006), o foco empírico da pesquisa está ligado aos comentários produzidos após o assassinato da vereadora do município do Rio de Janeiro, Marielle Franco, e do motorista Anderson Gomes, em março de 2018. Nesse decurso, sob a via do que temos trabalhado sob a noção de organização ordinária dos saberes socialmente oprimidos (O²S²O), procuramos aprofundar os estudos já realizados (Saldanha, Silva, Lima, Garcês, Romeiro, 2018) no território do corpus metainformacional produzido sobre o crime, com foco na presença política feminina de Marielle.

Palabras clave: Teoria crítica da organização do conhecimento. Marielle Franco – Brasil. Política - Brasil. Algoritmo racista.

Abstract. Focused on a critical theory of knowledge organization under construction, unfolding studies such as those by Martínez-Ávila, Daniel; Semidao, R. ; Ferreira, M. (2016) and Olson (2006), the empirical focus of the research is linked to the comments produced after the murder of city councilor in Rio de Janeiro, Marielle Franco, and driver Anderson Gomes, in March 2018. In this in the course of what we have been working under the notion of ordinary organization of socially oppressed knowledge (O²S²O), we seek to deepen the studies already carried out (Saldanha, Silva, Lima, Garcês, Romeiro, 2018) in the territory of the produced meta-informational corpus on crime, focusing on Marielle's female political presence.

Keywords: Critical theory of knowledge organization. Marielle Franco - Brazil. Politics - Brazil. Racist algorithm.

1 Introdução: algoritmos, racismo e a teoria crítica da organização do conhecimento

Conforme o pesquisador francês Ertzcheid (2018), a construção do movimento social *Les Gilets Jaunes* na França, a partir de 2018, tem uma relação fortíssima com o Facebook. O olhar sobre essa produção digital paralela à atuação do movimento nas ruas demonstra que o vocabulário utilizado nas páginas da rede social aponta para a presença de distintos dilemas sociais. Em outros termos, para se posicionar contra o movimento, um grupo social se utiliza da barbárie do racismo de classe como argumento. O pesquisador aponta especificamente para o estudo preliminar dos colegas franceses Brigitte Sebbah, Natacha Souillard, Laurent Thiong-Kay, Nikos Smyrniaos (2018). A pesquisa partiu de uma revisão da produção de comentários e de redes sociais sobre o movimento social.

Com foco em uma teoria crítica da organização do conhecimento em construção, desdobrando os estudos como os de Martínez-Ávila, Daniel; Semidao, R.; Ferreira, M. (2016) e Olson (2006), o foco empírico da pesquisa está ligado aos comentários produzidos após o assassinado da vereadora do município do Rio de Janeiro, Marielle Franco, e do motorista Anderson Gomes, em março de 2018. Nesse decurso, sob a via do que temos trabalhado sob a noção de organização ordinária dos saberes socialmente oprimidos (O²S²O), procuramos aprofundar os estudos já realizados (Saldanha, Silva, Lima, Garcês, Romeiro, 2018) no território do corpus metainformacional produzido sobre o crime, com foco na presença política feminina de Marielle.

A pesquisa dialoga com o questionamento do poder da linguagem e do papel da organização do conhecimento na produção contemporânea dos vocabulários do mal. Assim como Capurro (2019), a aproximação ao pensamento da filósofa Hannah Arendt (1999, 2007), no tocante à reflexão sobre o mal em expansão na sociedade e suas formas (no nosso caso, estruturalmente oriundas da e orientadas para a linguagem) de banalização. Para essa etapa atual da pesquisa, discutimos a condição da cidadania na globalidade mundial digital e o dilema (também ele essencialmente linguístico) dos algoritmos, bem como o poder de um *ethos* robótico que interessa sobremaneira a reflexão crítica da organização do conhecimento.

