

organizadores

Thiago Henrique Bragato Barros

Rita do Carmo Ferreira Laipelt

# Organização e Representação do Conhecimento em Múltiplas Abordagens

organizadores

Thiago Henrique Bragato Barros

Rita do Carmo Ferreira Laipelt

# Organização e Representação do Conhecimento em Múltiplas Abordagens

| São Paulo | 2022 |



Copyright © Pimenta Cultural, alguns direitos reservados.

Copyright do texto © 2022 os autores e as autoras.

Copyright da edição © 2022 Pimenta Cultural.

Esta obra é licenciada por uma Licença Creative Commons: Atribuição-NãoComercial-SemDerivações 4.0 Internacional - (CC BY-NC-ND 4.0). Os termos desta licença estão disponíveis em: <<https://creativecommons.org/licenses/>>. Direitos para esta edição cedidos à Pimenta Cultural. O conteúdo publicado não representa a posição oficial da Pimenta Cultural.

## CONSELHO EDITORIAL CIENTÍFICO

### Doutores e Doutoradas

Adilson Cristiano Habowski

*Universidade La Salle, Brasil*

Adriana Flávia Neu

*Universidade Federal de Santa Maria, Brasil*

Adriana Regina Vettorazzi Schmitt

*Instituto Federal de Santa Catarina, Brasil*

Aguimário Pimentel Silva

*Instituto Federal de Alagoas, Brasil*

Alaim Passos Bispo

*Universidade Federal do Rio Grande do Norte, Brasil*

Alaim Souza Neto

*Universidade Federal de Santa Catarina, Brasil*

Alessandra Knoll

*Universidade Federal de Santa Catarina, Brasil*

Alessandra Regina Müller Germani

*Universidade Federal de Santa Maria, Brasil*

Aline Corso

*Universidade do Vale do Rio dos Sinos, Brasil*

Aline Wendpap Nunes de Siqueira

*Universidade Federal de Mato Grosso, Brasil*

Ana Rosângela Colares Lavand

*Universidade Federal do Pará, Brasil*

André Gobbo

*Universidade Federal da Paraíba, Brasil*

Andressa Wiebusch

*Universidade Federal de Santa Maria, Brasil*

Andreza Regina Lopes da Silva

*Universidade Federal de Santa Catarina, Brasil*

Angela Maria Farah

*Universidade de São Paulo, Brasil*

Anísio Batista Pereira

*Universidade Federal de Uberlândia, Brasil*

Antonio Edson Alves da Silva

*Universidade Estadual do Ceará, Brasil*

Antonio Henrique Coutelo de Moraes

*Universidade Federal de Rondonópolis, Brasil*

Arthur Vianna Ferreira

*Universidade do Estado do Rio de Janeiro, Brasil*

Ary Albuquerque Cavalcanti Junior

*Universidade Federal de Mato Grosso, Brasil*

Asterlindo Bandeira de Oliveira Júnior

*Universidade Federal da Bahia, Brasil*

Bárbara Amaral da Silva

*Universidade Federal de Minas Gerais, Brasil*

Bernadette Beber

*Universidade Federal de Santa Catarina, Brasil*

Bruna Carolina de Lima Siqueira dos Santos

*Universidade do Vale do Itajaí, Brasil*

Bruno Rafael Silva Nogueira Barbosa

*Universidade Federal da Paraíba, Brasil*

Caio Cesar Portella Santos

*Instituto Municipal de Ensino Superior de São Manuel, Brasil*

Carla Wanessa do Amaral Caffagni

*Universidade de São Paulo, Brasil*

Carlos Adriano Martins

*Universidade Cruzeiro do Sul, Brasil*

Carlos Jordan Lapa Alves

*Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brasil*

Caroline Chioquetta Lorenset

*Universidade Federal de Santa Catarina, Brasil*

Cássio Michel dos Santos Camargo  
*Universidade Federal do Rio Grande do Sul-Faced, Brasil*

Christiano Martino Otero Avila  
*Universidade Federal de Pelotas, Brasil*

Cláudia Samuel Kessler  
*Universidade Federal do Rio Grande do Sul, Brasil*

Cristiane Silva Fontes  
*Universidade Federal de Minas Gerais, Brasil*

Daniela Susana Segre Guertzenstein  
*Universidade de São Paulo, Brasil*

Daniele Cristine Rodrigues  
*Universidade de São Paulo, Brasil*

Dayse Centurion da Silva  
*Universidade Anhanguera, Brasil*

Dayse Sampaio Lopes Borges  
*Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brasil*

Diego Pizarro  
*Instituto Federal de Brasília, Brasil*

Dorama de Miranda Carvalho  
*Escola Superior de Propaganda e Marketing, Brasil*

Edson da Silva  
*Universidade Federal dos Vales do Jequitinhonha e Mucuri, Brasil*

Elena Maria Mallmann  
*Universidade Federal de Santa Maria, Brasil*

Eleonora das Neves Simões  
*Universidade Federal do Rio Grande do Sul, Brasil*

Eliane Silva Souza  
*Universidade do Estado da Bahia, Brasil*

Elvira Rodrigues de Santana  
*Universidade Federal da Bahia, Brasil*

Éverly Pegoraro  
*Universidade Federal do Rio de Janeiro, Brasil*

Fábio Santos de Andrade  
*Universidade Federal de Mato Grosso, Brasil*

Fabrcia Lopes Pinheiro  
*Universidade Federal do Estado do Rio de Janeiro, Brasil*

Felipe Henrique Monteiro Oliveira  
*Universidade Federal da Bahia, Brasil*

Fernando Vieira da Cruz  
*Universidade Estadual de Campinas, Brasil*

Gabriella Eldereti Machado  
*Universidade Federal de Santa Maria, Brasil*

Germano Ehlert Pollnow  
*Universidade Federal de Pelotas, Brasil*

Geymeesson Brito da Silva  
*Universidade Federal de Pernambuco, Brasil*

Giovanna Ofretorio de Oliveira Martin Franchi  
*Universidade Federal de Santa Catarina, Brasil*

Handherson Leylton Costa Damasceno  
*Universidade Federal da Bahia, Brasil*

Hebert Elias Lobo Sosa  
*Universidad de Los Andes, Venezuela*

Helciclever Barros da Silva Sales  
*Instituto Nacional de Estudos e Pesquisas Educacionais  
Anísio Teixeira, Brasil*

Helena Azevedo Paulo de Almeida  
*Universidade Federal de Ouro Preto, Brasil*

Hendy Barbosa Santos  
*Faculdade de Artes do Paraná, Brasil*

Humberto Costa  
*Universidade Federal do Paraná, Brasil*

Igor Alexandre Barcelos Graciano Borges  
*Universidade de Brasília, Brasil*

Inara Antunes Vieira Willerding  
*Universidade Federal de Santa Catarina, Brasil*

Ivan Farias Barreto  
*Universidade Federal do Rio Grande do Norte, Brasil*

Jaziel Vasconcelos Dorneles  
*Universidade de Coimbra, Portugal*

Jean Carlos Gonçalves  
*Universidade Federal do Paraná, Brasil*

Jocimara Rodrigues de Sousa  
*Universidade de São Paulo, Brasil*

Joelson Alves Onofre  
*Universidade Estadual de Santa Cruz, Brasil*

Jónata Ferreira de Moura  
*Universidade São Francisco, Brasil*

Jorge Eschriqui Vieira Pinto  
*Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil*

Jorge Luís de Oliveira Pinto Filho  
*Universidade Federal do Rio Grande do Norte, Brasil*

Juliana de Oliveira Vicentini  
*Universidade de São Paulo, Brasil*

Julierme Sebastião Morais Souza  
*Universidade Federal de Uberlândia, Brasil*

Junior César Ferreira de Castro  
*Universidade de Brasília, Brasil*

Katía Bruginiski Mulik  
*Universidade de São Paulo, Brasil*

Laionel Vieira da Silva  
*Universidade Federal da Paraíba, Brasil*

Leonardo Pinheiro Mozdzenski  
*Universidade Federal de Pernambuco, Brasil*

Lucila Romano Tragtenberg  
*Pontifícia Universidade Católica de São Paulo, Brasil*

Lucimara Rett  
*Universidade Metodista de São Paulo, Brasil*

Manoel Augusto Polastreli Barbosa  
*Universidade Federal do Espírito Santo, Brasil*

Marcelo Nicomedes dos Reis Silva Filho  
*Universidade Estadual do Oeste do Paraná, Brasil*

Marcio Bernardino Sirino  
*Universidade Federal do Estado do Rio de Janeiro, Brasil*

Marcos Pereira dos Santos  
*Universidad Internacional Iberoamericana del Mexico, México*

Marcos Uzel Pereira da Silva  
*Universidade Federal da Bahia, Brasil*

Maria Aparecida da Silva Santandel  
*Universidade Federal de Mato Grosso do Sul, Brasil*

Maria Cristina Giorgi  
*Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brasil*

Maria Edith Maroca de Avelar  
*Universidade Federal de Ouro Preto, Brasil*

Marina Bezerra da Silva  
*Instituto Federal do Piauí, Brasil*

Michele Marcelo Silva Bortolai  
*Universidade de São Paulo, Brasil*

Mônica Tavares Orsini  
*Universidade Federal do Rio de Janeiro, Brasil*

Nara Oliveira Salles  
*Universidade do Estado do Rio de Janeiro, Brasil*

Neli Maria Mengalli  
*Pontifícia Universidade Católica de São Paulo, Brasil*

Patricia Biegging  
*Universidade de São Paulo, Brasil*

Patricia Flavia Mota  
*Universidade do Estado do Rio de Janeiro, Brasil*

Raul Inácio Busarello  
*Universidade Federal de Santa Catarina, Brasil*

Raymundo Carlos Machado Ferreira Filho  
*Universidade Federal do Rio Grande do Sul, Brasil*

Roberta Rodrigues Ponciano  
*Universidade Federal de Uberlândia, Brasil*

Robson Teles Gomes  
*Universidade Federal da Paraíba, Brasil*

Rodiney Marcelo Braga dos Santos  
*Universidade Federal de Roraima, Brasil*

Rodrigo Amancio de Assis  
*Universidade Federal de Mato Grosso, Brasil*

Rodrigo Sarruge Molina  
*Universidade Federal do Espírito Santo, Brasil*

Rogério Rauber  
*Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil*

Rosane de Fatima Antunes Obregon  
*Universidade Federal do Maranhão, Brasil*

Samuel André Pompeio  
*Universidade Estadual Paulista Júlio de Mesquita Filho, Brasil*

Sebastião Silva Soares  
*Universidade Federal do Tocantins, Brasil*

Silmar José Spinardi Franchi  
*Universidade Federal de Santa Catarina, Brasil*

Simone Alves de Carvalho  
*Universidade de São Paulo, Brasil*

Simoni Urnau Bonfiglio  
*Universidade Federal da Paraíba, Brasil*

Stela Maris Vaucher Farias  
*Universidade Federal do Rio Grande do Sul, Brasil*

Tadeu João Ribeiro Baptista  
*Universidade Federal do Rio Grande do Norte*

Taiane Aparecida Ribeiro Nepomoceno  
*Universidade Estadual do Oeste do Paraná, Brasil*

Taiza da Silva Gama  
*Universidade de São Paulo, Brasil*

Tania Micheline Miorando  
*Universidade Federal de Santa Maria, Brasil*

Tarcísio Vanzin  
*Universidade Federal de Santa Catarina, Brasil*

Tascieli Feltrin  
*Universidade Federal de Santa Maria, Brasil*

Tayson Ribeiro Teles  
*Universidade Federal do Acre, Brasil*

Thiago Barbosa Soares  
*Universidade Federal de São Carlos, Brasil*

Thiago Camargo Iwamoto  
*Pontifícia Universidade Católica de Goiás, Brasil*

Thiago Medeiros Barros  
*Universidade Federal do Rio Grande do Norte, Brasil*

Tiago Mendes de Oliveira  
*Centro Federal de Educação Tecnológica de Minas Gerais, Brasil*

Vanessa Elisabete Raue Rodrigues  
*Universidade Estadual de Ponta Grossa, Brasil*

Vania Ribas Ulbricht  
*Universidade Federal de Santa Catarina, Brasil*

Wellington Furtado Ramos  
*Universidade Federal de Mato Grosso do Sul, Brasil*

Wellton da Silva de Fatima  
*Instituto Federal de Alagoas, Brasil*

Yan Masetto Nicolai  
*Universidade Federal de São Carlos, Brasil*

## PARECERISTAS E REVISORES(AS) POR PARES

### Avaliadores e avaliadoras Ad-Hoc

Alessandra Figueiró Thornton

*Universidade Luterana do Brasil, Brasil*

Alexandre João Appio

*Universidade do Vale do Rio dos Sinos, Brasil*

Bianka de Abreu Severo

*Universidade Federal de Santa Maria, Brasil*

Carlos Eduardo Damian Leite

*Universidade de São Paulo, Brasil*

Catarina Prestes de Carvalho

*Instituto Federal Sul-Rio-Grandense, Brasil*

Eliisene Borges Leal

*Universidade Federal do Piauí, Brasil*

Elizabeth de Paula Pacheco

*Universidade Federal de Uberlândia, Brasil*

Elton Simomukay

*Universidade Estadual de Ponta Grossa, Brasil*

Francisco Geová Goveia Silva Júnior

*Universidade Potiguar, Brasil*

Indiamaris Pereira

*Universidade do Vale do Itajaí, Brasil*

Jacqueline de Castro Rimá

*Universidade Federal da Paraíba, Brasil*

Lucimar Romeu Fernandes

*Instituto Politécnico de Bragança, Brasil*

Marcos de Souza Machado

*Universidade Federal da Bahia, Brasil*

Michele de Oliveira Sampaio

*Universidade Federal do Espírito Santo, Brasil*

Pedro Augusto Paula do Carmo

*Universidade Paulista, Brasil*

Samara Castro da Silva

*Universidade de Caxias do Sul, Brasil*

Thais Karina Souza do Nascimento

*Instituto de Ciências das Artes, Brasil*

Viviane Gil da Silva Oliveira

*Universidade Federal do Amazonas, Brasil*

Weyber Rodrigues de Souza

*Pontifícia Universidade Católica de Goiás, Brasil*

William Roslindo Paranhos

*Universidade Federal de Santa Catarina, Brasil*

## PARECER E REVISÃO POR PARES

Os textos que compõem esta obra foram submetidos para avaliação do Conselho Editorial da Pimenta Cultural, bem como revisados por pares, sendo indicados para a publicação.

Direção editorial	Patricia Bieging e Raul Inácio Busarello
Editora executiva	Patricia Bieging
Coordenadora editorial	Landressa Rita Schiefelbein
Marketing digital	Lucas Andrius de Oliveira
Diretor de criação	Raul Inácio Busarello
Assistente de arte	Naiara Von Groll
Editoração eletrônica	Peter Valmorbida e Potira Manoela de Moraes
Imagens da capa	Bizkette1, Starline - Freepik.com
Tipografias	Swiss 721, Aileron, Libel Suit
Revisão	Maria Amália Cassol Lied
Organizadores	Thiago Henrique Bragato Barros Rita do Carmo Ferreira Laipelt

Dados Internacionais de Catalogação na Publicação (CIP)

O68

Organização e representação do conhecimento em múltiplas abordagens / Organizadores Thiago Henrique Bragato Barros, Rita do Carmo Ferreira Laipelt. – São Paulo: Pimenta Cultural, 2022.

Livro em PDF

ISBN 978-65-5939-561-3

DOI 10.31560/pimentacultural/2022.95613

1. Organização do conhecimento. 2. Metodologia.  
3. Arquivologia. I. Barros, Thiago Henrique Bragato  
(Organizador). II. Laipelt, Rita do Carmo Ferreira (Organizadora).  
III. Título.

CDD 020

Índice para catálogo sistemático:

I. Organização do conhecimento

Janaina Ramos – Bibliotecária – CRB-8/9166

**PIMENTA CULTURAL**

São Paulo · SP

Telefone: +55 (11) 96766 2200

livro@pimentacultural.com

www.pimentacultural.com



2 0 2 2

## Prefácio

A Organização do Conhecimento atua como um espaço interdisciplinar viabilizando pesquisas que oferecem respostas relacionadas à representação para recuperação, preservação e difusão da informação. Para tanto, pesquisadores e estudiosos se valem de diferentes tradições de pesquisa, posturas epistêmicas e abordagens metodológicas, trazendo uma riqueza singular à área.

Muitos são os atores que dão forma à Organização do Conhecimento e juntos partem da premissa de que a representação do conhecimento lida com contextos, domínios do conhecimento e comunidades de usuários.

Além da representação dos atributos de um documento e de seus contextos, a representação do conhecimento trabalha com abstrações da realidade, isto é, se vale de sistemas de organização do conhecimento, que atuam como modelos de mundo, para construir as pontes que serão trilhadas pelos usuários até os documentos. Uma vez recuperadas e acessadas, os usuários poderão se apropriar de informações e terão, assim, seu direito à informação garantido para exercer plenamente sua cidadania, tomar decisões, construir novos conhecimentos e vivências culturais.

As formas como essas pontes são construídas, e os sistemas de organização do conhecimento utilizados para sua estruturação, vêm sendo investigados por especialistas com diferentes formações. Para tanto, são desenvolvidas pesquisas que visam analisar ou propor processos, instrumentos ou produtos para organizar e recuperar informações e documentos. Na mesma medida, são empreendidas pesquisas sobre a própria área de Organização do Conhecimento e a literatura sobre sua conceituação, objetivos e diferentes correntes teóricas que a compõe é vasta.

No bojo da Ciência da Informação, os estudos vão de pesquisas clássicas sobre classificação em bibliotecas, presentes desde a gênese da Organização do Conhecimento, passando por temas permeados pelos impactos das tecnologias da informação e comunicação na organização, recuperação e uso da informação, chegando até estudos de classificação e descrição arquivísticas. É evidente que a Documentação e a Diplomática trouxeram elementos fundamentais para as discussões que ocorrem, atualmente, também no bojo da Organização do Conhecimento.

O Brasil tem presença constante e coesa nos debates acerca do ensino, pesquisa e atuação profissional no tocante à Organização do Conhecimento, nacional e internacionalmente, e a obra que aqui se apresenta é um excelente exemplo disso.

Com o intuito de oferecer uma coletânea de estudos sobre ou no âmbito da Organização do Conhecimento, foram reunidos autores brasileiros e estrangeiros, professores, pós-graduandos e profissionais especializados em diferentes temáticas. Destaca-se que o fio condutor da obra foi construído de forma sólida para acompanhar o leitor em seu percurso.

Assim, o leitor encontrará, na primeira parte da obra, estudos sobre aspectos epistemológicos da Organização do Conhecimento, isto é, estudos que investigam suas premissas, teorias, posturas epistêmicas e métodos, assim como seu percurso histórico, paradigmas e relações. A segunda parte trará ao leitor possibilidades de pensar sobre contextos aplicados a partir de pesquisas que analisaram ou desenvolveram processos, aplicaram ou avaliaram sistemas de organização do conhecimento ou propuseram produtos no âmbito da Organização do Conhecimento.

Os pesquisadores da área de Organização do Conhecimento a analisam, observam os seus atores e propõem agendas de

pesquisa. Nesse sentido, essa obra contém estudos teóricos que analisam as variações terminológicas do conceito de representação de assunto, bem como de que forma se dá a formação de conceitos partindo da sua existência mental e chegando à sua concepção como unidade de conhecimento.