2 Os algoritmos racistas e a ética dos robôs

A continuación se ofrecen ejemplos de presentación de tablas y de figuras e ilustraciones. Todas ellas deben ir numeradas secuencialmente. “La fonction du chercheur est alors de dévoiler les mécanismes d’invisibilisation de ces contraintes – dont l’exemple principal est la rhétorique de « l’algorithme ».” (Casilli, 2017)

Como lembra Capurro (2017), em sua discussão sobre a cidadania global digital, o mundo da *web*, mesmo atravessando fronteiras nacionais, não é um isolado do mundo

social amplo, mas parte dele. Diante disso, a cidadania global digital inverte as ideias do Iluminismo sobre a condição cidadã. O perigo da homogeneização da população mundial não está apenas em seu controle e manipulação, mas na exclusão de diferentes grupos e no desrespeito às diferenças culturais, histórias individuais e contingências que são a base da singularidade e da riqueza dos indivíduos e sociedades humanas. A distopia oposta a tal é o mútuo isolamento político, econômico ou cultural de indivíduos e sociedades, bem como seu desrespeito a qualquer tipo de responsabilidade pelo bem-estar comum e pela sustentabilidade do mundo físico e digital.

Essa homogeneização é dada cada vez mais pela ação e pelo poder em expansão dos algoritmos. A vida social é cada vez mais governada por algoritmos. Mas o que é um algoritmo?, indaga Capurro (2019). Trata-se na visão do filósofo de uma ferramenta digital para ajudar a encontrar soluções para problemas. As empresas de tecnologia da informação criaram poderosos algoritmos que permitem pesquisas personalizadas, criando perfis individuais e sociais que são uma base não apenas para a economia digital, mas também para processos políticos e sociais, local e globalmente. Os algoritmos, lembra Capurro (2019), estão no centro de todos os tipos de processos sociais, técnicas industriais e rotinas cotidianas. É nesse contexto que nos perguntamos sobre o papel dos algoritmos e do mal, elemento que será verificado pelas lentes da organização do conhecimento em corpora disponíveis na web, como o caso Marielle Franco e dos Gilets Jaunes.

No plano da banalização do mal apontada por Arendt (1999), podemos identificar a relevância dos questionamentos de Ertzcheid (2017). Ao se perguntar se “são os algoritmos racistas”, o pesquisador francês não tem dúvidas: os motores de busca online apresentam os piores e mais cruéis estereótipos racistas e de incitação ao ódio. O autor nos lembra que em 2016 a empresa Google completou seus 18 anos, o que inspirou uma reflexão sobre a “maturidade” da instituição, dado seu poder no mundo. Estaria uma empresa com tamanha influência moral no contexto internacional capaz de responder pelos seus atos? O contexto ético em crise na *web* e os riscos de uma moral robótica (ressuscitando os mitos da neutralidade das ferramentas lógicas) estão em cena.

Ertzcheid (2017) dá como exemplo, agora no cerne dos questionamentos arendtiano, uma busca como “o holocausto realmente existiu?” na plataforma Google. O resultado apresenta inúmeros resultados, fundados, no entanto, em um discurso oriundo de páginas neonazistas. Esta é uma resposta lógica, que parte de uma mínima condição de compreensão política. Porém o sistema “não é capaz” de afirmar isso, dado que seu compromisso (e sua pretensa “imaturidade”) tem como foco o lucro, e não a discussão. A questão para o autor está ligada a um princípio de “viés cognitivo” (*bias cognitive*) transposto para o viés algorítmico (*bias algorithmiques*).

Quando formulamos uma pergunta ao Google, estamos já em um território mental projetando uma resposta que desejaríamos confirmar (através de um sim ou não, e não de uma busca por compreensão). O estudo empírico de Ertzcheid (2017) demonstra que entre Google e Wikipédia avançamos para o abismo da barbárie através da linguagem. O poder dos algoritmos é influenciado, pois, pela ambição das

plataformas, independentemente do mal que circula no conteúdo. A conclusão do autor é que estamos desamparados diante dos desvios dos algoritmos.

Sabemos que os algoritmos não são racistas enquanto formas autônomas; racistas são seus criadores e usuários; e é, pois, por essa razão, que racistas se constituem os algoritmos, sendo a *web* um território de conhecimento tomado por contextos de manipulação, via linguagem, de estruturas de poder e de formas de propagação do mal.

3 Marielle Franco: da luta de uma mulher à barbárie

“Lugar de mulher é onde ela quiser.” (Franco, 2016)

Marielle Francisco da Silva (mais conhecida como Marielle Franco) aponta, no parágrafo final dos agradecimentos de sua dissertação, sobre o fato de ser uma mulher negra oriunda da favela da Maré, localizada no Rio de Janeiro. Intitulada “UPP – A Redução da Favela a Três Letras: uma Análise da Política de Segurança Pública do Estado do Rio de Janeiro”, a dissertação de Marielle Franco, defendida em 2014, no Programa de Pós-Graduação em Administração da Faculdade de Administração, Ciências Contábeis e Turismo da Universidade Federal Fluminense, busca demonstrar “as Unidades de Polícia Pacificadora (UPP), enquanto política de segurança pública adotada no estado do Rio de Janeiro, reforçam o modelo de Estado Penal” (FRANCO, 2014, p. 11).

Marielle Franco trabalhou dez anos como assessora parlamentar, e no ano de 2016, foi eleita como vereadora pelo Partido Socialismo e Liberdade (PSOL), pela coligação “Mudar é possível”. Ela foi eleita com 46.502 votos, sendo a quinta mais votada na cidade do Rio de Janeiro. A vereadora conquistou não só votos de pessoas oriundas das classes menos privilegiadas, mas também obteve mais de 250 votos de acadêmicos e professores (SARAIVA, 2018).

No dia 14 de março de 2018, aos 38 anos Marielle Franco foi executada! Seu veículo foi atingido por 13 tiros. Ela e o motorista, Anderson Pedro Gomes, morreram no local dos disparos. Em entrevista a um telejornal, a companheira de Marielle, Monica Tereza Benício, contou sobre os doze anos de relacionamento, incluindo o preconceito em relação à vivência homoafetiva, bem como da carreira da vereadora e de seu ativismo por diversas causas, entre elas, a defesa dos direitos LGBTQI+.

(Saldanha, Silva, Lima, Garcês, Romeiro, 2018)

4 O assassinato de Marielle Franco e o vocabulário do ódio em marcha acelerada

An ethics of algorithms deals with making this difference theoretically and practically between who and what we are by resisting the tendency to confuse or even to identify ourselves

(our selves) with masks that we give to ourselves or others give to us. (Capurro, 2019)

A partir do contexto do assassinato, direcionamos nosso olhar científico para a produção discursiva sobre Marielle Franco. A coleta de dados cumpriu as seguintes etapas: a) busca da Página do Portal G1 - O Portal de Notícias das Organizações Globo no *Facebook*; b) procura de Notícias de Marielle Franco; c) seleção de notícia do Domingo, após apresentação da reportagem sobre Marielle e familiares no jornal dominical; d) seleção a notícia do dia 18 de março de 2018, publicada às 22:03 minutos (após a reportagem ter sido apresentada no Jornal). Ao total, esta notícia obteve mais 11 mil reações e 809 compartilhamentos até o dia 23/03/2018, às 23 horas e 54 minutos. A partir dela, foram selecionados os comentários mais relevantes (critérios de relevância do algoritmo do *Facebook* no período de coleta a partir de perfil pessoal), segundo o próprio *Facebook*. Por fim, foram mapeados 406 (quatrocentos e seis) comentários. (Saldanha, Silva, Lima, Garcês, Romeiro, 2018)

O percurso nos permitiu demonstrar o panorama crítico das abordagens *folk* e a necessidade de construção de um olhar fundado na revisão constantes dos posicionamentos éticos de construção de linguagens e de sua reprodução na e para *web*. Trata-se de perceber como se efetiva um vocabulário que, dentro de um discurso de natureza transparente da ciência, seria tomado como alheio e inofensivo, mas que se estabelece justamente na contramão das construções sociais e do papel democrático das lutas políticas. (Saldanha, Silva, Lima, Garcês, Romeiro, 2018)

Para o desenvolvimento da pesquisa, as categorias crítico-discursivas adotadas como lentes para reconhecer e problematizar o desenvolvimento do discurso do mal no caso do assassinato do crime, com ênfase do caráter político do fato, sendo este o motivo da centralidade da mesma na pesquisa. São elas: a metáfora, a ordinariade, a desclassificação, a luta “das” classes.