Estudos interdisciplinares também são empreendidos e, nesta oportunidade, destacam-se as contribuições que as taxonomias podem trazer para construção de planos de classificação para documentos de arquivo, bem como as convergências entre a Organização do Conhecimento e a Arquivologia.

Para se pensar os processos, e o impacto das tecnologias de informação e comunicação neste fazer, estudos teóricos e aplicados são apresentados sobre a classificação automática de documentos de arquivo em sistemas informatizados de gestão de documentos e sobre análises terminológicas a partir das garantias literária, de usuário, cultural e outras.

Com vistas à ampliação dos recursos para compreensão da linguagem e uso dos sistemas de informação, estudos sobre análise de *logs* são empreendidos, tanto no âmbito dos sistemas de recuperação da informação, quanto para análise de reduções sintagmáticas no processo de indexação. Também são apresentados os benefícios que a observação da representação colaborativa em plataformas tecnológicas pode trazer para a área.

Os sistemas de organização do conhecimento são os instrumentos que viabilizam a comunicação, via linguagem, entre os usuários e os sistemas e, nesta obra, são pensados a partir da lógica usada para sua construção e, em uma perspectiva mais aplicada, é analisado o uso de softwares para criação de tesauros para arquivos.

Analisar os usuários, suas necessidades, comportamentos e práticas informacionais é fundamental para a Organização do

Conhecimento e essa premissa fica evidente no estudo sobre acesso e recuperação da informação em arquivos e bibliotecas. Tendências de pesquisa apontam para a necessária análise do engajamento dos usuários em redes sociais e, aqui, uma análise dessa abordagem em temas com viés político é realizada sob a ótica da folksonomia.

Reflexões sociais e políticas acompanham os movimentos de estudiosos e instituições que amparam a Organização do Conhecimento em sua história recente e, nesta obra, são apresentadas reflexões éticas no âmbito da Organização do Conhecimento e a presença das urgentes discussões sobre justiça social no bojo desta área. O foco é avançar na representação no que diz respeito aos domínios do conhecimento, principalmente os que necessitam de pesquisas atualizadas, e de comunidades de usuários, principalmente as que se encontram às margens dos sistemas de informação.

Representar o conhecimento de forma defensável exige um exercício constante de reflexões teóricas e metodológicas e a compreensão da responsabilidade dos profissionais e estudiosos da Organização do Conhecimento, sendo assim convido vocês, queridos leitores, a percorrerem as páginas dessa obra com uma postura ativa e crítica, afinal, como nos ensina Hope A. Olson, embora o objetivo possa não ser o de exercer poder, exceto o poder da recuperação da informação, atuamos no âmbito de um poderoso discurso cultural.

Bons estudos!

Profa. Dra. Suellen Oliveira Milani,  
Universidade Federal Fluminense

Niterói, 14 de março de 2022

## Apresentação

Esta obra foi organizada com o objetivo de apresentar as pesquisas desenvolvidas no âmbito da Organização do Conhecimento realizadas por membros (pesquisadores, egressos e estudantes) do Grupo de pesquisa Organização e Representação do Conhecimento Abordagens Linguísticas em Arquivos e Bibliotecas – ORCALAB/ UFRGS ou da rede por ele estabelecida.

Os textos foram sistematizados em duas partes no intuito de destacar na parte I Abordagem Epistemológica na Organização do Conhecimento aspectos teóricos e metodológicos relevantes para a área e, na parte II Abordagem Aplicada na Organização do Conhecimento, exemplos de estudos aplicados baseados em grande medida, mas não apenas, nas teorias e metodologias apresentadas na parte

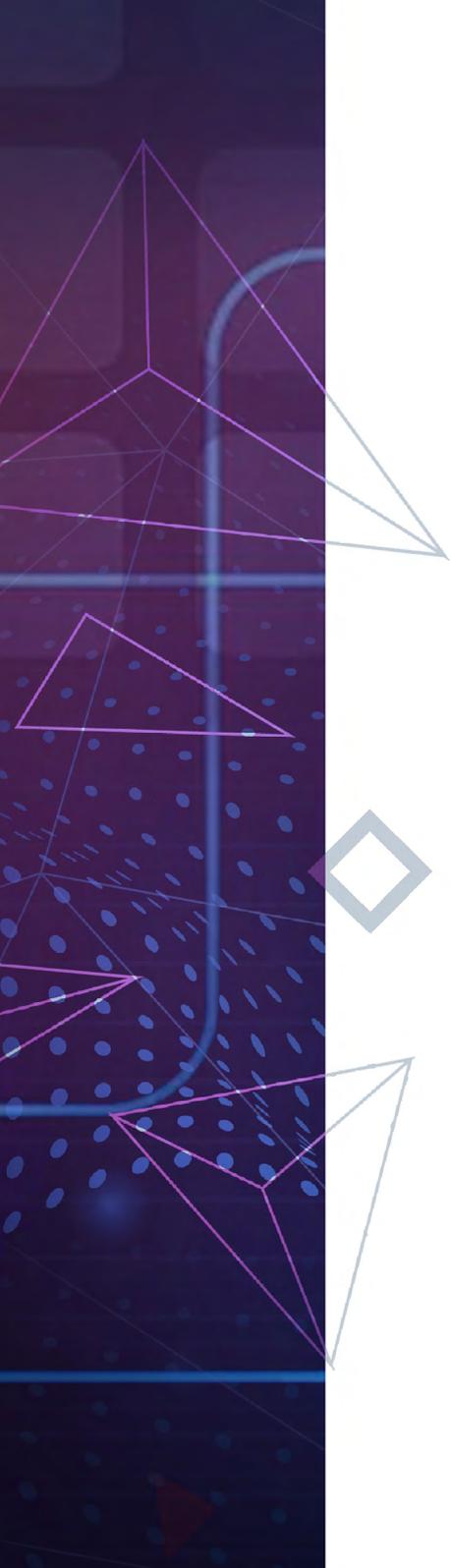
A rede de pesquisa formada a partir do ORCALAB possui pesquisadores de 9 Universidades Brasileiras (UFRGS, UFPA, UFSCAR, UNIRIO, UFSC, UNB, UFPE, UNESP E UFF) e 4 universidades estrangeiras (Universidad de la República, Uruguai, Universidade de Coimbra, Portugal, Universidad de Leon e University of Western Ontario, Canada) tratando-se assim de uma obra coletiva, colaborativa e que busca discutir temáticas da organização do conhecimento de grande relevância para área.

Não era nosso objetivo sermos exaustivos nem definitivos em relação às abordagens teóricas, metodológicas e temáticas dos textos devido às múltiplas possibilidades de pesquisas que a área de Organização do Conhecimento possibilita. Ainda assim, tentamos contemplar a maior diversidade possível para questões contemporâneas que permeiam a área e que necessitam ser discutidas, não necessariamente para encontrar respostas conclusivas, mas certamente como um convite à reflexão e a exploração de outros horizontes e perspectivas.

Convidamos todos a leitura e discussão nas múltiplas abordagens da organização do conhecimento na atualidade!

Thiago Henrique Bragato Barros e Rita do Carmo Ferreira Laipelt

Porto Alegre, 16 de maio de 2022



## Sumário

Parte 1

Abordagem epistemológica  
na organização do conhecimento

Capítulo 1

**Uma teoria “modal”  
de organização da informação ..... 19**

***A “modal” theory  
of information organization***

*Donald Grant Campbell*

Capítulo 2

**O assunto, a representação: uma análise  
com foco na variação terminológica a partir  
da série *Estudos Avançados em Organização  
do Conhecimento* (ISKO Brasil) ..... 44**

***The subject, representation: an analysis  
with focus on terminological variation from  
the series *Advanced Studies in Knowledge  
Organization* (ISKO Brazil)***

*Brisa Pozzi de Sousa*

Capítulo 3

**A indeterminação ontológica dos conceitos:  
interpretações linguísticas e psicológicas..... 94**

***The ontologic indetermination of concepts:  
linguistic and psychological interpretations***

*Luís Miguel Oliveira Machado  
Wilson Roberto Veronez Júnior  
Daniel Martínez-Ávila*

Capítulo 4

**Las garantías más allá del horizonte  
de la organización del conocimiento:**

proyecciones posibles..... 124

***Warrants beyond the horizon  
of knowledge organization:***

possible projections

*Mario Barité*

Capítulo 5

**Aspectos éticos em organização  
e representação do conhecimento:**

uma análise de sua presença na literatura  
internacional da ciência da informação ..... 149

***Ethical aspects in knowledge organization:***

*an analysis about the theme in the international  
literature in information science*

*Isadora Victorino Evangelista*

*José Augusto Chaves Guimarães*

Capítulo 6

**Metodologia de análise de logs  
na ciência da informação:**

revisão de literatura e melhores práticas..... 181

***Logs analysis as methodology  
for studies on information science:***

*literature review and best practices*

*Luciana Monteiro-Krebs*

*Rita do Carmo Ferreira Laipelt*

*Rafael Port da Rocha*

Capítulo 7

**A classificação automática de documentos de arquivo é uma solução para os problemas que os usuários encontram com a classificação funcional? Algumas reflexões e caminhos a percorrer ..... 221**

***Is automatic classification of records a solution to the problems users encounter with functional classification? Some thoughts and ways forward***  
*Renato Tarciso Barbosa de Sousa*

Parte 2

Abordagem aplicada na organização do conhecimento

Capítulo 8

**Reduções sintagmáticas e indexação: uma análise a partir de logs de pesquisa dos usuários de um sistema de recuperação da informação..... 252**

***Syntagmatic reductions and indexing: an analysis from user search logs of an information retrieval system***  
*Isabel Cristina Pereira dos Santos*

Capítulo 9

**As taxonomias no contexto da arquivologia ..... 281**

***Taxonomies in the context of archives***  
*Carine Melo Cogo Bastos*  
*Thiago Henrique Bragato Barros*

Capítulo 10

<b>Reflexão sobre folksonomia e engajamento político .....</b>	<b>296</b>
--------------------------------------------------------------------	------------

***Reflection on folksonomy  
and political engagement***

*Talita Morgana Arruda Tavares  
Fabio Assis Pinho*

Capítulo 11

<b>Organização e recuperação da informação: desafios na Arquivologia e na Biblioteconomia.....</b>	<b>320</b>
--------------------------------------------------------------------------------------------------------	------------

***Organization and information retrieval:  
challenges in Archival Science and Librarianship***

*Andréa Fontoura da Silva  
Bruna Carballo Dominguez de Almeida  
Joana Peregrina Hernandes*

Capítulo 12

<b>Desdobramentos da transcrição colaborativa nas abordagens teórico-metodológicas da organização do conhecimento .....</b>	<b>341</b>
-------------------------------------------------------------------------------------------------------------------------------------	------------

***Developments of collaborative transcription  
in knowledge organization  
theoretical-methodological approaches***

*Camila Monteiro de Barros  
Edgar Bisset Alvarez*

Capítulo 13

**O software Thesa como SOC no contexto  
arquivístico: uma proposta de modelo teórico ..... 365**

***Thesa software as SOC in the archival  
context: a proposal for a theoretical model***

*Adriana Carla Ribeiro dos Santos*

*Raí Rocha Costa*

*Marcos Oliveira da Costa*

Capítulo 14

**A justiça social na organização  
do conhecimento..... 391**

***Social justice in Knowledge Organization***

*Lucas Andrade Sá Corrêa*

*Natália Bolfarini Tognoli*

Capítulo 15

**Organização do conhecimento e Arquivologia:  
diálogos e prospecções ..... 418**

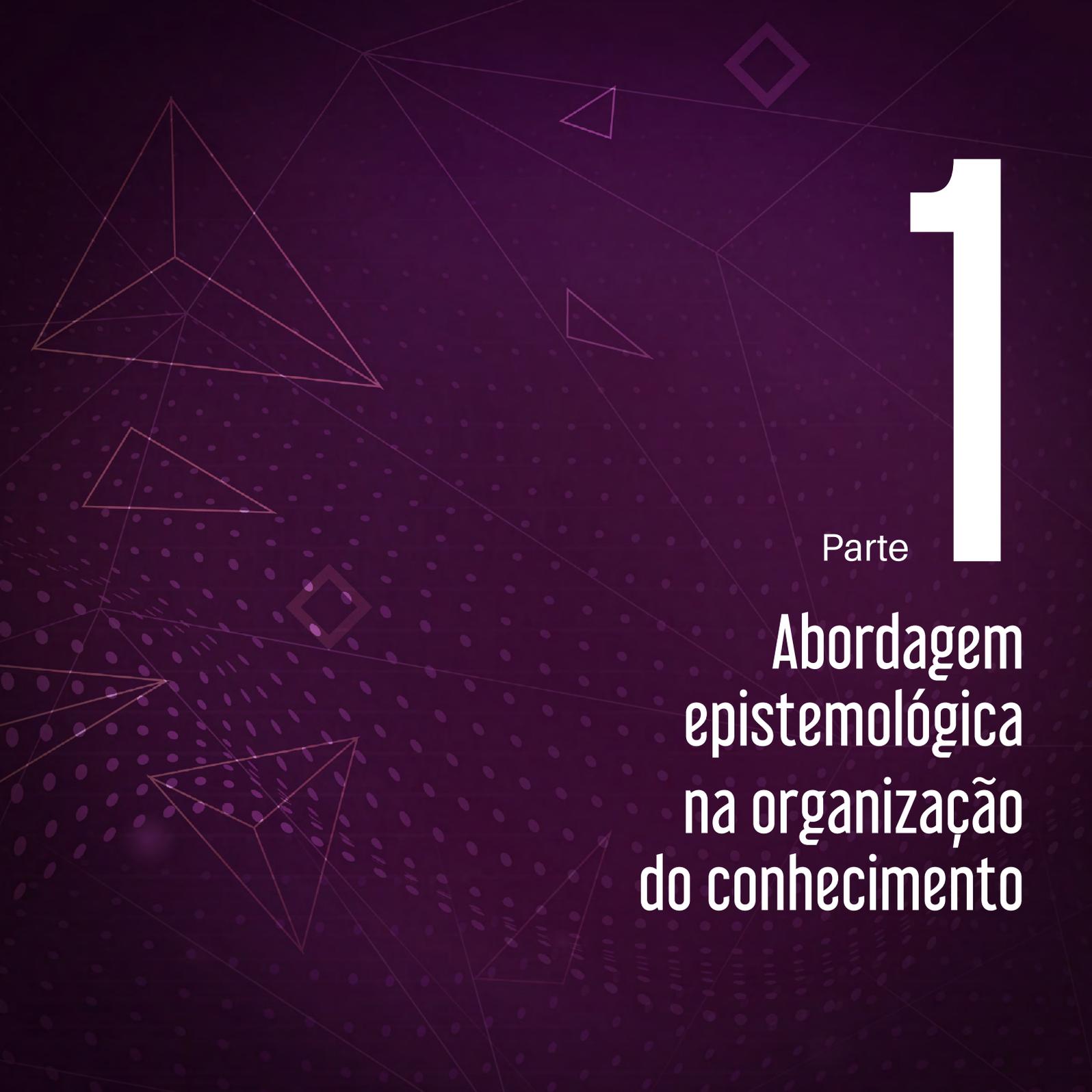
***Knowledge organization and Archival Science:***

*dialogues and prospects*

*Leolíbia Luana Linden*

**Sobre os organizadores ..... 442**

**Sobre os autores ..... 443**



# 1

Parte

## Abordagem epistemológica na organização do conhecimento

# 1

Donald Grant Campbell

Uma teoria “modal”  
de organização da informação

*A “modal” theory  
of information organization*

DOI: 10.31560/pimentacultural/2022.95613.1



**Abstract:**

This paper applies Northrop Frye's theory of modes to the development of professional tools of information organization. In *Words with Power*, Frye distinguishes between two modes of discourse. "Description" is a discourse common in scientific writing, in which the author strives to describe some external reality as faithfully and accurately as possible. Controlled vocabularies such as the Library of Congress Subject Headings strive for such accuracy, seeking to create uniform and unique headings that are coextensive with their subjects. "Dialectic," on the other hand, is a discourse common in philosophical argument, where meaning is generated internally, by the rules and patterns of the language and the argument. As simple controlled vocabularies evolve into thesauri and ontologies, they create internal rules of relationship between the two terms: relationships of equivalence, hierarchy, and association. Meaning in controlled vocabularies, therefore, emerges through a combination of what Frye calls centripetal and centrifugal force. On the one hand, the terms of a vocabulary refer outward, "centrifugally," to the external world, establishing a firm and useful relationship between the index term and the phenomenon to which it refers. On the other hand, the terms refer inwardly, "centripetally," to the network of relationships within the vocabulary, establishing firm and useful distinctions between each index term and other terms to which it is related.

**Keywords:** Modes; controlled vocabularies; discourse.

### **Resumo:**

*Esta pesquisa aplica a teoria dos “modos”, de Northrop Frye, ao desenvolvimento de ferramentas profissionais de organização da informação. Em Words with Power, Frye distingue entre dois tipos de discurso: “Descrição” e “Conceptual”. “Descrição” é um discurso comum na escrita científica, em que o autor se esforça para descrever uma realidade externa da forma mais fiel e precisa possível. Vocabulários controlados, como os termos de indexação da Biblioteca do Congresso, buscam essa precisão para criar termos uniformes e únicos, que são coextensivos com os assuntos descritos. “Conceptual”, por outro lado, é um discurso comum na argumentação filosófica em que o significado é gerado internamente pelas regras e padrões da linguagem e do argumento. À medida que os vocabulários controlados evoluem para os tesouros e ontologias, eles criam regras internas das relações entre os termos: relações de equivalência, hierarquia e associação. O significado, nos vocabulários controlados, então, surge por uma combinação do que Frye chama de uma força centrípeta e uma força centrífuga. De um lado, os termos de um vocabulário referem-se ao mundo externo, para fora, estabelecendo uma relação firme e útil entre o termo e o fenômeno ao qual ele representa. De outro lado, os termos referem-se de forma centrípeta, para as relações dentro do vocabulário, estabelecendo as distinções entre cada termo e os outros termos relacionados.*

**Palavras-Chave:** Modos; vocabulários controlados; discurso.

## 1 INTRODUCTION

The concepts of “bias” and “neutrality” continue to vex us in Information Organization, particularly in relation to the tools we design for information access. In the decades, since Sanford Berman first began to challenge the Library of Congress to revise its subject headings to remove implicit biases (1971), theorists and practitioners of information organization have become increasingly aware that the “neutrality” to which libraries have traditionally aspired is illusory and even dangerously mistaken.

While most of us in the field accept that absolute neutrality is impossible, the strategies for dealing with the problem vary. Some, like Rick Szostak, suggest that information professionals circumvent the contested ideas in different fields by diving below them and focusing on the non-contested building blocks such as terminology, which he calls “basic concepts”:

It has generally not been difficult to render expert advice in terminology (i.e. basic concepts) that will have a broadly shared understanding — well enough for the purposes of guiding diverse users to documents or objects — outside of the field of the expert. The classificationist need not and should not take sides on theoretical disputes within a field but should rather seek a structure that has a place for any phenomenon identified either theoretically or empirically in the field (SZOSTAK, 2019).

This approach does not sit well with Birgir Hjørland, who argues that we should design and use such tools as classification systems in full awareness of the biases that prevail, not only in the fields we are classifying, but in ourselves as information theorists and professionals. Taking his cue from overtly feminist analyses of information organization systems, Hjørland urges us to engage explicitly with scholarly debates in the field we are classifying, and to profess our own epistemological positions in the process:

In order to classify a given domain one must examine how it is classified according to contemporary knowledge (including different views), to discuss the basis, the epistemological assumptions, and which interests are served by proposed classifications and finally to suggest a motivated classification. This is a highly qualified scholarly activity, and it normally implies taking sides in scholarly disputes at many levels, including the epistemological level. ... Besides studying the fields to be classified, we also need to study our own assumptions in KO and LIS (HJØRLAND, 2020, p. 474).

In this paper, I support Hjørland's position. I also support his dry understatement: "important assumptions are often hidden, making their study more difficult." Biases are notoriously difficult to identify, in ourselves and in others, no matter how honest or self-aware we may be. Even the most disarmingly frank confession of an epistemological position can be self-deceptive: some biases are easier to confess than others, and some are deeply cloaked in denial. The consequences of these biases are also difficult to identify: too often we assume that good intentions produce good outcomes. If the road to hell is paved with good intentions, many roads to intolerance, injustice and atrocity begin with simple attempts to place things in categories and put labels on them.

This study offers a theoretical approach whereby we might take up Hjørland's challenge to engage with biases both in the fields we classify and in ourselves. I begin with "mode": a concept that emerged in both philosophy and music, and became a cornerstone of the literary and linguistic theories of Northrop Frye. I then apply the concept of mode as a metaphorical tool to explore the biases of three different knowledge organization tools: *The Library of Congress Subject Headings (LCSH)*, the National Library of Medicine's *Medical Subject Headings (MeSH)*, and the *Tesouro de Folclore e Cultura Popular Brasileira*, produced by the Centro de Folclore e Cultura Popular Brasileira. The comparison will involve two subjects: mental illness, and alternative medicine.

## 2 THE CONCEPT OF THE MODE

The concept of “the mode” originated in philosophy as a principle of arrangement and combination. For Aristotle, mode refers to the specific combination of propositions within a syllogism necessary for that syllogism to be valid (GARBER, 1999, p. 577). Spinoza defines mode as the qualities of an entity that are received from some other entity; hence, human beings are modes of the divine substance which is God (SPINOZA, 2000, p. 75). In these cases, “mode” refers to a process of creating complex structures by combining simple or basic elements. But the concept of “mode” implies that these structures are not random: they emerge from consistent patterns and principles.

Music theorists through the ages have used the term “mode” to denote widely-accepted patterns and arrangements of tones that we traditionally associate with certain qualities and moods. Some music theorists refer to the major and minor keys of Western music as modes, with the “major” key typically associated with happiness and confidence, and the minor sadness and reflection (RAMOS & Mello, 2021). But modes predate the standard tonal scales of Western music, and many different musical cultures retain a set of standard modes: systems of arranged tones and melodic formulas which have “ethical, emotional and cosmological connotations” (KOLINSKI, 2000). As in philosophy, the term “mode” in music refers to the underlying patterns and elemental principles that silently govern and control creativity, both intellectual and musical.

In 1957, the Canadian literary critic Northrop Frye adopted the term “mode” in the first essay of his seminal work, *The Anatomy of Criticism* (1971). Frye argues that literary works can be classified according to a hierarchically-arranged sequence of modes: myth, romance, tragedy, comedy and satire. Later, Frye revised and adapted his theory of modes in an ambitious attempt to understand the relationship between literature and the Bible, and it is this adaptation of modal theory that I will

be using in this analysis. In *Words with Power* (1990), Frye argues that human discourse is conventionally governed by four primary modes:

- **descriptive:** in which the author attempts to describe the world beyond the written work;
- **conceptual:** in which the author strives to produce an argument that conforms to rules of internal consistency;
- **rhetorical:** in which the author urges the reader to some form of action;
- **imaginative:** in which the author takes the outer world and transforms it into some form of artistic or literary expression (FRYE, 1990, p. 3-29).

This study uses the first two modes, descriptive and conceptual, to analyze the different patterns and arrangements that we find in knowledge organization tools. The descriptive mode is most commonly associated with scientific and empirical research: for the writer in the descriptive mode, validity is achieved through faithful fidelity to the outer world that the author is describing: “The overriding criterion of descriptive writing is, speaking practically, objective truth” (FRYE, 1990, p. 5). The conceptual mode, on the other hand, defines and measures value according to internal rather than external standards: the logical coherence of the argument, the validity of its propositions, the clarity of the writing, and the coherence of its argumentative structure:

The conceptual writer, like the descriptive one, is searching for whatever objective truth words can give him, and he is still appealing to the conscious mind and its sense of objectivity. But he is searching for it within the verbal order he is constructing, and this shows itself in an intense tightening up of the narrative movement. It is most important that sentence B should ‘follow’ from sentence A, and rules of logic are developed to ensure that such following is correct throughout (FRYE, 1990, p. 9).

Clearly, these two modes of discourse are not mutually exclusive. As any scholarly writer will attest, a piece of empirical research must strive, at all times, to tell the truth, while at the same time following a recognizable process of illustration and persuasion. Nonetheless, the concept of descriptive and conceptual modes provides a useful means of exploring the tensions in many forms of discourse. Frye makes frequent use of a concept in physics to describe such tensions:

- **centrifugal force:** which propels the reader's attention outward from a center;
- **centripetal force:** which pulls the reader's attention towards the center (FRYE, 1990, p. 3).

The descriptive mode has a centrifugal momentum, drawing the reader's attention from the text to the outer world, while the conceptual mode has a centripetal momentum, drawing the reader's attention to the internal integrity of the writing itself.

Useful as this distinction might be, it can be complex, and even treacherous. Authors who set their plays and films in "real time," insisting that the time for the audience match the time of the narrative, are appealing to a descriptive credibility while observing tightly rigorous rules of composition. On the other hand, a human figure in a cubist painting by Picasso, while bearing no resemblance to how a human looks in the outer world, is nonetheless expressing, through faceted techniques of composition, a truth about human figures in the real world that might not be obvious to the naked eye.

### 3 MODES IN INFORMATION ORGANIZATION

The metaphor of centripetal and centrifugal force has been used before to describe the multidisciplinary nature of knowledge organization research (CAMPBELL, 2002). This paper, however, links the

metaphor to Frye's definition of descriptive and conceptual modes, in hopes that it will provide a means of illuminating the complex position of knowledge organization tools such as controlled vocabularies, classifications, thesauri and ontologies in relation to disputes over neutrality, bias, veracity and consistency.

In designing and maintaining tools for organizing information in various fields of study, knowledge organization experts work within both descriptive and conceptual modes. On the one hand, they frequently need to address the world outside their tools, often in areas of contest and dispute. Such disputes might include debates between qualitative and quantitative methods in the field of social sciences; evolution and creationism in theology; conventional and alternative therapies in medicine. They must also address the distortions and injustices perpetrated by their systems, particularly in relation to marginalized or underserved populations, through overt or covert ideological assumptions or biases.

On the other hand, access tools such as controlled vocabularies, classifications and ontologies require conceptual focus as well. Elaine Svenonius, for instance, argues that such tools look inward to a collection of documents, rather than outward to what those documents signify:

Subject language terms differ *referentially* from words used in ordinary language. The former do not refer to objects in the real world or concepts in a mentalistic world but to subjects. As a name of a subject, the term *Butterflies* refers not to actual butterflies but rather to the set of all indexed documents about butterflies. In a natural language the extension, or extensional meaning, of a word is the class of entities denoted by that word, such as the class consisting of all butterflies. In a subject language the extension of a term is the class of all documents about what that term denotes, such as all documents about butterflies (SVENONIUS, 2009, p. 130).

To what degree do the descriptive and conceptual modes overlap in KO tools? This preliminary exploration takes two examples, involving two concepts:

- **Mental Illness:** defined by the American Psychiatric Association (S.d.) as “health conditions involving changes in emotion, thinking or behavior (or a combination of these)... associated with distress and/or problems functioning in social, work or family activities” (<https://www.psychiatry.org/patients-families/what-is-mental-illness>);
- **Traditional Medicine:** defined by the World Health Organization (2021) as “the sum total of the knowledge, skill, and practices based on the theories, beliefs, and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health as well as in the prevention, diagnosis, improvement or treatment of physical and mental illness” (<https://www.who.int/health-topics/traditional-complementary-and-integrative-medicine>).

Both concepts, “Mental illness” and “Traditional medicine,” may be said to have a basis in the field of medicine, but both appear in publications beyond medicine as well. The three tools are very different in origin and purpose:

- **Library of Congress Subject Headings (LCSH):** As the Library of Congress’s principal list of subject headings, this is the oldest of the three tools: a pre-coordinate list of subjects, determined by the literary warrant of the LC’s own collection, and thus representing a vast range of different subjects, fields and professions;
- **Medical Subject Headings (MeSH):** As the National Library of Medicine’s controlled vocabulary for its vast database of medical research, MeSH is also a pre-coordinate list of subject headings, but more rigorously designed from a tree-structure classification, and designed primarily for electronic access through databases such as PubMed.
- **Tesouro de Folclore e Cultura Popular Brasileira (TFCBP):** Created by the Centro Nacional de Folclore e Cultura Popular,

this multidisciplinary thesaurus provides index terms for a wide range of documents and museum artifacts.

Both LCSH and *MeSH* have headings representing the phenomenon of mental illness. Both *MeSH* and *TFCBP* contain terms representing the phenomenon of traditional medicine.

#### 4 “MENTAL ILLNESS” IN THE LIBRARY OF CONGRESS, AND THE NATIONAL LIBRARY OF MEDICINE

Both the *Library of Congress Subject Headings* (LCSH) and the *Medical Subject Headings* (MeSH) contain a heading for mental illness. Both offer a subject heading that is co-extensive with the concept, and both contain the following standard features of controlled vocabularies:

- an authorized heading;
- syndetic references, both from equivalent terms to the authorized term, and from the authorized term to broader, narrower and related terms; and
- a scope note which defines the extent of the subject represented by the authorized term.

The Library of Congress selects as its authorized heading the phrase, **Mental illness**; the National Library of Medicine opts for the phrase, **Mental Disorders**. The LC heading suggests a single phenomenon: an “illness” which afflicts the mind, and implicitly requires that we distinguish it from illness of the body. The National Library of Medicine, on the other hand, adopts a plural term that suggests a range of phenomena classified together.

The NLM's list of equivalent terms for **Mental Disorders**, such as "Behavior Disorders," "Mental Illness," "Psychiatric Diseases" and "Psychiatric Illness," are all terms that one might reasonably expect to encounter in the medical literature that forms the collection of the National Library of Medicine. In addition, since mental disorders do, indeed, arise for many different reasons and from many different causes, the list of narrower terms is extensive, including such terms as **Anxiety Disorders, Dissociative Disorders, Neurocognitive Disorders** and **Sleep Wake disorders**. In this sense, one can argue that the MeSH heading operates credibly within the descriptive mode: the heading and its references refer centrifugally to an external reality borne out by the medical research to which it provides access.

The LC heading, **Mental illness**, is less credible when measured by what we know empirically of mental disorders. The narrower terms—**Astrology and mental illness, Dual diagnosis, Genius and mental illness**, and **Neurobehavioral disorders**—make no pretence of completion or comprehensiveness: they are merely the terms that reflect the literary warrant of the Library of Congress itself, and hence operate primarily in the conceptual mode: they tell us less about mental illness than they do about the library's collection. One equivalent term—"Madness"—is hardly reliable as a referent to what we now understand about mental disorders. It does, however, refer to a concept that has a powerful social history: the concept of "madness" as represented in such fields as literature, the fine arts and law as a specific condition and the antithesis of reason and sanity. The descriptive, centrifugal quality of the heading, therefore, exists in the form of an anticipation of cultural norms that Charles Cutter once called "the public's habitual way of looking at things" (1904, p.6). It refers, not necessarily to the physical reality of mental illness, but to an implied cultural construction assumed to be held by those who use the library. Furthermore, these cultural constructions evolve over time, and as outdated terms like "madness" are replaced by more appropriate ones, the old headings are generally preserved as equivalent terms, not because they are necessarily

equivalent, but because they provide important links to older materials. In this sense, the Library's UF terms for **Mental illness** partly represent a synchronic set of equivalents and partly a diachronic history of the terminological development of the concept within the Library's collection.

The scope notes of the two headings show the sharpest contrast. The MeSH heading is entirely descriptive, and consists of a definition of mental disorders:

Psychiatric illness or diseases manifested by breakdowns in the adaptational process expressed primarily as abnormalities of thought, feeling, and behavior producing either distress or impairment of function (NATIONAL LIBRARY OF MEDICINE, 2021).

The LCSH scope note, by contrast, operates entirely in the conceptual mode in its focus on disambiguating the various authorized terms that the user might confuse with **Mental illness**:

Here are entered popular works and works on social aspects of mental disorders. Words on legal aspects of mental illness are entered under Insanity (Law). Systematic descriptions of mental disorders are entered under Psychology, Pathological. Works on clinical aspects of mental disorders, including therapy, are entered under Psychiatry (LIBRARY OF CONGRESS, S.d.).

The scope note serves to emulate the internal order and consistency of the subject headings, to ensure that the user does not use the **Mental illness** to find works on the treatment of mental disorders through psychotherapy.

In the case of mental illness, therefore, the two headings both provide coherence through different combinations of the descriptive and conceptual modes. The National Library of Medicine provides conceptual, centripetal coherence through its carefully-constructed tree structure and comprehensive cross-references. It provides descriptive, centrifugal coherence through its scope note and choice of authorized terms that reflect scientific and medical consensus on the

empirical truth of mental disorders. The Library of Congress, on the other hand, makes tentative centrifugal gestures towards a vaguely-  
-implied cultural view, but its scope note and references focus almost entirely inward, illuminating the coherence and order of the tool itself. This contrast makes perfect sense, given that the Library of Congress caters to a far less well-defined user base, and that its collection contains works that address mental illness, not just from a medical standpoint, but also from imaginative and artistic standpoints as well.

This example might tempt us to conclude that collections supporting empirical research, such as bio-medical collections, tend to emphasize the descriptive mode over the conceptual mode, and are far more likely to be “neutral” and less defined by cultural norms and biases. Unfortunately, the matter is more complicated, as this second example will show.

## 5 “TRADITIONAL MEDICINE” IN THE NATIONAL LIBRARY OF MEDICINE AND THE TESAURO DE FOLCLORE E CULTURA POPULAR BRASILEIRA

Our first example compared a specialized vocabulary to a general or universal one. The second example uses the concept of “traditional medicine” to compare two specialized vocabularies, both containing the concept, but in very different contexts. For this second example, we will once again use the *Medical Subject Headings* as one vocabulary. The other vocabulary, however, comes from a tool very different from the *Library of Congress Subject Headings*: we will use the *Tesouro de Folclore e Cultura Popular Brasileira*, a thesaurus designed to support libraries, archives and museums devoted not just to documents, but to various and diverse artifacts related to the history of Brazilian popular culture.

The National Library of Medicine's treatment of this concept has less centrifugal momentum than its treatment of mental disorders. The subject heading is **Medicine, Traditional**, and it falls under the general heading of **Therapeutics** and the more specific heading of **Complementary therapies**. The scope note for **Complementary therapies** places greater emphasis on disambiguation than on definition:

Therapeutic practices which are not currently considered an integral part of conventional allopathic medical practice. They may lack biomedical explanations but as they become better researched some (PHYSICAL THERAPY MODALITIES; DIET; ACUPUNCTURE) become widely accepted whereas others (humors, radium therapy) quietly fade away, yet are important historical footnotes. Therapies are termed as Complementary when used in addition to conventional treatments and as Alternative when used instead of conventional treatment (DeCS, 2022)

The scope note defines this heading as, in some respects, a waiting area, containing headings that represent practices which do not proceed from conventional allopathic medical practice, and which are not as yet, supported by empirical bio-medical research. The NLM concedes that some of the therapeutics included under this head, such as acupuncture, may well acquire such support, and will presumably be moved elsewhere in the MeSH tree structure. Others, such as the concept of the "humors," are retained only for their historical value. In addition, the NLM treats "Complementary therapies" and "Alternative therapies" as equivalents, but with an important proviso. The preferred term, **Complementary medicine**, suggests that, at least for the users of PubMed, such therapies would complement standard medical procedures rather than replace them. The entry term, "Alternative medicine," refers to the use of such therapies *instead of* conventional methods, a practice the NLM presumably would not expect its user to prefer.

Within **Complementary therapies** we find **Medicine, Traditional** as a term which applies to documents dealing with

Systems of medicine based on cultural beliefs and practices handed down from generation to generation. The concept includes mystical and magical rituals (SPIRITUAL THERAPIES; PHYTOTHERAPY); and other treatments which may not be explained by modern medicine (DeCS, 2022)

When it comes to traditional methods, the NLM provides increased conceptual scaffolding, drawing the user's attention centripetally inward, seeking to clarify potential confusion. This very centripetal effort, paradoxically, provides some descriptive insight: the headings, it seems, are referring outward, not necessarily to an accepted view of medical "reality," but to a contested and conflicted area of medical practice, marked by a division between conventional "Western" medicine and a range of alternate fields of knowledge and practice which draw varying degrees of skepticism. And the conceptual, centripetal mechanisms of the MeSH headings reflect this conflict.

The *Tesouro de Folclore e Cultura Popular Brasileira*, on the other hand, subsumes medicine within a commitment to folkore and popular culture. And, like the MeSH headings, the thesaurus is designed to sustain both a descriptive mode, with centrifugal fidelity to an external "reality," and a conceptual mode, with a concomitant centripetal tendency to internal order. In so doing, the thesaurus distinguishes between thing and process. **Medicamento** refers to the substances used to effect cures, while the activities associated with medicine appear within two different facets. **Farmacopéia popular** refers to the production of medicines through traditional techniques, and is contained in the facet, **Atividade produtiva**. **Medicina popular**, on the other hand, is contained within the generic term, **Atividade ritual**: a term that includes such activities as dance, religious practice and storytelling. Traditional medicine, in the *Tesouro*, rests not within the context of Western biomedical research; instead, it is classed as an activity of sacred and/or profane character, marked by a tradition that contains special meaning for a specific cultural group.

The heading itself, **Medicina popular**, might surprise someone used to MeSH headings, in that it embraces empiricism in its scope note:

Atividade ritual que abrange o conjunto de práticas adotadas pelo povo, com base no conhecimento empírico, destinadas a prevenir ou curar males físicos e espirituais. Envolve componentes herdados que vão sendo reinterpretados e somados a elementos resultantes de diferentes influências. São empregados materiais e recursos tais como vegetais, animais e minerais isoladamente ou em composições. Entre as formas adotadas para a realização das curas estão procedimentos de caráter religioso e mágico. Nesses casos são comuns gestos, rituais, passes, rezas bentinhas, escapulários, etc. para reforçar as terapias (CENTRO NACIONAL DE FOLCLORE E CULTURA POPULAR, 2021).

Activities and practices of a religious and magical nature co-exist with practices based on empirical knowledge: all such activities are cultural, social and ritualistic activities.

Whether or not one agrees with this perspective, the Thesaurus, in its centrifugal dimensions, exerts as great a fidelity to its external reality as the MeSH headings do to theirs. Indeed, the Thesaurus succeeds in clarifying what the MeSH headings fail to acknowledge: that a commitment to bio-medical empirical evidence is, in itself, a cultural activity and an epistemological commitment that shapes the very coherence of the headings themselves. The MeSH headings confine “culture to the heading, Medicine, Traditional, insisting that is headings refer outward to a reality governed and defined by empirical evidence, ignoring the fact that such a commitment is, in itself, a cultural gesture of belief.