O conceito de metáfora se torna a centralidade de uma perspectiva sociocultural para a compreensão dos saberes que transversalizam a sociedade, a partir dos estudos sobre significantes e significados da teoria barroca de Emanuele Tesauro (1670), atendendo-se para, na atualidade, os estudos metodológicos e teóricos, por exemplo, em Evelyn Orrico (2001) e Fábio Pinho (2014).

A ordinariade: pelo léxico wittgensteiniano - o “segundo” Wittgenstein (1979) - a via de uma língua ordinária responde pelo reconhecimento da construção social da realidade pela linguagem tecida em cada comunidade falante, em diálogo com os trabalhos iniciais de Bernd Frohmann (1990) sobre a indexação o potencial de percepção de concepções de mundo distintas sendo tecidas a partir da linguagem ordinária.

A desclassificação: sob o desenvolvimento de um olhar desconstrucionista de García Gutiérrez (2011), encontramos a dinâmica de problematização crítica dos atos classificatórios; retomando uma teoria logológica (que toma o discurso como centralidade), com foco nos tropos, na percepção as condições dramáticas de constituição das classificações nas extremidades das fraturas sociais.

A luta “das” classes: sob uma configuração dialética da análise crítica do desenvolvimento da OC, a luta das classes remonta um marxismo de fundo linguístico, que encontra Bakhtin sob a via da filosofia da linguagem, Pierre Bourdieu (2008) sob a via sociológica, e Robert Estivals (1978, 1990), sob a perspectiva do marxismo esquematista. Essa luta é reconhecida como o conjunto de forças que interagem, na linguagem, para a manutenção das práticas de opressão e a anulação das diferenças, representadas como anomalias da vida social. Essas condicionantes materiais da existência são multiplicadas na linguagem – ou seja, nas classes que isolam e extinguem nomes, coisas e pessoas nas classificações –, constituindo muralhas simbólicas mais eficazes do que a própria demarcação das fronteiras entre classes.

A arbitrariedade: que consiste na escolha “natural” de todo sistema de organização do conhecimento, ou seja, todo sistema é arbitrário na medida que estabelece uma estrutura hierarquizada que reproduz a vida social em suas distinções e as reforça. No caso da teoria dos tesouros, tal condição aparece, por exemplo, tanto na escolha e exclusão de termos, como nas modalidades de estruturação dos termos gerais e termos específicos. (saldanaha et al, 2018)

A produção dos comentários no corpus coletado revela a intensa construção do discurso de ódio, a produção de notícias falsas sobre a trajetória de vida e a atuação política de Marielle Franco, o ataque às diferenças políticas, a profusão de injúrias contras diferentes classes e movimentos sociais (a luta do movimento negro, dos movimentos de descriminalização das drogas, a luta pela perspectiva social de desenvolvimento governamental), sintetizados na tentativa de desconstrução da pauta da Declaração dos Direitos Humanos (ONU, 1949). A corpus demonstra ainda o papel de robôs na reprodução massiva de desinformação, junto da desconstrução de quaisquer preceitos da cidadania (dentro ou fora da *web*).

Os efeitos discutidos do corpus levam-nos ao posicionamento de três categorias fundacionais do papel político da OC via abordagem de uma leitura da ordinariade da organização dos saberes oprimidos: crítica, denúncia e memória. (Saldanha, Silva, Lima, Garcês, Romeiro, 2018). No entanto, igualmente, para além dessas categorias, o foco do desenvolvimento das teorias críticas em OC recai aqui sobre o dilema de uma cidadania global digital e o mal representado por e para os algoritmos, máquinas de reprodução dos caminhos que nos distanciam das condições mínimas da dignidade humana.