## 6 DISCUSSION AND CONCLUSIONS

At first sight, neither of these examples appears to suggest anything we don't already know. We have seen scholars analyze classification systems and controlled vocabularies as texts, using strategies

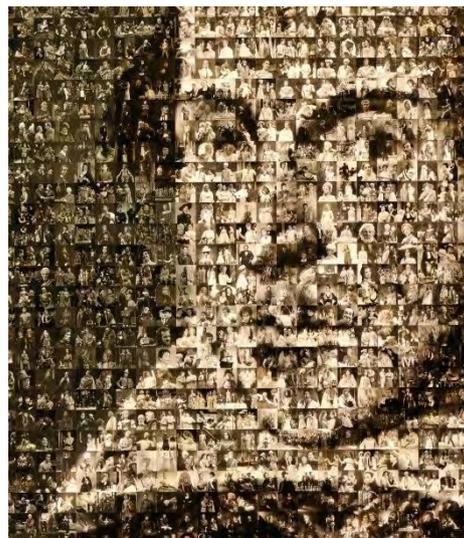
such as text linguistics (BEGHTOL, 1986). Nor should the differences between the tools surprise us. The Library of Congress does not profess to be a sophisticated tool of bio-medical research, and a thesaurus of folklore and popular culture will naturally refer to belief systems and traditional narratives rather than to bio-medical research.

What, then, does the concept of “mode” bring to the study of KO tools and concepts? While these examples make no pretence of being a thorough and complete study, they provide some glimpses into the nature of our tools that could prove helpful in negotiating some of the vexed controversies surrounding bias and neutrality.

If we look at the definitions which open Spinoza’s *Ethics*, we find that he distinguishes “mode” from the concepts of “attribute” and “substance.” A substance is “that which is in itself and is conceived through itself,” while an “attribute” is “that which intellect perceives of substance”; the “mode” constitutes “the affections of substance, or, that which is in something else, through which it is also conceived” (SPINOZA, 2000, p. 75). For Donald Garret (2017), the mode of a substance “constitutes a ‘way’ in which the existence of the substance is expressed (p. 14). A mode, in this sense, is something that reveals the presence and nature of something else: for Spinoza, particular things in the world are modes of God’s attributes (GARRETT, 2017, p. 97). Everything particular offers, besides itself, a fragmentary glimpse of something larger, greater, to which it refers.

We have become accustomed, in modern popular culture, to the use of “picture mosaics,” in which a collection of images is arranged on a massive collection of small tiles to assemble into a larger image, as in this poster from the Stratford Shakespeare Festival of Canada, in which tiny photographs from the various productions are arranged to form a composite image of Shakespeare himself (see Figure 1).

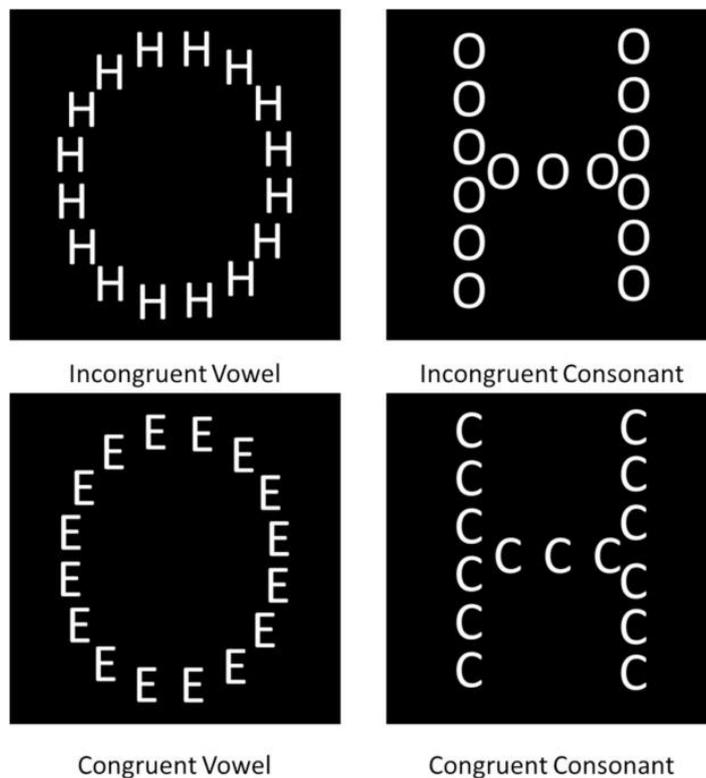
Figure 1 – Picture Mosaic from the Stratford Shakespeare Festival of Canada



Source: <https://www.pinterest.com/pin/402157441701090491/>.  
Accessed on: 10/05/2021.

Studies of global-local processing in psychology have shown us that a complex relationship exists between how we perceive things individually and how we perceive their parts in a greater whole. Of particular interest is the issue of congruence and incongruence between the small and larger image, such as the Navon hierarchical letters experiment, which tests the recognition of vowels and consonants in instances where the large figure is not congruent with the smaller one (CHAMBERLAIN *et al.*, 2017, p. 249):

Figure 2 – Navon Selective Attention Task



Source: Chamberlain, *et al.*, 2017.

For the examples in this study, a philosophically-based theory of modes might help us discern the relationship between local and global patterns of relationship. Both the *Medical Subject Headings* and the *Tesauro* display primarily a descriptive mode: while their structures of internal coherence are constructed with care and precision, the scope notes reveal a commitment to harmonizing the tool's terms and references with a reality that lies beyond it. In the case of *MeSH*, the terms refer to a body of knowledge dominated by biological-medical empirical research. In the case of the *Tesauro*, the terms refer to cultural concepts

rather than the empirically-perceived physical world; however, the descriptive mode is no less prominent.

*The Library of Congress Subject Headings*, on the other hand, illustrate a primarily conceptual mode. While the Library attempts to use accurate and up-to-date terminology, the scope notes suggest that the Library's collection prevails over the outer world that the collection: the Library is using scope notes, at least in these examples, to disambiguate authorized headings with overlapping meanings. And the Used-For terms, in addition to providing terms that can currently be considered "equivalent," also serve to maximize access to the Library's historical collection through the inclusion of outmoded terms.

What, then, is the figure in the design in each of these tools? What is the global image, or the substance, that emerges from the modes of these tools? In the case of *MeSH*, we would say that the substance is the bio-medical research. In *TFPCB*, the substance is Brazilian popular culture throughout the country's history. In the case of *LCSH*, we would say that the substance is the Library's own collection.

If we examine the concept of "mode" in musical terms, further possibilities present themselves. A mode is a prescribed arrangement of notes that provides the building blocks for a variety of melodies. However diverse the melodies might be, their conformance to a particular mode lends them a powerful effect on the listener, not just on their own merits, but through their signification of a larger whole with distinct characteristics (KOLINSKI, 2000).

The larger whole conveyed by a particular mode could, for instance, express an allegiance to a specific community or a specific set of beliefs. The tightly-controlled modes of Gregorian chants, for instance, demonstrated adherence to Christian faith, while the Indian raga, which permitted free improvisation, often expressed an individual's exploration of a personal relationship with a deity (POWERS, 1958, p. 451).

The mode could be an expression of a community-held sentiment, in which the composer or singer offers up a personally-felt expression of a communal experience. The Jewish nigun mode, for instance, expresses the sentiments of exile: “[the nigun] aims to express the unexpressible, to give voice to that which is too intimate to be uttered in words . . . There is a mystic idea that also tunes are in exile, and may be liberated by leading them back to serve a holy purpose” (AVENARY, 1964, p. 61).

Most important for our purposes, modes can be abused: manipulated for their evocative power to inspire false expectations and inappropriate emotions in the listener. Plato, in *The Republic*, urges that the ideal society banish modes that drain citizens of courage and energy, such as the “wailing” Lydian mode and the “slack” Ionian mode (PLATO, 1968, p. 77). Vaughan Williams’s use of modes such as the Dorian gave his treatments of English folk music a unique and haunting quality. However, some music theorists have argued that in so doing Vaughan Williams and other composers of the English Folksong Revival movement misrepresented English folk music by deliberately using older modes: “Holding up the modes as the defining characteristic of ‘true’ folk songs . . . the revivalists essentially denied that folk song was of the *urban* working class and so at a single stroke devalued and marginalized the culture of an urban proletariat that they unconsciously sought to control” (ONDERDONK, 1999, p. 611).

If modes can be deceptive in the way they evoke the larger body that they partially reveal, perhaps modal theory can be used to help us understand why the controversies around neutrality and bias persist in knowledge organization tools. When the Library of Congress first altered its syndetic references in the Subject Headings, changing them from “SEE” and “SEE ALSO” references to “BT,” “NT,” “RT” and “UF” references, Mary Dykstra accused the Library of disguising the subject headings as a thesaurus (DYKSTRA, 1988, p. 42). Thesaural syndetic references are a distinct aspect of the tool’s conceptual mode: carefully-structured scaffolding that gives the thesaurus its internal coherence and

order. By adopting the same reference style, Dykstra argued, the Library of Congress was evoking the aura of thesaural coherence, without actually providing that coherence. A modal theory might provide a means of exploring that discordance between the expectations that a tool implicitly inspires and the results that a tool can actually deliver.

At the end of his treatment of political and apolitical epistemologies in knowledge organization, Hjørland challenges the field to adopt more hermeneutical and reflective methods of inquiry: “the most politicized epistemologies may be those that do not realize that there can be no such thing as a non-political epistemology: it is better to have explicit subjectivity than subjectivity disguised as objectivity” (HJØRLAND, 2020, p. 474). This paper accepts that challenge in a small way. I suggest that a theory of modes, adapted to the field of knowledge organization, enhances our vocabulary for understanding and articulating the hidden patterns in both our tools and our assumptions. Such a theory might also help us recognize how the cognitive authority that our tools inspire may be misplaced, and that we are creating and presenting tools within a modal context that arouses expectations that we cannot satisfy.

## 7 ACKNOWLEDGEMENTS

I wish to acknowledge the generous assistance of Nair Hassan, Portuguese language instructor, and Helen Jacob-Stein of the Royal Conservatory of Music, in formulating the ideas in this paper.

## 8 REFERENCES

AMERICAN PSYCHOLOGICAL ASSOCIATION. **What Is Mental Illness?** *S.d.* <https://www.psychiatry.org/patients-families/what-is-mental-illness>. Accessed on: 7 Dec. 2021.

AVENARY, Hanoch. The Hasidic Nigun—Ethos and Melos of a Folk Liturgy. **Journal of the International Folk Music Council**, v. 16, Cambridge University Press, Jan. 1964, p. 60–63. Available on: <https://doi.org/10.2307/835078>.

BEGHTOL, Clare. Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents. **Journal of Documentation**, v. 42, n. 2, Jan. 1986, p. 84–113. Available on: <https://doi.org/10.1108/eb026788>.

CAMPBELL, D. Grant. Centripetal and Centrifugal Forces in Bibliographic Classification Research. **Proceedings of the 13th ASIS SIG/CR Classification Research Workshop**, 2002. p. 8–15.

CENTRO NACIONAL DE FOLCLORE E CULTURA POPULAR. **Tesouro de Folclore e Cultura Popular Brasileira**. Available on: <http://www.cnfcp.gov.br/tesouro/apresentacao.html>. Accessed on: 12 Dec. 2021.

CHAMBERLAIN, Ruth van der Hallen; REBECCA, Hanna Huygelier; CRUYS, Sander van de; WAGEMANS, Johan. Local-Global Processing Bias Is Not a Unitary Individual Difference in Visual Processing. **Vision Research**, v. 141, Dec. 2017, p. 247–57. Available on: <https://doi.org/10.1016/j.visres.2017.01.008>.

CUTTER, Charles Ammi. **Rules for a Dictionary Catalog**. U.S. Government Printing Office, 1904.

DYKSTRA, Mary. LC Subject Headings Disguised as a Thesaurus. **Library Journal**, v. 113, n. 4, 1988, p. 42–46.

FRYE, NORTHROP. **Anatomy of Criticism: Four Essays**. Princeton: Princeton paperback ed., Princeton University Press, 1971.

FRYE, Northrop. **Words with Power: Being a Second Study of “The Bible and Literature.”** Penguin Books, 1990.

GARBER, Daniel. **“Mode.” The Cambridge Dictionary of Philosophy**. Second edition. Cambridge: Cambridge University Press, 1999.

GARRETT, Donald. **The Indiscernibility of Identicals and the Transitivity of Identity in Spinoza’s Logic of the Attributes**. Spinoza’s Ethics: aCritical Guide, edited by Yitzhak Y. Melamed: Cambridge University Press, 2017. p. 12–42. Available on: <https://doi.org/10.1017/9781316339213.002>. Accessed on: 12 Dec. 2021.

HJØRLAND, Birger. “Political Versus Apolitical Epistemologies in Knowledge Organization.” **Knowledge Organization**, v. 47, n. 6, 2020, p. 461–85. Available on: <https://doi.org/10.5771/0943-7444-2020-6-461>. Accessed on: 12 Dec. 2021.

HUCKE, Helmut. Toward a New Historical View of Gregorian Chant. **Journal of the American Musicological Society**, v. 33, n. 3, p. 437–67, 1980. Available on: <https://doi.org/10.2307/831302>. Accessed on: day month year.

KOLINSKI, Mieczyslaw. “Mode”. **Encyclopedia Britannica**, 5 Oct. 2000. Available on: [academic-eb-com.proxy1.lib.uwo.ca/levels/collegiate/article/mode/110124](https://academic-eb-com.proxy1.lib.uwo.ca/levels/collegiate/article/mode/110124). Accessed on: 12 Dec. 2021.

LIBRARY OF CONGRESS. **Library of Congress Subject Headings**. Classification Web, *S.d.*. Available on: <https://classweb.org/min/minaret?table=1&menu=/Menu/&mod=Search&app=Auth>. Accessed on: 12 Dec. 2021.

NATIONAL LIBRARY OF MEDICINE. **Medical Subject Headings - Home Page**. U.S. National Library of Medicine, <https://www.nlm.nih.gov/mesh/meshhome.html>. Accessed on: 12 Dec. 2021.

ONDERDONK, Julian. Vaughan Williams and the Modes. **Folk Music Journal**, v. 7, n. 5, English Folk Dance + Song Society, p. 609–26, 1999.

PLATO. **The Republic of Plato**. Translated by Allan Bloom. *S.l.*: Basic Books, 1968.

POWERS, Harold S. Mode and Raga. **The Musical Quarterly**, v. 44, n. 4, Oxford University Press, p. 448–60, 1958.

RAMOS, Danilo; MELLO, Eduardo K. S. Communication of Emotions in Music between Brazilian Composer, Performers, and Listeners - ProQuest. **Psychomusicology: Music, Mind and Brain**, v. 31, n. 1, Mar. 2021, p. 1–17.

SPINOZA, Benedictus de. **Ethics**. Oxford: Oxford University Press, 2000.

SVENONIUS, Elaine. **The Intellectual Foundation of Information Organization**. London: MIT Press, 2009.

SZOSTAK, Rick. **Basic Concepts Classification (BCC)**. 2019. Available on: <https://www.isko.org/cyclo/bcc..> Accessed on: 12 Dec. 2021.

WORLD HEALTH ORGANIZATION. Traditional, Complementary and Integrative Medicine. 2021. Available on: <https://www.who.int/westernpacific/health-topics/traditional-complementary-and-integrativemedicine>. Accessed on: 12 Dec. 2021.

# 2

Brisa Pozzi de Sousa

**O assunto, a representação:**  
uma análise com foco na variação  
terminológica a partir da série *Estudos  
Avançados em Organização  
do Conhecimento* (ISKO Brasil)

*The subject, representation:*  
an analysis with focus on terminological  
variation from the series *Advanced  
Studies in Knowledge Organization (ISKO Brazil)*

DOI: 10.31560/pimentacultural/2022.95613.2

### Resumo:

O entendimento a respeito da representação de assunto torna-se primordial na atualidade, sobretudo, pelo constante fluxo e acesso à informação, situação que envolve os conteúdos que estão em documentos de diferentes formatos, como também, na web. É possível notar que aspectos relacionados à representação de assunto tornaram-se, de fato, uma questão estratégica evidenciada pelos formatos de metadados. A organização e o acesso à produção de informação em diferentes documentos, variando os suportes e lugares de armazenamento, desde a forma impressa, como em livros, até o que pode estar na web em repositórios digitais ou, qualquer outro sistema de informação, são preocupações da Organização do Conhecimento. Considerando esse contexto, o propósito da investigação é conjecturar variações terminológicas sobre a representação de assunto. O objetivo é explorar quais termos são abordados em estudos de Organização do Conhecimento nacionalmente. Busca-se, nas publicações dos eventos da Sociedade Brasileira para Organização do Conhecimento (ISKO Brasil), denominada *Série Estudos Avançados em Organização do Conhecimento*, atualmente composta por cinco volumes, artigos que apresentam discussões sobre a representação de assunto, a fim de compreender como e quais termos a área contempla. O *corpus* de análise desse estudo foi constituído por nove artigos que explicitaram terminologicamente em seus conteúdos a expressão representação de assunto. Pela análise realizada do *corpus*, constatou-se que a expressão representação de assunto, segundo as publicações da ISKO Brasil, tem como elemento recorrente o processo de indexação, que envolve o uso de diferentes instrumentos – inclusive os de classificação – e geram produtos variados. Entretanto, é preciso considerar a existência das diversas abordagens teóricas, construídas em momentos históricos e culturais distintos e que estão presentes na literatura.

**Palavras-Chave:** Representação de assunto; assunto; variação terminológica.

**Abstract:**

*The understanding of the subject representation becomes essential nowadays, especially due to the constant flow and access to information, a situation that involves the contents that are in documents of different formats, as well as on the web. It is possible to notice that aspects related to the subject representation became, in fact, a strategic issue evidenced by the metadata formats. The organization and access to the production of information in different documents, varying the supports and storage places, from the printed form, such as books, to what can be on the web in digital repositories, or any other information system, is a concern of the Knowledge Organization. Considering this context, the purpose of the investigation is to conjecture terminological variations on the subject representation. The objective is to explore which terms are covered in Knowledge Organization studies nationally. The publications of the events of the Brazilian Society for Knowledge Organization (ISKO Brazil), called Advanced Studies in Knowledge Organization, currently composed of five volumes, are searched for articles that present discussions on subject representation, in order to understand how and which terms the area contemplates. The corpus of analysis of this study consisted of nine articles that terminologically explained in their contents the expression subject representation. Through the analysis of the corpus, it was found that the expression subject representation, according to ISKO Brazil publications, has as a recurrent element the indexing process, which involves the use of different instruments – including classification ones – and generate different products. However, it is necessary to consider the existence of different theoretical approaches, built in different historical and cultural moments, and that are present in the literature.*

**Keywords:** Subject representation; subject; terminological variation.

## 1 INTRODUÇÃO

Nos dias de hoje, a palavra assunto pode se relacionar a diferentes contextos, como, por exemplo, “o assunto que está na mídia”, “um assunto da vida particular”, “ir a um buscador e recuperar um assunto”, “o assunto de um e-mail”, dentre uma infinidade de possibilidades. Já assunto, com enfoque nos estudos que relacionam a Organização do Conhecimento, assume diferentes aspectos que ajudam a pensar a sua manipulação, mas deve ser entendido com as perspectivas teóricas que o sustenta e não isoladamente.

Este trabalho lança olhar para o assunto e sua manipulação pela representação, com foco nas variações terminológicas. Embora qualquer reflexão envolva um recorte teórico para o entendimento de um quadro referencial, no caso da presente investigação, o assunto, ao recorte temporal, juntam-se elementos que integram, mas também delimitam a pluralidade do que se quer compreender. Olhar para as terminologias de qualquer área, em um determinado período, ajuda a entender como elas se expressam por meio de linguagem própria, o que, por sua vez, permite o fortalecimento da sua identidade.

De forma ampla, conjecturar as variações terminológicas sobre a representação de assunto é o propósito da investigação. De forma mais específica, explorar quais termos são abordados em estudos da Organização do Conhecimento nacionalmente, compõe o objetivo do presente trabalho. Busca-se, nas publicações dos eventos da Sociedade Brasileira para Organização do Conhecimento (SKO Brasil), denominada *Série Estudos Avançados em Organização do Conhecimento*, atualmente composta por cinco volumes (dos respectivos anos: 2012, 2013, 2015, 2017 e 2019), artigos que apresentam discussões sobre a representação de assunto, a fim de compreender como e quais termos a área contempla.

Seguindo esse objetivo, a ideia é identificar o que a comunidade ISKO Brasil tem tratado ao abordar a manipulação do assunto, ou seja, a representação e sua respectiva variação terminológica, com o propósito de identificar resultados que auxiliam o entendimento sobre o que é sinalizado quando se fala de assunto.

## 2 DELIMITAÇÃO SOBRE ASSUNTO

Não há uma única abordagem sobre o que seja assunto, pois sua compreensão varia de acordo com a influência teórica considerada. Mas, tendo em vista a importância da sua manipulação, que resulta na representação temática, ou seja, na representação de assunto, é fato visualizar as diferentes vertentes que se caracterizam de forma geográfica a partir de duas principais influências: anglo-saxã e europeia. Nesse contexto, o acúmulo de saber da Biblioteconomia fornece subsídios para a compreensão de aspectos significativos ao assunto.

De forma geral, pensar a representação de documentos em ambientes informacionais, como bases de dados, catálogos, repositórios, dentre outros, visa à recuperação do registro que, por sua vez, remete ao documento, possibilitando seu acesso. Vários atributos podem ser representados em um registro, que é composto por partes e, para sua elaboração, uma estrutura deve ser seguida de acordo com a política informacional institucionalizada. A representação de assunto em um registro resulta em um conjunto de termos que exercem a função de ponto de acesso a esse registro. Há mais de um século, foi proposta a entrada por assunto em fichas catalográficas compondo um lugar no catálogo. Com o registro on-line, a função da entrada por assunto continuou relevante, pois o contexto automatizado dos computadores não dispensa os fundamentos anteriores de organização da informação, mas reafirma a necessidade de resgate

do que foi construído ao longo do tempo e olhar para isso reforça a compreensão sobre como a área está pautada.

Usando como exemplo os catálogos de bibliotecas, por vezes, foram compreendidos como ferramentas para inventariar acervos, mas, na Antiguidade e em grande parte da Idade Média, a história da Catalogação mostra que não é explícita a diferença entre catálogos e inventários. Foi, no final do século XV, pelo avanço das atividades bibliográficas e, a partir do século XVIII em diante, que o catálogo passou a ser uma ferramenta de busca e recuperação. Esse contexto é permeado pelo desenvolvimento e crescimento de bibliotecas no contexto europeu.

Nesse desenvolvimento, a produção dos registros é realizada pelos aspectos do assunto abarcado pelo processo de catalogação. A Catalogação inclui a Catalogação Descritiva e a Catalogação de Assunto, as quais respondem pela construção completa do catálogo. No entanto, segundo Guimarães (2009), o aspecto temático (ou do assunto) envolve mais de uma perspectiva: além da Catalogação de Assunto, existe a Indexação e a Análise Documentária.

Vertentes constituíram-se com suas respectivas abordagens, mas elas não existem separadamente, tornando necessária a reflexão sobre o que é mais consistente e funcional, haja vista que essas diferenças resultam na formação de conceitos da área. Sem esse entendimento, além de uma confusão conceitual, os fundamentos acabam restringidos a um patamar de processamento técnico, pecando contra a potencialidade de instaurar uma efetiva comunicação entre documentos e usuários(as).

Contemporaneamente, no Brasil, para a representação da informação, são adotados os termos “representação descritiva” e “representação temática”, que indicam as etapas para a construção dos registros de informação, quanto aos processos específicos em causa. Por exemplo, a primeira ocupa-se dos aspectos formais que permitem identificar o documento, a qual também pode ser chamada de catalogação

descritiva ou somente catalogação. Já a segunda opera com os processos de indexação e elaboração de resumos. Mas, em alguns currículos dos cursos de Biblioteconomia no Brasil, por exemplo, é possível encontrar os termos Representação Descritiva e Representação Temática de forma articulada, respectivamente, com a Catalogação e com a Indexação. No entanto, alguns instrumentos, como os sistemas de classificação, podem compor os currículos enquanto Sistemas de Organização do Conhecimento (SOCs), mas esse termo é empregado para abordar várias linguagens (como as de indexação e as de classificação). Nesse sentido, é preciso considerar as diferenças entre os processos e os instrumentos.

Importante ressaltar que a classificação de documentos trabalha com vários atributos, dentre eles, o assunto. A diferença entre classificação e indexação pode ser evidenciada pela constatação de que a primeira é um método para a ordenação ou arranjo de documentos, do qual decorre que vários de seus atributos são considerados; já, na indexação, opera-se com a atribuição de conteúdos temáticos (assuntos) ao documento. Desse modo, os produtos também são distintos: notações classificatórias (a depender do sistema adotado) e termos de indexação, respectivamente.

Ainda relativo ao contexto brasileiro, o termo “representação temática” também é adotado de forma ampla para referir-se aos processos de indexação e elaboração de resumos. A despeito da complexidade teórica e prática, a construção de um registro de informação gera representações que referenciam o documento por diferentes aspectos, dentre eles o temático. Assim, na indexação, pontos de acesso por assunto remetem ao registro que, por sua vez, referenciam os documentos representados. Na mesma medida de importância, um texto objetivo e conciso do conteúdo do documento pode ser elaborado no registro, sendo chamado de resumo, com o propósito do(a) usuário(a) selecionar documento(s) que gostaria de acessar. Eventualmente, as palavras do resumo podem compor o índice de busca, mas ele não existe em função disso.

Com essa breve exposição, é possível observarmos diferentes nuances que envolvem o entendimento sobre assunto e, por consequência, sua representação. Na mesma medida, há variações terminológicas em jogo.

Compreender os termos empregados por pesquisadores e pesquisadoras nos eventos da Sociedade Brasileira para Organização do Conhecimento (ISKO Brasil), para caracterizar uma reflexão conceitual sobre a representação de assunto, é uma forma de contribuição. Dito de outra forma, busca-se compreender como pesquisadores e pesquisadoras do capítulo brasileiro da ISKO relacionam a representação de assunto com as possíveis variações terminológicas.

Fato é que não há uma única abordagem sobre o que seja o assunto. Sua manipulação ocorre através da representação e sua compreensão varia de acordo com a influência teórica considerada.

Tendo em vista a importância dessa representação, as ideias serão evidenciadas pela “decomposição do assunto” a partir da influência do *Classification Research Group* (CRG). Considera-se que os estudos do referido grupo englobam o que é possível entender como desdobramento do pensamento anglo-saxão no século XX, ou seja, o desdobramento da vertente Catalogação de Assunto. Esse desdobramento é considerado, pois, seguindo um recorte geográfico para delimitação do entendimento a ser alcançado, estudos iniciais no contexto anglo-saxão do século XIX marcaram a interseção de teóricos como Anthony Panizzi, Charles Jewett e Charles Cutter. Entretanto, para fins de recorte teórico, geográfico e temporal e pelo limite de páginas que esse trabalho requer, a vertente Catalogação de Assunto não será explorada.

Para contextualizar a variação terminológica em jogo, a perspectiva da Indexação faz parte do recorte bibliográfico. Não há pretensão de cobrir a totalidade de autores/publicações do CRG, mas de lançar olhar para a variação terminológica que envolve a representação de assunto, que, por consequência, esbarra nas variações terminológicas sobre assunto.

### 3 O ASSUNTO, A REPRESENTAÇÃO: DO DESCRITOR A OUTROS TERMOS

As apropriações efetivas de terminologias de qualquer área envolvem esforços intelectuais do seu coletivo. Assim, vários pesquisadores e pesquisadoras trabalharam e trabalham na tentativa de fornecer escopo científico a elementos teóricos que são basilares.

Para tratar do escopo científico sobre o assunto, na perspectiva da Indexação, consideramos Foskett (1973), que cita Eric J. Coates, o sistema PRECIS (*PREserved Context Indexing System*), de Derek Austin e Jason Ferradane, como referenciais contemporâneos que envolveram a abordagem temática da informação.

Coates, em 1960, publicou o livro *Subject Catalogues Headings and Structure*, que abordou a ordem dos cabeçalhos compostos e buscou compreender a razão que levou Kaiser a optar pelo concreto ao invés do processo como ponto de entrada dos cabeçalhos. Seu desafio consistiu em buscar a aplicação da sua teoria no *British Technology Index*, pois foi o redator responsável do projeto. Coates estabeleceu a ordem de importância das entradas pela fórmula Coisa-Parte-Material-Ação.

O PRECIS foi idealizado por Derek Austin, a partir de 1968, e consistiu na mecanização da produção de índices em inglês, como também em outras línguas, caracterizando-se por ser um sistema de indexação pré-coordenado, feito para produzir automaticamente os índices de assunto da *British National Bibliography* (FUJITA, 1988). Nesse sentido, o avanço “[...] que o PRECIS [...] [representou foi a] elaboração de um enunciado básico que pode ser processado pelo computador, a fim de proporcionar um conjunto completo de cabeçalhos.” (FOSKETT, 1973, p. 62). O sistema foi elogiado por propiciar ao usuário visualizar de forma completa cada cabeçalho.

Já Farradane, segundo Foskett (1973), elaborou um sistema de operadores relacionais embasado em conceitos sobre o processo de aprendizagem. Com base na psicologia infantil, adotou nove operadores relacionais que vinculava termos através de aspectos delimitados no tempo e no espaço.

O movimento que congrega esses autores é o de institucionalização dos estudos de Ranganathan, na Inglaterra, a partir de 1952, realizado pelo *Classification Research Group* (CRG). O grupo contou com a participação de inúmeros pesquisadores, dentre eles Derek Austin (Inglaterra), Eric Coates (Estados Unidos), Jason Farradane (Canadá), Ingetraut Dahlberg (Alemanha), dentre outros nomes (FUJITA, 1988).

Para Metcalfe (1973), Farradane teria introduzido o isolado de Ranganathan como um termo técnico, pois, em 1950, ele apresentou o que chamou de teoria científica da classificação e suas aplicações práticas. Já Coates, no *British Technology Index*, ainda segundo Metcalfe (1973), considerou o artigo de Ranganathan<sup>1</sup> sobre cabeçalho de assunto e análise de facetas (discutido por Hjørland [2017]), mas não incluiu qualquer terminologia de Ranganathan em seu livro (*Subject Catalogues Headings and Structure*). Todavia, Coates abordou Melvil Dewey (criador da CDD), Cutter e Kaiser como autores que foram anteriores a Ranganathan, no sentido de ter havido uma continuidade das discussões entre eles, apesar dos diferentes aspectos que os distanciam e as particularidades teóricas que os sustentam. Mas, foi em Kaiser que Coates viu a solução “[...] do problema da importância relativa dos componentes de um assunto composto no ponto em que Cutter tinha deixado” (METCALFE, 1973, p. 324, tradução nossa).

Em sentido amplo, a partir de Guimarães (2008, p. 80), é possível inferir que o quadro teórico apresentado indicou

1 RANGANATHAN, S. R. **Prolegomena to library classification**. Bombay: Ásia Publishing House, 1961.

[...] as experiências investigativas de Kaiser (EUA, 1911) relativamente à composição do binômio concreto/processo para a estruturação de enunciados de assunto e o trabalho teórico de Ranganathan (Índia, 1933) quanto a uma análise em facetas – Personalidade, Matéria, Energia, Espaço e Tempo – do conteúdo temático dos documentos.

Igualmente há de se registrar as experiências investigativas do *Classification Research Group* (Inglaterra, 1952), que ampliou para doze as cinco categorias fundamentais de Ranganathan, e de Derek Austin (Inglaterra, 1968) e do grupo de Bangalore (Índia, 1969) que, por meio dos sistemas PRECIS e POPSI, lançaram as bases para uma aplicação da análise facetada em sistemas automatizados.<sup>2</sup>

Embora haja outros referenciais que tenham contribuído com diferentes discussões no contexto temático, o ponto em comum foi a transição de algo explícito no documento com fins de organização dos acervos. A decomposição e os possíveis relacionamentos entre os assuntos foram percebidos à medida que o contexto linear de sintaxe dos cabeçalhos (de assunto) não deram conta de conferir sentido à padronização das formas de entrada.

Por outro lado, os cabeçalhos de assunto não se tornavam eficazes, pois sua elaboração foi marcada pelo aspecto morfológico da palavra. Taylor (1995, p. 484, tradução nossa) exemplifica que “[...] o conceito de busca por assunto permaneceu com má reputação”. A autora relata que, na conferência da *American Library Association* (ALA), em 1994, houve um movimento que contestou a ênfase dada ao uso da *Library of Congress Subject Headings* (LCSH).

Taylor (1995) ainda menciona que Ray R. Larson apontou, em artigo, na mesma conferência da ALA, a obriedade sobre o declínio nas porcentagens de uso do índice de assuntos da LCSH e a compensação das

2 POPSI (*POstulated-based Permuted Subject Indexing Language*) originou-se de estudos realizados, a partir de 1969, por A. Neelamegham e colaboradores no *Documentation Research and Training Center* de Bangalore, Índia. Neelamegham fez parte do *Classification Research Group* (FUJITA, 1988).

buscas realizadas no índice de palavras-chave por título. A autora continua explanando que, desde 1985, a LC busca reestruturar regras para a LCSH, como, por exemplo, de caminhos hierárquicos para os termos dos cabeçalhos. Em 1991, a LC realizou uma conferência com o propósito de cooptar sugestões sobre como poderiam melhorar o sistema existente de subdivisões da LCSH e a ação resultou em mudanças na referida lista.

A discussão que Taylor (1995) propõe envolver um olhar permanente para o registro de assunto no catálogo, visando à busca por conteúdo temático dos documentos pelos usuários. Não se falou na possibilidade de se mudar o instrumento, mas isso não era usual, pois seria uma demanda de alto custo.

Ademais, presume-se que o contexto propiciou demanda por linguagens pós-coordenadas. Foskett (1973) refere-se que o tipo de linguagem de indexação reflete o resultado do processo na saída da informação para o usuário, mas também na análise para a entrada do assunto. As listas de cabeçalhos de assunto são linguagens pré-ordenadas, pois os termos de busca empregados no sistema de recuperação prescrevem a combinação. Já uma linguagem pós-coordenada permite a combinação de termos no sistema de recuperação, como acontece com o uso dos tesauros. Nota-se que o assunto no sistema de recuperação evidencia-se pelo controle do vocabulário na forma apresentada pela linguagem.

No mínimo, dois caminhos podem ser apontados em relação ao aspecto da representação de assunto: um que visa coletar termos para construir um vocabulário controlado, portanto uma linguagem; e outro que envolve a sistematização de aspectos metodológicos da indexação de assuntos. Porém, ainda com a estrutura das listas de cabeçalhos de assunto, sobretudo da LCSH, sustentando entre os cabeçalhos apenas relações de sinonímia (*see – ver*) e de associações (*see also – ver também*), predominou a necessidade de mecanismos que dessem conta de aperfeiçoar o sistema de recuperação com vínculo semântico.

Simões (2017, p. 166) afirma que “as remissivas foram-se alterando ao longo do tempo, quer no que respeita à nomenclatura, ao princípio e à função.” A autora alerta que a nomenclatura alterou-se para relações semânticas.

Mas, as relações para um vocabulário envolvem um controle, que

[...] não significa apenas padronizar as formas de entrada (significantes) ou registrar a sinonímia: significa operar sobre a significação (significante/significado) dos termos, o que depende inteiramente de delimitações de sentido conferidas pela rede de relações lógico-semânticas amparadas por definições (LARA, 2001, p. 6).

Com a difusão do uso do computador e das fichas perfuradas, Calvin Mooers, na década de 1950, introduz o uso do termo descritor (*descriptor*). Fundou a *Zator Company* e desenvolveu um sistema próprio de recuperação da informação, o *Zatocoding*, no contexto dos estudos de *information retrieval* (IR).

Segundo Mooers (1951, p. 21, tradução nossa), seu sistema foi de “[...] especial importância para o documentalista porque lhe dá o poder de uma nova ferramenta de especificação de ideias para a recuperação de informações armazenadas.” Seu propósito foi de agilizar a recuperação por assunto, através da sobreposição de padrões aleatórios de códigos no registro. Esses padrões de código formavam a representação de descritores por ideias, ou seja, o conhecimento sobre o assunto do documento a ser registrado em informação.

Em linhas gerais, o sistema operava através da

[...] memória de filme com o campo de codificação de cada quadro possuindo 216 posições que podem ser marcadas por opacidades. Assumindo uma coleção de 5.000.000 documentos – comparáveis à Biblioteca do Congresso – um padrão do *Zatocoding* de 8 pontos por assunto pode ser usado. O campo de codificação pode conter até 18 padrões de assunto, e

seleções podem ser feitas em qualquer combinação desses padrões. O tamanho do vocabulário descritivo do *Zatocoding* é ilimitado (MOOERS, 1951, p. 22, tradução nossa).

Nesse panorama, o tratamento empregado pelo sistema *Zatocoding*, através das fichas perfuradas, usava as suas margens e máquinas seletoras que, a partir de uma lista de codificação, registrava os assuntos. O padrão da codificação proporcionava a recuperação pelo código numérico de cada descritor.

O emprego do termo “descritor” por Mooers ocorreu para uma aplicação específica do seu sistema, com funcionamento independente, mas ele entendia que os “[...] descritores [...] constituem as melhores unidades conceituais possíveis (científica ou tecnologicamente) para o trabalho específico de recuperação” (MOOERS, 1985, p. 253, tradução nossa). De acordo com Henderson (1996, p. 660, tradução nossa),

Mais tarde Mooers observou que o neologismo, “descritor”, sofreu um triste destino: o termo foi introduzido para “sintetizar um conceito cuidadosamente elaborado no uso de unidades de significado para recuperação”, mas veio a ser usado como sinônimo de entradas de índice de qualquer tipo. Por exemplo, desde a publicação dos anais da Conferência Internacional sobre Informação Científica, realizada em Washington, D.C., em 1958, o termo “descritor” foi equiparado a “termos de índice” e “palavras-chave” no índice dos anais. Na visão de Mooers, isso era “a maior corrupção semântica.”

No entanto, conforme exposto por Henderson (1996), o termo “descritor”, empregado por Mooers, inicialmente na concepção do seu sistema de recuperação, baseado em um processo mecânico, perdeu-se em meio ao emprego de outros produtos documentários. De fato, Mooers demonstrou interesse sobre métodos mecânicos de recuperação por assunto em coleções de documentos, como demonstram suas investigações sobre recuperação da informação.

À medida que sua investigação deflagrou interesse, conforme trabalho apresentado no ano de 1962, na *University of Minnesota*, mas reimpresso em 1985, Mooers abordou a linguagem de indexação em sistemas de recuperação da informação. Para ele, a linguagem de indexação é o meio responsável por mediar as informações armazenadas nos documentos, tornando-se “[...] a ponte, a conexão entre os usuários e as informações” (MOOERS, 1985, p. 249, tradução nossa). Ele demonstrou interesse por essas linguagens, mas centrou-se em entender como as máquinas poderiam manipulá-las.