5 Considerações críticas: o longo e contínuo assassinato de Marielle Franco

“Algorithms know nothing about hesitation. In fact, they know nothing at all and they do not learn. They are heteronomous. They are not played by the world, but by human

designers and users. The question about who we are is about the being of the who.” (Capurro, 2019)

“Ela desatinou / Desatou nós” (Strassacapa, 2016)

A pesquisa, sob a construção teórica aberta de uma organização ordinária dos saberes socialmente oprimidos colocou em ênfase o papel de denúncia que a organização do conhecimento, através de suas teorias e métodos, pode desenvolver. Através do estudo qualitativo da produção do discurso de ódio sobre a vereadora assassinada Marielle Franco, percebemos o risco de um vocabulário do mal a caminho passível de influenciar e alterar linguagens documentárias, sob o viés da barbárie.

Junto ao pensamento capurriano (Capurro, 2017), procuramos ferramentas críticas, como a organização do conhecimento, para, na era digital, lutar por um ethos transcultural com componentes democráticos que promova ativamente a experiência intercultural e um tratado internacional para o mundo cibernético no qual, seguindo a proposta de Kant, as diferentes partes interessadas concordam livremente. As ferramentas da OC permitem a análise e a promoção de diferenças culturais que subjazem implícita ou explicitamente as regras alfandegárias e comportamentais no país. físico e no mundo digital.

O caso Marielle é emblemático neste sentido, dado que não bastasse a crueldade envolvida em todo o processo, percebe-se que seu assassinato é longo e contínuo, influenciado por uma robótica do mal e pela contínua manipulação algorítmica. O caso revela justamente a dinâmica trágica dos modos de produção, organização, representação do conhecimento na *web* e o papel da constituição das teorias críticas da organização do conhecimento.

Referencias

- Arendt, Hannah. (1999). *Eichmann em Jerusalém: um relato sobre a banalidade do mal*. São Paulo: Cia das Letras.
- Arendt, Hannah. (2007). *A condição Humana*. Rio de Janeiro: Forense universitária.
- Bakhtin, Mikhail. *Marxismo e filosofia da linguagem*. São Paulo: Hucitec, 2006.
- Bourdieu, Pierre. *A Economia das trocas linguísticas: o que falar quer dizer*. 2. ed. São Paulo: Edusp, 2008.
- Butler, Judith. Performatividad, precariedad y políticas sexuales. *AIBR*. Revista de Antropología Iberoamericana, v. 4, n. 3, 2009.
- Capurro, Rafael. (2019). Enculturating Algorithms. In: Proceedings of the International Conference: Artificial Intelligence & Regulation, LUISS, Rome.
- Capurro, Rafael. (2017). Citizenship in the Digital Age. In: Toni Samek and Lynette Shultz (eds.): *Information Ethics, Globalization and Citizenship*. Essays on Ideas to Praxis. Jefferson NC: McFarland. 11-39.

- Casilli, Antonio. (2017). Préface. In.: Ertzcheid, Olivier. *L'appétit des géants* : pouvoir des algorithmes, ambitions des plateformes. Paris: C&F éditions.
- Eco, Umberto. *Semiótica e filosofia da linguagem*. Lisboa: Instituto Piaget, 1984.
- Ertzcheid, Olivier. (2017). *L'appétit des géants* : pouvoir des algorithmes, ambitions des plateformes. Paris: C&F éditions.
- Ertzcheid, Olivier. (2018). #giletsjaunes : de l'algorithme des pauvres gens à l'internet des familles modestes. 30 novembre 2018. Disponible en : https://www.affordance.info/mon_weblog/2018/11/gilets-jaunes-internet-familles-modestes-facebook.html. Accès en: 02 jan. 2019.
- Estivals, Robert. Les langages et leurs interrelations: quelques axes pour une théorie sémiologique de la communication. *Revue de Bibliologie: schéma et schématisation*, n. 33, p. 8-16, 1990.
- Estivals, Robert. Luttas de classe et schématisation. *Schéma et schématisation*, n. 9, p. 5-10, 1978.
- Franco, Marielle. (2018). Aos bastardos da PUC-Rio, com carinho. In: ANTENORE, Armando. Aos bastardos da PUC-Rio, com carinho. *Folha de São Paulo*, 17 mar. 2018. Disponível em: <<http://piaui.folha.uol.com.br/aos-bastardos-da-puc-com-carinho/>>. Acesso em: 03 jun. 2018.
- Franco, Marielle. (2014). *UPP – A Redução da Favela a Três Letras*: uma Análise da Política de Segurança Pública do Estado do Rio de Janeiro. 2014. 136 p. Dissertação (Mestrado em Administração) – Universidade Federal Fluminense. Niterói, RJ.
- Frohmann, Bernd. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, v. 46, n. 2, jun. 1990.
- García Gutiérrez, Antonio. Desclassification in Knowledge Organization: a post-epistemological essay. *Transformação*, Campinas, v.23, n.1, p. 5-14, jan./abr., 2011.
- García Gutiérrez, Antonio. La organización del conocimiento desde la perspectiva poscolonial: itinerarios de la para consistencia. *Perspectivas em Ciência da Informação*, v.18, n.4, p. 93-111, out./dez. 2013.
- ONU. Declaração universal dos direitos humanos. Paris: ONU, 1949.
- Orrico, E.G.D. *Binômio Linguística - Ciência da Informação*: abordagem teórica para elaboração de metafiltro de recuperação da informação. 2001. 206 f. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia / Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.
- Martínez-Ávila, Daniel; Semidao, R.; Ferreira, M. (2016). Methodological Aspects of Critical Theories in Knowledge Organization. *Knowledge Organization*, v. 43, p. 118-125.

- Olson, H. A. (2006). Codes, costs, and critiques: The organization of information in *Library Quarterly*, 1931-2004. *Library Quarterly*. 76(1): 19-35.
- Pinho, Fabio A. Metafiltro para controle terminológico de metáforas no domínio da homossexualidade masculina. *Ci. Inf.*, Brasília, DF, v. 41 n. 1, p.120-133, jan./abr., 2014.
- Saldanha, G. S. Sobre a O²S²O, de Tesouro à Bourdieu: linguagem simbólica e a organização ordinária dos saberes socialmente oprimidos. In: Lucas, Elaine R. de O.; Silveira, Murilo A. A. da. *A Ciência da informação encontra Pierre Bourdieu*. Recife: Ed. Universitária da UFPE, 2017.
- Saldanha, Gustavo Silva; Silva, Franciéle Carneiro Garcês da; Lima, Graziela Santos; Garcês, Dirnéle Carneiro; Romeiro, Nathalia Lima. (2018). Quem matou Marielle? Organização do conhecimento e os caminhos do tesouro do mal. Enancib 2018. *Anais...* Londrina: UEL. p. 1-20.
- Saraiva, Jacqueline. (2018). Saiba quem era Marielle Franco, vereadora assassinada a tiros no Rio. *Em.com.br*. 15 mar. 2018. Disponível em: <https://www.em.com.br/app/noticia/politica/2018/03/15/interna_politica,944288/saiba-quem-era-marielle-franco-vereadora-assassinada-a-tiros-no-rio.shtml>. Acesso em: 01 jul. 2018.
- Sebbah, Brigitte; Souillard, Natacha; Thiong-Kay, Laurent; Smyrniaios, Nikos. (2018). *Les gilets jaunes, des cadrages médiatiques aux paroles citoyennes*. Rapport de recherche préliminaire. 26 novembre 2018. Disponible : <https://pt.scribd.com/document/394250648/Rapport-Gilets-Jaunes>. Accès en: 02 jan. 2019.
- Strassacapa, Juliana. (2016). Triste, louca ou má. *Francisco El Hombre* (Álbum).
- Tesouro, Emanuele. *Il canocchiale aristotelico*. Turim, 1670.
- Vignaux, Georges. *O demônio da classificação*. Lisboa: Instituto Piaget, 2000.