Para esse autor, uma lista de cabeçalho de assunto, por exemplo, determinaria um arquivo científico de palavras com sintaxe específica, conforme “*Lubrication – high temperature – non-hydrocarbon*”. Explica que esse cabeçalho possui diferentes conceitos unitários, como “*lubrication*”, “*high temperature*”, “*non-hydrocarbon*”, que permitem outras combinações, mas que ele chamou de descritor. Propõe o descritor como composto por duas partes: um rótulo ou etiqueta (*label*) e a definição (*definition*), que pode ser um verbo ou uma palavra. A partir do rótulo e da definição, considerou alguns princípios básicos úteis à recuperação da informação.

Segundo Mooers (1985), a linguagem de indexação auxiliaria o isolamento dos conceitos através de um rótulo (*label*), exemplificando com o termo “*lubrication*” que pode ser um *descriptor label* (rótulo do descritor), mas que também precisaria, segundo ele, de um significado. Por sua vez, esse significado relacionado com o sistema de recuperação não estaria baseado em definições de dicionários, pois “[...] toda organização ou laboratório com um sistema de recuperação de informações tem seu ponto de vista especial sobre cada tópico ou conceito de importância para ele. Portanto, este ponto de vista especial deve ser escrito na definição do rótulo.” (MOOERS, 1985, p. 253, tradução nossa).

Seria possível, a todo descritor, a inclusão de sua definição anexada ao rótulo como uma ideia pronta, o que o autor denominou ideia empacotada (*packaged idea*) “[...] como um conceito que deve ser

usado principalmente para recuperação em algum sistema de recuperação específico." (MOOERS, 1985, p. 253, tradução nossa). Com isso, haveria a constituição de uma lista, com rótulos e definições acopladas a um conjunto de seis características definidas.

A primeira característica refere-se ao vocabulário com os rótulos associados a um conjunto pequeno de descritores (250 a 350), sendo que, em casos especiais, esse número poderia ser maior, chegando a mil. A segunda é a do uso de um descritor ou mais para formar um conjunto a ser adotado na caracterização dos documentos. A terceira é a de que cada descritor seria definido por uma nota de escopo (*scope note*). A quarta é a necessidade de sutileza, segundo denominação de Mooers, pois o conjunto de descritores, como também cada um deles individualmente, deveriam ser adaptados para um trabalho específico de recuperação. A quinta característica é a de que seria desejável organizar o conjunto de rótulos dos descritores em uma folha de papel, com o propósito de programar e exibir, agrupar ou classificar; a classificação serviria apenas para individualizar os descritores, não implicando uma hierarquia que, na visão de Mooers, seria incômoda e deveria ser evitada. A sexta e última característica é a escolha dos descritores para a busca. Ao escolher um, dois ou três descritores, seria possível usá-los de forma concomitante, como um "conjunto prescrito" (*prescribing set*), com a finalidade de conduzir a busca pela informação desejada. A pesquisa, considerando a prescrição de descritores, aprimoraria a busca dos documentos, tornando-a mais específica.

De acordo com Mooers (1985), seu sistema não tinha equivalência com o *Uniterm* (Unitermo), criado por Mortimer Taube, em 1952, o qual adotou como método de indexação a extração de palavras diretamente do texto. Para o autor, a maior dificuldade do Unitermo seria o controle de sinônimos, pois o sistema baseou-se em palavras ao invés de conceitos. Da mesma forma, as duas propostas não se aproximavam do tesouro, em função da ausência de relacionamentos entre os termos, controle terminológico, definições, dentre outros aspectos a serem considerados.

De forma geral, o trabalho de Mooers relacionou a linguagem de indexação à recuperação da informação, pois considerou os descritores a partir da coleção. O descritor foi o fundamento para distinguir o que ele propunha e que não era idêntico ao contexto de outros sistemas, como aqueles constituídos por cabeçalhos de assunto, palavras-chave, dentre outros elementos de controle. Mas, seu interesse primeiro foi o de subsidiar um sistema que tivesse a capacidade de melhorar a recuperação da informação.

Considerando o exposto, é possível notar que a busca por sistematização de aspectos metodológicos, com base no trabalho de Mooers, tem relação direta com a representação de assunto para recuperação. Entretanto, a repercussão do termo descritor marca de forma significativa a Indexação. Compreender os termos que compõem a constituição científica de uma área é um contínuo exercício de revisão dos seus fundamentos.

Na subseção seguinte, o foco está na ampliação de termos em torno do assunto e, por consequência, na variação terminológica que deflagra dispersão de abordagens. Subjacente a esta variação terminológica encontram-se aspectos distintos elaborados em função de abordagens diferenciadas sobre o mesmo processo: a representação do assunto.

### **3.1 VARIAÇÃO TERMINOLÓGICA: *ABOUTNESS*, *TOPIC*, *CONTENT*, *SUBJECT* – ASSUNTO, TEMA, MATÉRIA – CONTENIDO, MATERIA EPÍGRAFE, TEMA**

Na década de 1970, discussões relacionadas à análise de documentos para transformá-los em representações seguiram o que alguns pesquisadores chamaram de *aboutness*. Guedes (2009), a partir de Moraes e Guimarães (2006), explica que inicialmente foi Hutchins

(1977; 1978) o responsável em dar destaque ao termo. Por sua vez, Hutchins (1977) indica responsabilidade a Fairthorne (1969).

A preocupação inicial de Hutchins (1977, p. 17) partiu de como compreender o “tópico de um documento” (*topic of a document*) no contexto de um sistema de informação. Conforme o autor, uma resposta superficial seria entender esse tópico como assunto de uma entrada em um índice, com correspondência ao documento. Mas, a descrição do assunto simplesmente não contempla uma fórmula ou equação, tornando-se instigante alcançar “o que o documento trata” (*what the document is about*).

Segundo o autor, a linguagem documentária (*documentary language*) adotada pode ser um facilitador da “formulação de descrições de assunto” (*formulation of subject descriptions*), pois o propósito seria expressar “todo o tópico” (*whole topic*) do documento. De outro modo, o tópico de um documento (*the topic of a document*) envolve a sumarização do seu conteúdo para a recuperação em um sistema, de forma independente da linguagem documentária empregada.

Ao considerar o documento em um contexto temático específico, sua possibilidade de recuperação será abarcada por esse contexto, porém, antes das ferramentas e recursos disponibilizados, há uma descrição particular do conteúdo. Nesse entendimento, Hutchins (1977, p. 18, tradução nossa) interroga-se: “O que queremos dizer com o conteúdo de um documento? Para responder a isso, precisamos ser claros sobre a distinção entre o ‘sentido’ de uma expressão linguística e a ‘referência’ dessa expressão.”

Ao vincular o contexto linguístico, Hutchins (1977, p. 18, tradução nossa) explica que “o sentido de uma palavra determina o alcance de seus potenciais referentes.” De forma isolada, “uma palavra tem sentido, mas não tem um referente real; pode ter um referente apenas em um contexto textual específico.”

O autor traça uma distinção entre o sentido de um texto e a sua referência. Explica que o entendimento do texto não implica as referências ou o que o autor teve em mente ao escrevê-lo e, se a preocupação consistir na busca da verdade ou falsidade do que foi escrito, o conhecimento dos sentidos das expressões não se torna suficiente. Cita, como exemplo, Bertrand Russell e a condição de “rei da França é careca”. Hutchins (1977) explica não ser possível dizer se a afirmação é verdadeira ou falsa fora de uma situação referencial específica (*outside a particular referential situation*). O referente seria a França do século XVIII e, com isso, haveria a possibilidade de testar a verdade ou a falsidade da declaração. “Assim, enquanto o sentido permanece constante, a referência varia – e assim também, em consequência, o valor de verdade” (HUTCHINS, 1977, p. 18, tradução nossa). O autor completa, com base em Fairthorne (1961), que os indexadores centram atenção no sentido dos textos que “significa o conteúdo dos documentos” (*this is what is meant by the content of documents*) e, para ele, assim deve ser. Afirma que o sentido do texto independe do autor, como se alcançasse uma existência autônoma, “como parte do que Karl Popper chamou de ‘Mundo 3’” (*as part of what Karl Popper has called ‘World 3’*) (HUTCHINS, 1977, p. 18).

Com isso, o autor afirma que o problema linguístico de sumarização necessita de compreensão da estrutura do texto. Isso envolve uma progressão temática e semântica no que Fairthorne (1969), segundo Hutchins (1977), estabeleceu haver uma distinção entre o caráter extensional de um texto e sua intencionalidade. O primeiro é definido pelos tópicos de partes componentes de um texto, os tópicos de seus parágrafos, seções, capítulos etc. O último é o tópico do texto como um todo, representando algo a mais que os tópicos de suas partes.

De acordo com Beghtol (1986), para Fairthorne, o aboutness extensional (*extensional aboutness*) “é o assunto inerente ao documento” (*is the inherent subject of the document*); já o aboutness intencional

(*intensional aboutness*) é a razão ou propósito para o qual o documento foi adquirido de acordo com o interesse dos usuários. Explica que qualquer documento possui mais de um assunto, dependendo de quem o solicita, por que e para quê. Nesse contexto de forte aproximação com o assunto, ainda segundo Beghtol (1986), com base em Boyce (1982)<sup>3</sup>, os termos “*topicality*” (*aboutness*) e “*informativeness*” (*meaning*), concatenam-se ao significado do documento, que não é algo pronto e fixo, pois variará de acordo com diferentes fatores, dentre eles, sociais, culturais, institucionais etc. Esse recorte demonstra que, no *aboutness* extensional e no “*topicality*”, permanece a ideia de algo pronto, estável, enquanto, no *aboutness* intencional e “*informativeness*”, há presença de uma ideia mutável e não fixa. Cabe destacar a dificuldade de tradução dos termos para língua portuguesa.

Ademais, o objetivo de Beghtol (1986) foi aproximar o *aboutness* dos sistemas de classificação bibliográfica, pois, segundo a autora, os classificacionistas projetam os sistemas, realizam todo um trabalho conceitual e notacional, mas não consideram os significados particulares intrínsecos, uma vez que o documento muda quando um significado particular é atribuído a ele ou quando são usados de modo particular por usuários.

Segundo Lancaster (2004, p. 13), o *aboutness*, traduzido no seu livro por “*atinação*”, busca subsidiar a expressão “de que trata um documento” que, para ele, é sinônimo de “tem por assunto”. De acordo com o autor, “estas expressões talvez não sejam muito precisas e não é fácil definir ‘trata de’ [*about*] e ‘tem por assunto’ [*aboutness*]. Apesar disso, são expressões que soam aceitáveis para a maioria das pessoas, sendo por elas compreendidas.” Lancaster (2004, p. 13) ainda afirma ser difícil explicar a situação no que tange à indexação de assuntos, e autores como “Beghtol (1986) e Hutchins (1978) recorreram à linguística do

3 BOYCE, B. Beyond topicality: a two-stage view of relevance and the retrieval process. *Information Processing and Management*, 18, p. 105-109, 1982.

texto ao examinar essa questão”, porém houve autores que recorreram ao recorte probabilístico e até outros que afirmaram tratar de um aspecto “‘intratável’, visto ser tão difícil decidir do que trata um documento.”

O assunto, considerado como um dos atributos do documento, pois é uma característica que permite a estrutura do registro de informação compondo sua natureza temática, é a forma de expressar os potenciais epistemológicos do documento, porém, isso se dá conforme a ótica de quem o descreve. Nessa concepção, Hjørland (1992) considera o conceito de assunto (*subject* ou *subject matter*) em dois aspectos: idealístico subjetivo e idealístico objetivo.

O primeiro aspecto envolve percepções ou visões de um ou mais indivíduos, pois assuntos são subjetivamente compreendidos e, assim, o seu conceito reside no estudo das mentes de algumas pessoas, como, por exemplo, os autores ou usuários de documentos. “Do ponto de vista da epistemologia, o idealismo subjetivo é caracterizado por tornar a percepção e o pensamento independente de uma maneira subjetivista. O positivismo é o representante mais comum do idealismo subjetivo.” (HJØRLAND, 1992, p. 174, tradução nossa). Esse ponto de vista enfatiza a interpretação do documento, acarretando o que Hjørland (1992) nomeia ser uma concepção agnóstica do assunto (*agnostic conception of subject*), a qual leva à impossibilidade de saber como ele deve ser determinado.

O autor faz crítica ao idealismo subjetivo, pois aponta ser impossível determinar assuntos examinando a mente de autores, usuários ou qualquer outro grupo específico de pessoas. No documento há

[...] as intenções do autor, seu ponto de vista do assunto, e outros novos aspectos que a ele se relaciona. Isso deu origem ao conceito de “*aboutness*” na literatura da biblioteconomia e ciência da informação, interesse que, a meu ver, representa um beco sem saída, uma tentativa de escapar das dificuldades do conceito de assunto [...] (HJØRLAND, 1992, p. 174, tradução nossa).

O que chama atenção no extrato acima é a dificuldade sobre o conceito de assunto. O termo *aboutness*, que não possui tradução consensual para o português do Brasil, não deve ser considerado sinônimo de assunto. Situação parecida acontece no espanhol, pois se encontra dificuldade em estabelecer definição para *contenido* e *materia*, quando se refere à palavra assunto em inglês: *subject*.

Quanto ao aspecto idealístico objetivo, o assunto é enfatizado a partir da análise teórica, de um ponto de vista absoluto, independentemente da consciência humana. Nessa concepção, os documentos compartilham ideias que estão expressas em um assunto e “essas ideias existem fora da consciência humana (ou dentro dela como percepções *a priori*) e também são anteriores aos conceitos individuais expressos nos documentos individuais.” (HJØRLAND, 1992, p. 177, tradução nossa). Ainda segundo Hjørland (1992), o aspecto idealístico objetivo expressa que o assunto pode ser identificado de maneira independentemente do contexto que será empregado, mesmo a partir de um sistema de classificação.

A fim de compreender o que seja o assunto dos documentos, torna-se necessário algum tipo de análise, pois o objetivo é realizar a representação temática: “[...] a informação é procurada em bibliotecas ou bases de dados em que os documentos (transportadores de informação) são registrados por assunto” (HJØRLAND, 1992, p. 179, tradução nossa). Acontece que a construção do conceito de assunto não elimina ou descarta seu caráter intangível e, com isso, a dificuldade em defini-lo. Duas razões, segundo Hjørland (1992), constituem esse quadro: os documentos possuem infinitas propriedades e estas não podem ser estabelecidas de uma única vez. Sendo assim, apesar do registro de informação possibilitar a recuperação temática, o assunto depende da análise de fatores contextuais.

O termo “assunto”, no contexto das listas de cabeçalhos, foi bastante adotado de uma forma mais prática que teórica, como produto do

processo de catalogação de assunto. Em relação a esse termo, no Dicionário da Língua Portuguesa (2004, p. 166), assunto é definido como “matéria de que se trata; objecto; tema; motivo (Do lat. *assumptu*-, <<assumido>>).” No Dicionário de Latim-Português (FERREIRA, 1987, p. 139), a etimologia *assumptus* refere-se ao *assumo* - “1. Tomar, tomar para si, tomar consigo, associar a. 2. Adaptar, atribuir, arrogar, aplicar.”

No dicionário Aulete (on-line), sua origem também é identificada a partir do latim “*assumptus*, *a*, *um*”, mas no sentido em “que se elevou; ELEVADO”. Outra definição: “aquilo sobre o que se conversa, fala ou escreve; TEMA; MATÉRIA” (AULETE, s.d.). Indica-se primeiro sua origem latina, vinculando-se às regras da língua portuguesa, pois, pelo radical, é possível incluir, por exemplo, afixos<sup>4</sup> para formação de outras palavras. Observam-se também outros conceitos, como “tema” e “matéria”.

Faria e Pericão (2008, p. 111), no Dicionário do Livro, definem assunto como “aquilo que, numa obra literária, constitui o conteúdo de pensamento sobre o qual se exerceu o talento criador do autor. Aquilo sobre o que se aplicou a reflexão do escritor numa obra científica ou didáctica. O tema, matéria ou argumento de uma obra, mencionado ou não no título.” Essa definição de um dicionário especializado sobre o livro, seja ele apresentado em suporte tradicional ou eletrônico, aborda o assunto relacionado ao documento desde a escrita ao conteúdo intelectual, entendido como “tema, matéria”.

Já no dicionário de Biblioteconomia, em idioma espanhol, a definição para assunto, que corresponde à palavra *materia* é: “assunto, tema ou argumento que uma obra literária é composta, etc.” (BUONOCORE, 1976, p. 298, tradução nossa). Com exceção dos dois primeiros dicionários, que não são especializados, ou seja, são dicionários de língua, as respectivas definições trazem em comum os termos tema e matéria.

4 “Os **afixos** são elementos que se agregam ao radical para modificar-lhe o significado. Os **afixos** que se antepõem ao radical chamam-se **prefixos**; os que a ele se pospõem, **sufixos**.” (CUNHA, on-line).

Para Hjørland (1992, p. 172, tradução nossa), “uma ligação muito estreita existe entre o que os assuntos são, e como estamos a conhecê-los.” De acordo com a abordagem até aqui adotada, é possível observar nitidamente as variações terminológicas para abordá-lo.

A relação entre a ideia de um assunto e os documentos foi sistematizada no trabalho de Cutter pelo catálogo dicionário e, segundo Miksa (2012, p. 15, tradução nossa), constitui-se como “[...] referente primário [...] [e] seu *status* formal em relação ao domínio classificatório natural do conhecimento humano. Em contraste, a ideia de um assunto como um atributo de um livro ou de qualquer documento tendia a ser de menor importância.” Portanto, a ideia de assunto, naquele momento, não foi uma proposição teórico-metodológica de elementos de manipulação do texto.

Levando em consideração a palavra assunto em inglês, *subject*, no dicionário de Biblioteconomia e Ciência da Informação, Reitz (2014, on-line, tradução nossa), define:

Qualquer um dos tópicos ou temas de uma obra, explicitamente declarado no texto ou título ou implícito em sua mensagem. [No processo de] catalogação em bibliotecas, é atribuído a um livro ou outro item um ou mais cabeçalhos de assunto como pontos de acesso, para ajudar os usuários a localizar seu conteúdo por assunto. Nos serviços de indexação e resumos, os cabeçalhos designados para representar o conteúdo de um documento são chamados de descritores. Abreviação subj. Veja também: *aboutness* e análise de assunto.

Em um sentido mais geral, qualquer tópico de estudo ou discussão, tema expresso por escrito, ou objeto ou cena retratado na pintura, desenho, fotografia, etc. Na Classificação Decimal de Dewey (CDD), os assuntos são organizados por disciplina. Porque um assunto pode ser estudado em mais de uma disciplina (exemplo: casamento em lei, psicologia, religião, sociologia, etc.), as escolhas feitas na classificação são governadas por regras.

Nesse extrato, tópico e tema são igualmente identificados como assunto. O autor resgata a tradição, segundo Cutter, de como os assuntos dos livros podem ser representados com uso de palavras sistematizadas em cabeçalhos. Apresenta também os descritores adotados nos serviços de indexação e resumos (em geral, sistemas especializados) e faz referência aos sistemas de classificação bibliográfica, como a CDD, referindo-se aos assuntos, agrupados por disciplinas, mas sem especificar o processo que esse instrumento contempla. Quanto aos termos relacionados, cita *aboutness* e análise de assunto. A definição é ampla e abarca o que já foi tratado neste trabalho. Portanto, aproxima as vertentes Catalogação de Assunto e Indexação. Se, “para Cutter, a estabilidade dos assuntos depende do processo social em que seu significado é estabilizado em um nome ou designação” (HJØRLAND, 2017, p. 56, tradução nossa), o cenário muda de configuração ao observar conceitos como *aboutness* e análise de assunto.

Hjørland (2001) explica não ser fácil entender o conceito de assunto (*subject*), pois seu contexto histórico imprime abordagem no processamento de livros em catálogos de bibliotecas. O autor pondera ao especificar que o assunto de um documento é o objetivo de um serviço de informação, definindo-o “[...] como os *potenciais* epistemológicos ou informativos dos documentos. [Afirma ainda que] A melhor análise de assunto é aquela que faz o melhor prognóstico do uso futuro do documento” (HJØRLAND, 2001, p. 776, itálico do autor, tradução nossa). O autor destaca a palavra potencial (*potentials*), que significa possibilidade, algo que pode vir a ser, portanto não real (AULETE, online; s.d.), pois deverá corresponder ao uso do documento.

Desse modo, Hjørland (2001) relaciona outros termos que por vezes são considerados sinônimos, como *aboutness*, tópico (*topic*), tema (*theme*) e conteúdo (*content*). Com exceção do primeiro, os demais corroboram as orientações que seguem. Ao abordar o termo “tópico”, o autor usa como explicação ele ser uma parte específica em

relação ao que um cientista trabalha em um campo mais amplo e, assim, remete à concepção do específico em relação a uma cobertura ampla. Para explicar o termo “tema”, faz aproximação com o que é individualizante e caracterizador. Cita-o como um elemento que direciona uma conferência ou um número especial de periódico e, portanto, fornece o aspecto central sobre o que está sendo abordado. Já em relação ao “conteúdo”, o autor faz aproximação com um texto, indicando que ele tem semelhança com um comentário que pormenoriza um determinado assunto. Sendo assim, aproxima à análise de conteúdo a determinação de assuntos dos documentos. Essas definições básicas podem caracterizar e relacionar os termos, porém são limitadas frente às demandas que poderiam ser colocadas para discussão da noção de assunto.

A conjugação do assunto como parte do registro de informação em catálogos e bases de dados deve reconhecer aspectos conceituais presentes no documento, os quais variam com a experiência de quem realiza a análise, o contexto, o público que fará uso, dentre outros fatores. Desse modo, a defesa dos assuntos serem potenciais epistemológicos ou informativos dos documentos (HJØRLAND, 2001) concorre para uma forma que seja procedimental e significativa, como, por exemplo, o assunto gerado a partir da análise.

Embora Hjørland (1992, p. 187, tradução nossa) explore diferentes linhas epistemológicas para estruturar o conceito do assunto como um elemento informativo, pensar a sua descrição envolve “[...] avaliação e atribuição de prioridades para as propriedades de um documento com relação a uma categorização e uma descrição do assunto desse documento.” Se, em termos gerais, essa descrição é precedida por alguma forma de análise, ela é “[...] contínua e decisiva para a ‘visibilidade’ do documento em bibliotecas e bases de dados, e assim para seu papel potencial no desenvolvimento futuro do conhecimento.”

Se o assunto abarca, de algum modo, adequação para indicar o que se quer representar e recuperar nos documentos, a análise

de assunto, também denominada, conforme Dias e Naves (2013), análise conceitual, análise temática, análise documental e análise de conteúdo, deve almejar implicações significativas para a constituição de uma representação contextualizada. Entretanto, é importante destacar que essa variação terminológica tem reflexo a partir de diferentes vertentes teóricas, mas também se relaciona à imprecisão que assumem as palavras conceito, tema, assunto, conteúdo e matéria. Além disso, há distinções de significado relacionado aos termos “análise de assunto” e “análise documental”.

Com base na citação acima e a partir do livro de Dias e Naves (2013, p. ix), encontra-se na “Apresentação” um alerta ao leitor para “os problemas de variação terminológica” que, segundo os autores, é percebida, inclusive, na literatura de língua estrangeira. Mas, na literatura nacional, o cenário não é tão diferente. A despeito de os autores fazerem o alerta, no decorrer do livro, percebe-se ausência de uma discussão sobre essas diferenças. Da mesma forma que, no inglês, é difícil estabelecer uma única definição para *subject* e *aboutness*, fato similar acontece no espanhol, em relação às palavras *contenido* e *materia*.

Na Catalogação, o termo *epígrafe*, em espanhol, é usado em linhas gerais para definir palavra ou palavras que são empregadas nos cabeçalhos das fichas de assunto dos catálogos dicionários, expressando o que a obra trata (ROVIRA, 1966; LITTON, 1971; BUONOCORE, 1976). Penna (1945) preferiu usar *encabezamientos de materia* em detrimento de *epígrafes de materia*, da mesma forma que Litton (1971) realizou essa escolha. Buonocore (1976, p. 189) destaca que os *encabezamientos de materia* ou *asuntos* recebem o nome de *epígrafe*, traduzido para o português brasileiro como “rubrica”.

O principal responsável em propagar a definição de *epígrafe*, segundo Rovira (1966), foi Jorge Aguayo, objetivando que a Biblioteca formasse seu próprio vocabulário em espanhol. Na visão de Rovira (1966), pareceu aceitável usar *epígrafe* em detrimento de

*encabezamiento de materia*, entretanto, a incorporação do termo evidenciou-se na literatura de Cuba, país que Aguayo desenvolveu longo trabalho na *Universidad de La Habana*.

Buonocore (1976, p. 198, tradução nossa), ao definir o termo *epígrafe*, demonstrou reflexão sobre a prática profissional iniciada no século XIX e explica que:

Embora os termos título e *epígrafe* sejam sinônimos em linguagem natural, a terminologia profissional os diferencia. O título é a palavra ou frase com a qual o autor denomina oficialmente sua obra e como decide estampar, como se sabe, na capa desta. A *epígrafe*, por outro lado, é o título atribuído pelo catalogador da obra e para isso ele escolhe a palavra mais específica e expressiva do conteúdo desta, usando uma lista previamente preparada – a de Lasso de la Vega, a Sears, por exemplo, – ou, caso contrário, mantendo um registro dos cabeçalhos usados no catálogo da biblioteca para padronizar critérios e estabelecer relações entre as diferentes *epígrafes*.

Nesse panorama e mesmo com o exemplo de Aguayo que buscou consolidar o termo citado em língua espanhola, o fato não alterou a influência do arcabouço anglo-saxão na Catalogação. Considerando a obra de Cutter antiquada em muitos aspectos, Rovira (1966, p. 7, tradução nossa) expressou que os autores de textos da área “[...] se limitavam a aconselhar a adoção de listas impressas, a ensinar seu uso e a dar algumas indicações práticas e princípios gerais deduzidos destas.”

Em Portugal, constata-se, a partir de Peixoto (1962, p. 26), referência à *epígrafe* como sinônimo de encabeçamento ou cabeçalho de assunto, compondo “[...] a palavra ou palavras que, postas na parte superior das fichas de matérias, nos dão o tema ou assunto concreto [...]”. Ainda, para o autor, “a expressão *terminológica* do assunto” (p. 26, destaque do autor) é a *epígrafe* e, “a expressão *numérica*” (p. 27, destaque do autor) é a classificação, disposta a partir da notação (também numérica) de um sistema classificatório, como, por exemplo,

a Classificação Decimal Universal (CDU). Desse modo, o autor trata das modalidades de *epígrafes*, conforme definições a seguir:

1. Epígrafes temáticas que designam a principal expressão terminológica do contexto da espécie bibliográfica a ser tratada catalogicamente. Ex.: Educação.
2. Epígrafes geográficas, que designam a expressão terminológica do espaço geográfico do contexto. Ex.: Portugal – Educação.
3. Epígrafes cronológicas ou históricas, que designam a expressão terminológica da época histórica ou cronológica. Ex.: Renascimento; Século XIII.
4. Epígrafes onomásticas ou biográficas, que designam uma pessoa individual ou colectiva tratada no seu aspecto biográfico ou histórico, etc., usando-se sempre o mesmo nome, quer se trate do catálogo de autores ou de assuntos. Ex.: Tomás de Aquino (Santo).
5. Epígrafes de título ou didascálicas, que se usam para designar como tema ou assunto uma obra apenas conhecida pelo título. Ex.: Diário de Notícias; Funchal e as lendas marítimas (PEIXOTO, 1962, p. 27).

Para Peixoto (1962), a disposição do catálogo alfabético de matérias (ou seja, de assuntos) é oriunda da tradição anglo-saxã do catálogo dicionário, que é a fusão do catálogo de autores, títulos e de matérias. Sobre o catálogo dicionário, Peixoto (1962, p. 26) explica que bibliotecários europeus mostraram-se “[...] relutantes em o aceitar. [...] nota-se uma certa tendência para o abandonar, [...] em virtude de ele crescer desmedidamente. [...] outros afirmam que o catálogo-dicionário, ao juntar o de autores e o de matérias, confunde duas fases da investigação, que são bem distintas.” De fato, a tradição europeia separou a elaboração do catálogo de assunto, por meio da construção do catálogo sistemático.

Litton (1971) explica que, pelo exame do livro, é possível indicar seu assunto (*materia*) e, assim, elegê-lo em uma lista de cabeçalho, o que compatibilizaria, segundo ele, a natureza da publicação. Contudo, também explicitou que, pelo cabeçalho, é possível descrever “[...] o conteúdo dos livros registrados no catálogo da biblioteca apenas em grau relativo. De todo modo, são mais uniformes e corretos para a interpretação do conteúdo das obras quando a instituição adota e adere estritamente a uma lista autorizada” (LITTON, 1971, p. 81, tradução nossa). Nessa condição, o assunto do documento seria preconcebido a partir da lista de cabeçalho mantendo vinculação com um plano explícito de palavras do documento. Essa relação, permeada por “[...] variações praticamente ilimitadas dos temas abordados nos livros, [e] a análise destes através dos cabeçalhos e das tabelas de classificação não é tão suscetível à formulação de um corpo de regras [...]” (LITTON, 1971, p. 77, tradução nossa).

No âmbito das fichas, a representação temática dos documentos comportou a identificação do que se denominou *contents* (conteúdo). Taylor (1948) abordou que essa compreensão envolveu aspectos de uma obra possuir vários volumes e, assim, a identificação requerer uma nota de conteúdo (*contents note*). Outra situação para o reconhecimento desse aspecto, segundo a autora, seria o de uma obra apresentar capítulos escritos por diferentes autores. Ainda segundo Taylor (1948, p. 91, tradução nossa, destaque da autora), na ficha catalográfica:

A nota de conteúdo é iniciada meio centímetro abaixo da última linha da entrada. A palavra *conteúdo* começa na margem interna, é indicada em itálico [...] e seguida por dois pontos. Depois vêm os detalhes do conteúdo, sendo estes levados de volta para a margem externa e cada item separado por um longo traço. Um centímetro para cima a partir da parte inferior da ficha deve ser deixado em branco, exceto para as palavras (continua na próxima ficha), pois a parte inferior é difícil de ser vista quando se está em pé em uma gaveta do catálogo. Ao continuar numa segunda ficha, numere esta (2), repita o cabeçalho e o título brevemente, seguidos por um espaço de dois centímetros e *Conteúdo (continuação)*. Deixe meio centímetro e comece a listar o restante do conteúdo sem qualquer recuo de parágrafo na margem interna.

A identificação do *contents* (conteúdo) na ficha catalográfica alcançou relativa estabilidade através da nota de conteúdo, e não propriamente no tange à representação temática. Por sua vez, Buonocore (1976, p. 143, tradução nossa), com base em Jorge Aguayo, afirma que a descrição do *contenido* “[...] é o índice colocado nos livros seguindo a mesma ordem dos capítulos.” Neste caso, ele está se referindo ao recurso que chamamos mais frequentemente de sumário. O conteúdo, através da nota concebida logicamente por palavras extraídas do documento, apresentou uma proposta de padronização envolvendo a pontuação e o lugar de apresentação na ficha, porém não houve embasamento metodológico para essa construção.

Ainda, no que tange às fichas, a entrada secundária, explica Litton (1971), rege o desdobramento, ou seja, a elaboração de fichas secundárias, a partir da ficha principal ou única (*tarjeta única*). Nessa tradição, ao identificar o assunto (*materia*), quem catalogou deve encontrá-lo na lista de cabeçalho e, assim, indicá-lo no lugar da entrada secundária na ficha principal. Após essa escolha e seguindo o padrão da ficha secundária de assunto,

O cabeçalho selecionado é transcrito no espaço acima da linha do autor, a partir da segunda margem. Para destacá-lo, uma dessas duas regras é geralmente seguida: a) as letras são todas escritas em maiúsculas e em preto; e b) todas as letras são transcritas em maiúsculas e em vermelho.

Quando um cabeçalho não se encaixa na linha a ele destinada, terá que elevá-lo mais para a borda da ficha, de modo que a segunda parte possa ocupar a segunda linha a partir da terceira margem.

[...]

Quando um cabeçalho é composto por várias partes, elas podem ser separadas umas das outras por meio de: a) um ponto e dois espaços; e b) uma linha com dois traços (nunca um) (LITTON, 1971, p. 138, tradução nossa).

De modo análogo, conforme Penna (1945, p. 85, tradução nossa) expressa, “[...] as fichas de assunto são as que representam os livros de acordo com seu conteúdo.” Para o autor, essa aproximação parte das *fichas de materia* (fichas de assunto) de um catálogo dicionário se relacionarem com o modo de ser do catálogo sistemático através dos seus símbolos classificadores (*símbolos clasificadores*). Ainda, segundo o autor, antes de decidir esse símbolo e o cabeçalho, quem cataloga deve buscar o *tema* (tema) tratado na obra através da leitura do seu prefácio, sumário, dentre outros elementos e, se necessário, recorrer à consulta de terceiros, ou a fontes adequadas como bibliografias e outros catálogos. Após determinar o tema da obra, o autor explica que as tarefas de indicação do símbolo classificador e do cabeçalho não são idênticas, pois no primeiro caso “[...] será necessário ter um raciocínio que nos leve exatamente ao lugar do sistema que, pela disposição lógica das ciências, corresponda ao livro que se busque localizar [...]”. Já para o segundo, “[...] devemos procurar a palavra que expresse de forma mais determinante o conteúdo do livro [...]” e dessa forma ambos os processos “[...] se diferem de forma tão fundamental.” (PENNA, 1945, p. 85, tradução nossa).

A disposição das fichas nesse tipo de catálogo segue a ordem alfabética agrupando autores, títulos e assuntos, mas os cabeçalhos de assunto (*encabezamientos de materia*) assumem relevância “[...] pois são os que agrupam em determinadas partes do catálogo todas as fichas de obras que tratam do mesmo tema” (PENNA, 1945, p. 85, tradução nossa). Ainda, segundo esse autor, o conteúdo seria a substância do texto formando as partes do livro e, através do seu exame, se determinaria o tema. Portanto, a partir do tema, o assunto seria indicado pelo cabeçalho e, representado na ficha, congregaria o primeiro vinculando-o com a descrição controlada do instrumento. Com esse aspecto, depreende-se que o assunto seria o resultado do tema, pois o primeiro é esquematizado no cabeçalho e o segundo resultaria da análise.

A despeito da relação apontada por Penna (1945), a atribuição de um símbolo classificador ao documento forneceria uma localização no acervo, de acordo com o mapeamento das ciências dado no sistema bibliográfico e, com o cabeçalho, o uso de uma palavra específica resultaria no assunto. Em ambos os casos, a representação continuou privilegiando aspectos genéricos dos instrumentos e se baseou no documento como um todo.

Sharp (1964), ao abordar as tarefas de classificação e catalogação de assunto, explica que a entrada secundária de assunto nas fichas fornece um caminho que amplia o processo de análise por conta das referências cruzadas que podem ser construídas. Critica apontando que houve frequente negligenciamento na relação dessas construções e que “muito já foi dito sobre os méritos de disposição dos assuntos nas respectivas formas dos catálogos dicionário e classificado [...]” (SHARP, 1964, p. 334, tradução nossa). Mas o autor, sob o viés anglo-saxão, também se restringiu metodologicamente em torno da prática tão difundida e igualmente já apontada nos livros de Catalogação (PENNA, 1945; TAYLOR, 1948; MANN, 1962; LITTON, 1971). A sistematização para o assunto fornecida por Cutter, na forma de um cabeçalho, foi influente na construção da vertente da Indexação, mas tomou outra direção a partir da ideia de índice e das metodologias de indexação.

Nessa diferença entre vertentes, o processo de indexação é comum, mas a ideia de uma entrada secundária, não é. Nos catálogos em fichas ou on-line, há entrada secundária de assunto, mas, nas bases de dados científicas, elas não existem como tal, pois se trabalha com pontos de acesso de assunto. Com isso, o ponto de acesso tido como principal empregado por um nome, termo ou signo são utilizados para encontrar uma entidade ou um conjunto de entidades em uma ficha catalográfica e seus desdobramentos são possibilitados pelas entradas secundárias. Porém, com as bases de dados, vários pontos podem ser disponibilizados para a recuperação e a palavra principal perdeu a função original.

A exposição efetuada permite verificar que o desdobramento da vertente anglo-saxã é constituído por fragmentos, mas com aspectos comuns, como o da entrada por assunto. Ademais, o assunto é um atributo relevante, mas prevalece a dispersão e, por consequência, a variação terminológica em torno da sua compreensão. Do ponto de vista da organização, instrumentos condicionam o assunto, como a lista de cabeçalhos, que centram na morfologia e na sintaxe, enquanto o tesouro na semântica. Do ponto de vista da representação, o assunto é construído e não extraído do documento.

#### 4 BUSCA DOS OLHARES SOBRE A REPRESENTAÇÃO DE ASSUNTO NA SÉRIE ESTUDOS AVANÇADOS EM ORGANIZAÇÃO DO CONHECIMENTO (ISKO BRASIL)

Com base na exposição teórica levantada, é possível observar a variação terminológica que envolve o termo “assunto”. Essa tentativa de trabalhar o escopo científico sobre o assunto e, por consequência, sua representação, pauta-se em elementos aparentemente similares ou complementares, mas pertencentes a diferentes perspectivas que convergem e divergem ideias sobre o que é o assunto. Por consequência, há relação direta com o resultado da manipulação do assunto, que se dá pela representação.

A motivação em identificar a presença do termo “representação de assunto” e se, junto a ele, há variação terminológica, observando o levantamento feito – descritor; *aboutness*, *topic*, *content*, *subject* – assunto, tema, matéria – *contenido*, *materia epígrafe*, *tema* – tem como intuito compreender qual(is) termo(s) marca(m) presença na literatura dos eventos do capítulo brasileiro da ISKO. Essa tentativa de olhar para os termos que coexistem na literatura das cinco edições

dos eventos nacionais é um movimento que auxilia a (re)conhecer o que é trabalhado quando se trata da representação de assunto.

Para identificar os trabalhos, foram realizadas buscas de forma automática, usando o recurso localizar, nos cinco arquivos em PDF, que compõem os volumes da Série *Estudos Avançados em Organização do Conhecimento*<sup>5</sup>. A expressão empregada foi “representação de assunto”, incluindo o plural “assuntos”, de modo a verificar a ocorrência. É oportuno explicitar que trabalhos com a referida expressão somente nas considerações finais ou com citação para exemplificar algo uma única vez não foram selecionados para compor o *corpus* da análise. Seguindo esse critério há dois trabalhos no volume 5 que não foram selecionados.

A partir da busca realizada, foram recuperados, nos volumes 1, 2 e 5, um trabalho em cada e, nos volumes 3 e 4, três em cada, totalizando nove trabalhos, conforme sistematização no quadro 1.

**Quadro 1 – Identificação do *corpus* para análise**

Ano do evento/ volume/fonte	Autor(es)(as)	Título do trabalho	Uso da expressão representação de assunto	Presença dos termos: descriptor; <i>about- ness, topic, content, subject</i> – assunto, tema, matéria – <i>contenido, materia epígrafe, tema</i>
2012, v. 1  (GUIMARÃES; DODEBEI, 2012)	Carlos H. Marcondes	Organização e representação do conhecimento em ambientes digitais: as relações entre ontologia e Organização do Conhecimento	Representação de assuntos compos- tos (Ranganathan)	assunto

<sup>5</sup> Todos os volumes estão disponíveis no site da ISKO Brasil (<https://isko.org.br/>), na aba publicações. Observar que as entradas das referências dos volumes variam de acordo com os(as) organizadores(as).

2013, v. 2  (DODEBEI; GUI- MARÃES, 2013)	Fabio Assis Pinho; Suellen de Oliveira Milani	Metáfora e ortofe- mismo na represen- tação de assunto	Representação de assunto - substi- tutos documentais (por exemplo: índices, resumos e notações de classificação); poder de nomear (Olson, 2002)	assunto; conteúdo tópico dos documentos
2015, v. 3  (GUIMARÃES; DODEBEI, 2015)	Graziela Mar- tins de Medei- ros; Leolíbia Luana Linden; Luciane Paula Vital; Marisa Bräscher	A representa- ção de assunto no escopo da arquivologia: uma análise de artigos científicos nacionais e internacionais	Representação de assunto; indexação de assuntos (Norma Brasileira de Des- crição Arquivística –NOBRADE)	assunto; organização e repre- sentação da infor- mação arquivística
	Gustavo Silva Saldanha; Naira Christo- foletti Silveira	Poole, o índice e as fraturas: indexação e publicações seriadas nos Estados Unidos do oitocentos	Indexação de as- suntos (Lancaster); Representação de assuntos (plural); Representação temática	assunto; termos extraídos de um documen- to – descritores;
	Vera Lucia Ribeiro Guim; Mariângela Spotti Lopes Fujita	As linguagens de indexação e a análise de domínio	Representação de assuntos (Classi- ficação Decimal Universal – CDU)	conteúdo; assuntos (plural)

2017, v. 4  (PINHO; GUIMARÃES, 2017)	Maria Carolina Andrade e Cruz; Mariângela Spotti Lopes Fujita; Luciana Beatriz Piovezan dos Santos	Linguagem de indexação no contexto da política de indexação: estudo em bibliotecas universitárias	Representação de assuntos; Tratamento Temático da Informação; Representar o conteúdo dos documentos; Indexação de assuntos	assunto; conteúdos atribuídos aos documentos da biblioteca; conteúdo temático do documento;
	Helen de Castro Silva Casarin; Thaís Guedes Ferreira; Suellen de Oliveira Milani	As crianças como comunidades de usuários e os Sistemas de Organização do Conhecimento: uma análise da literatura	Representação de assunto	assunto
	Suellen de Oliveira Milani; José Augusto Chaves Guimarães	Problemas relacionados a biases em Sistemas de Organização do Conhecimento: perspectivas para a representação de assunto	Representação de assunto; Atribuição de rótulos (Olson, 2002)	assunto; conteúdo
2019, v. 5  (BARROS; TOGNOLI, 2019)	Marcos Luiz Cavalcanti de Miranda; Fábio Gomes da Silva	O Islamismo na CDD e CDU: religião e cultura periféricas nos esquemas de classificação bibliográfica	Representação de assunto; Classificação de assunto	assunto; assuntos

Fonte: Elaborado pela autora, 2021.

Cabe destacar o quantitativo de trabalhos em cada volume: no 1, há 39 trabalhos, com participação de 72 autores(as) (GUIMARÃES; DODEBEI, 2012, p. 18); no 2, 46 trabalhos e participação de 85 autores(as) (DODEBEI; GUIMARÃES, 2013, p. 14); no 3, 75 trabalhos (GUIMARÃES; DODEBEI, 2015, p. 16), mas sem indicação do número de autores(as). Já a contabilização de trabalhos do volume 4 (com 46) e do volume 5 (com 56) não constam nos prefácios e foram levantados manualmente pela autora.

A partir da leitura integral dos nove artigos recuperados, foi possível identificar, mesmo que de forma breve, como as autoras e os autores empregaram a representação de assunto (penúltima coluna do quadro 1) nas suas abordagens. Também, por esse movimento, foi possível identificar se há presença de variação terminológica, conforme apontado na última coluna do referido quadro. Nesse panorama, a análise do *corpus* levantado segue na próxima subseção.

#### 4.1 ANÁLISE E POSSÍVEIS DISCUSSÕES

Após a leitura e análise dos nove artigos, constata-se que há presença de variação dos termos para referir-se a assunto e, por consequência, sua representação, conforme os apontamentos a seguir:

O artigo recuperado no volume 1 (MARCONDES, 2012) propõe um esboço das relações entre Organização do Conhecimento (OC), a partir da Teoria da Classificação Facetada de Ranganathan, e Ontologia, buscando o diálogo interdisciplinar. O autor indica trabalhar os fundamentos de ambas as disciplinas e identificar o papel da OC na gestão do conhecimento em ambientes digitais. O autor busca relacionar a representação de assunto no contexto web. Não se constatou presença de variação terminológica. O trabalho foi vinculado ao eixo denominado “A dimensão epistemológica da organização do conhecimento”.

No volume 2, o artigo recuperado (PINHO; MILANI, 2013) tem como objeto de discussão a ausência de precisão terminológica no processo de indexação, situação que pode levar a vieses na representação, envolvendo outros diferentes problemas, como os relativos ao preconceito, proselitismo, questões de gênero, categorizações dicotômicas, dentre outros. Pinho e Milani (2013) abordam que o processo de indexação é permeado por subjetividade e o(a) profissional, ao organizar a informação, a rotula. Ressaltam que essa ação não pode

ser dissociada do contexto cultural e ideológico de quem indexa e do sistema de informação. A representação de assunto, para o autor e a autora, consiste em atribuir nomes ou rótulos ao conteúdo tópico dos documentos. Consta-se que há presença de outro termo para tratar do assunto, que comunga com a proposta do estudo. Ele foi vinculado ao eixo denominado “A dimensão social, cultural e política da Organização e Representação do Conhecimento”.

Já no volume 3, Medeiros, Linden, Vital e Bräscher (2015) abordam a representação de assunto no escopo da Arquivologia, indicando haver escassez de aprofundamento teórico nesse escopo. Verificaram como a representação de assunto em documentos de arquivo é tratada nas publicações da área de Ciência da Informação. É possível notar que a expressão organização e representação da informação arquivística é empregada de forma sinônima a representação de assunto. Esse trabalho diferencia-se por focar a visão da Arquivologia e foi vinculado ao eixo denominado “A dimensão aplicada da Organização e Representação do Conhecimento”.

O segundo artigo recuperado no volume 3, de Saldanha e Silveira (2015), trabalha elementos epistemológicos para a discussão de conceitos, de métodos, de abordagens e de experimentos dos estudos informacionais. Objetiva discutir, a partir de um enfoque teórico, o desenvolvimento de ferramentas de recuperação da informação no âmbito da produção seriada no século XIX, considerando a figura de William Frederick Poole e o seu Índice Poole. De acordo com Saldanha e Silveira (2015), o trabalho de Poole marcou a consolidação de um pensamento americano em “recuperação temática de artefatos”, principalmente no âmbito dos periódicos. Por terem como foco um instrumento específico da indexação (o índice), o autor e a autora empregam outras expressões que tem relação com o processo em jogo. O trabalho foi vinculado ao eixo “A dimensão epistemológica da Organização e Representação do Conhecimento”.

No último artigo recuperado do volume 3, Guim e Fujita (2015) abordam diferentes linguagens de indexação, pré e pós-coordenadas, sendo: Classificação Decimal de Dewey (CDD), Classificação Decimal Universal (CDU), tesouro e cabeçalhos de assunto. Enfocam serem linguagens construídas artificialmente que visam, de maneira sucinta, representar o conteúdo do documento. Aliam a importância da análise de domínio como embasamento teórico para atualização e construção de linguagens de indexação. Enfocam a expressão representação de assuntos ao instrumento CDU. O trabalho foi vinculado ao eixo “A dimensão epistemológica da Organização e Representação do Conhecimento”.

Tratando do volume 4, o artigo de Cruz, Fujita e Santos (2017) aborda as linguagens de indexação e sua função no processo de tratamento temático dos documentos. Enfocam como sua aplicação repercute no resultado da indexação e a importância da política de indexação nesse contexto, pois ela garante que os critérios preestabelecidos para o processo de representação sejam seguidos por quem realiza o processo. Investigam os procedimentos adotados quanto ao uso das linguagens de indexação, no contexto das bibliotecas universitárias, olhando para a inconsistência no controle de vocabulário, que causa imprecisão na representação temática nos catálogos das bibliotecas e, conseqüentemente, disparidades na recuperação da informação. O trabalho apresenta variedade de expressões com termos que compõem a perspectiva da indexação. Foi vinculado ao eixo “Dimensão Aplicada da Organização do Conhecimento”.

Continuando no volume 4, o artigo de Casarin, Ferreira e Milani (2017) relaciona os estudos sobre comportamento informacional e como, por meio deste, é possível viabilizar a identificação de características e necessidades de grupos específicos de indivíduos a fim de trazer subsídios aos processos de organização do conhecimento. O estudo propõe-se a analisar a produção científica da área de organização do conhecimento, no que diz respeito às crianças e, para isso, realizou-se

um levantamento bibliográfico em periódicos internacionais especializados. As autoras relacionam a representação de assunto à organização do conhecimento e, por consequência, aos sistemas que dão base a essa representação. Nesse contexto, incluem os sistemas de classificação, listas de cabeçalhos de assunto, listas de autoridades, tesouros, ontologias, taxonomias, terminologias e outros instrumentos que servem para a representação de assunto de um documento por meio de uma estruturação semântica (HODGE, 2000). O trabalho foi vinculado ao eixo “Dimensão Política e Social da Organização do Conhecimento”.

O último artigo recuperado do volume 4, de Milani e Guimarães (2017), aborda dilemas éticos no processo de representação de assunto. Sinalizam que o uso de instrumentos, tais como esquemas de classificação e linguagens de indexação, juntamente à representação de assunto não realizada de uma forma defensável, pode promover, censurar, omitir e distorcer informações. Nesse contexto, a autora e o autor direcionam que o(a) bibliotecário(a) precisa considerar valores éticos e barreiras culturais e linguísticas, pois estará diante de dilemas que exigirão tomadas de decisão éticas. A partir do momento em que o(a) bibliotecário(a) reconhece o fato de que dilemas éticos não resolvidos podem causar danos às comunidades de usuários(as), sua postura perante tal circunstância deverá ser respeitosa. Considerando relatos de artigos extraídos da literatura internacional especializada em organização do conhecimento, o trabalho propõe uma reflexão sobre a forma pela qual os sistemas de organização do conhecimento podem fomentar dilemas éticos na atuação do(a) bibliotecário(a). Milani e Guimarães (2017) apontam que a representação de assunto insere-se no âmbito da organização do conhecimento, que é um marco teórico conceitual para a Biblioteconomia e Ciência da Informação. O trabalho foi vinculado ao eixo “Dimensão Política e Social da Organização do Conhecimento”.

O último trabalho que compõe o *corpus* selecionado e está no volume 5, de Miranda e Silva (2019), investiga a existência de desvios

na representação de assuntos referentes às culturas não alinhadas à cultura ocidental, no que tange o Islamismo. Emprega a 23ª edição da Classificação Decimal de Dewey e a 2ª Edição Padrão Internacional em Língua Portuguesa da Classificação Decimal Universal. Aborda a construção do conceito de Islamismo no Ocidente, considerando as transformações de sentido moduladas pelo contexto sociocultural, conferindo, ao Islã, a imagem da violência e realiza estudo comparativo junto aos dois sistemas de classificação mencionados. Os resultados do estudo dos esquemas de classificação demonstraram a existência de desvios (biases) de representação na Classificação Decimal de Dewey, evidenciados tanto na seleção terminológica quanto em sua estrutura conceitual; e as correções de biases de representação pela abordagem classificatória (analítico sintética e facetada) da Classificação Decimal Universal. É possível indicar que a representação de assunto está alinhada à classificação. O trabalho foi vinculado ao eixo “Dimensão Política e Social da Organização do Conhecimento”.

Com base no *corpus* selecionado, observa-se que o elemento recorrente à representação de assunto é o processo de indexação, que envolve o uso de diferentes instrumentos – inclusive os de classificação – e geram produtos variados. Sendo assim, a recorrência vai ao encontro do cenário teórico levantado.

Também, conforme apontado no referencial teórico, é preciso considerar a existência das diversas abordagens, construídas em momentos históricos e culturais distintos, e que estão presentes na literatura. Assim, é importante compreender como esses elementos convergem, considerando a fundamentação que os envolvem, mas sem esquecer das particularidades teórica que os individualizam.

Confirma-se com a análise que no mínimo dois amplos caminhos podem ser apontados em relação ao aspecto da representação de assunto: um que visa refletir/coletar/estruturar termos para construir um vocabulário controlado, portanto uma linguagem; e outro

que envolve a sistematização de aspectos metodológicos da própria indexação, que se dá pelos assuntos. O processo não é isento de neutralidade e fundamentos denominados éticos estão presentes.

Nenhum dos trabalhos recuperados tratam de elementos conceituais do assunto e/ou da sua representação. Também não empregam termos distantes dos que foram elencados para abarcar assunto, mas foi recorrente o uso do termo conteúdo. Sobre a expressão “representação de assunto”, foi interessante observar os termos correlacionados. Ademais, com exceção do trabalho de Medeiros, Linden, Vital e Bräscher (2015), todos estão vinculados ao contexto da Biblioteconomia.

## 5 PARA FINALIZAR: BREVES CONSIDERAÇÕES

(Re)conhecer o que é trabalhado sobre a representação de assunto, através das publicações da comunidade do capítulo brasileiro da ISKO, é um grande desafio. Se, por um lado, a sistematização desse esforço pode prover contribuição para a consolidação do que é produzido em um determinado momento, por outro, a análise inicial requer que outras sejam realizadas.

Nas pesquisas iniciais para esta investigação, percebeu-se que a palavra assunto deflagra diferentes nuances da representação, que é temática, fazendo perceber outros termos também adotados para nomear este tipo de representação. Deve-se considerar a existência de abordagens diversas, construídas em momentos históricos e culturais distintos, e que estão presentes na literatura da área. De acordo com a exploração da literatura realizada, vertentes constituíram-se com suas respectivas abordagens, mas elas influenciam-se umas as outras, pois, apesar dos fundamentos que as distinguem, elas se aproximam por um propósito: a representação de assunto.

Analisar a variação terminológica que envolve o assunto e, por consequência, sua representação e relacionar o referencial trabalhado, sinaliza diferentes tempos e lugares. Resgatar esse cenário colabora para o fortalecimento da Organização do Conhecimento ao demonstrar que a representação de assunto envolve a busca por teorias e metodologias rigorosamente fundamentadas, de base intelectual. Além disso, há forte relação com a Biblioteconomia.

No que tange as denominações dos eixos que englobam os trabalhos da Série *Estudos Avançados em Organização do Conhecimento*, cabe destacar alteração de nomenclaturas, especificamente na expressão Organização e Representação do Conhecimento para Organização do Conhecimento. Também é importante salientar que houve alteração dos formatos dos trabalhos que compõem os volumes da referida série: de resumos expandidos passaram para texto completo.

É desejo que a apropriação realizada nesse trabalho possibilite outras discussões, pois, além de ser tema do projeto de pesquisa<sup>6</sup> coordenado pela autora, abrem-se portas para o aprofundamento de críticas às questões em torno do assunto e da sua representação.

## 6 REFERÊNCIAS

AULETE Digital. Disponível em: <http://www.aulete.com.br/assunto>. Acesso em: 22 nov. 2021.

BARROS, Thiago Henrique Bragato; TOGNOLI, Natalia Bolfarini (orgs). **Organização do conhecimento responsável**: promovendo sociedades democráticas e inclusivas. Belém: ED. da UFPA, 2019. Série: Estudos Avançados em Organização do Conhecimento, v. 5. Disponível em: <https://isko.org.br/>

6 Título do projeto: A representação de assunto entre o contexto analógico ao digital: perspectivas teórico-prática dos processos, produtos e instrumentos. Cadastrado no 'Portal da Pesquisa' da Diretoria de Pesquisa, vinculada a Pró-Reitoria de Pós-Graduação, Pesquisa e Inovação (PROPGPI) da UNIRIO: [http://sistemas.unirio.br/projetos/projeto/index?ID\\_PROJETO=9978](http://sistemas.unirio.br/projetos/projeto/index?ID_PROJETO=9978).

publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/. Acesso em: 20 out. 2021.

BEGHTOL, Claire. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. **Journal of Documentation**, London, v. 42, n. 2, p. 84-113, june. 1986.

BUONOCORE, Domingo. **Diccionario de Bibliotecologia**. 2. ed. aum. Buenos Aires: Marymar, 1976.

CASARIN, Helen de Castro Silva; FERREIRA, Thaís Guedes; MILANI, Suellen de Oliveira. As crianças como comunidades de usuários e os Sistemas de Organização do Conhecimento: uma análise da literatura. *In*: PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves (orgs). **Memória, tecnologia e cultura na organização do conhecimento**. Recife: Ed. UFPE, 2017. Série: Estudos Avançados em Organização do Conhecimento, v. 4. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

CRUZ, Maria Carolina Andrade e; FUJITA, Mariângela Spotti Lopes; SANTOS, Luciana Beatriz Piovezan dos. Linguagem de indexação no contexto da política de indexação: estudo em bibliotecas universitárias. *In*: PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves (orgs). **Memória, tecnologia e cultura na organização do conhecimento**. Recife: Ed. UFPE, 2017. Série: Estudos Avançados em Organização do Conhecimento, v. 4. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

CUNHA, Celso. Gramática básica do português contemporâneo. *In*: AULETE Digital. **Gramática básica**. Disponível em: <http://www.aulete.com.br/site.php?mdl=gramatica>. Acesso em: 15 out. 2021.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto**: teoria e prática. Brasília: Briquet de Lemos, 2013.

DICIONÁRIO da língua portuguesa: dicionários editora. Porto: Porto Editora, 2004.

DOBEDEI, Vera; GUIMARÃES, José Augusto Chaves (orgs). **Complexidade e organização do conhecimento, desafios de nosso século**. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. Série: Estudos Avançados em Organização do Conhecimento, v. 2. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

FAIRTHORNE, Robert A. Content analysis, specification and control. **Annual Review of Information Science and Technology**, v. 4, p. 73-109. 1969.

FARIA, Maria Isabel; PERICÃO, Maria da Graça. **Dicionário do livro**: da escrita ao livro electrónico. Coimbra: Almedina, 2008.

FERREIRA, António Gomes. **Dicionário de latim-português**. Porto: Porto Codex, 1987.

FOSKETT, Antony Charles. **A abordagem temática da informação**. São Paulo; Brasília: Editora da Universidade de Brasília; Editora Polígono, 1973.

FUJITA, Mariângela Spotti Lopes. Sistema de indexação precis. I – PRECIS: perspectiva histórica e técnica do seu desenvolvimento e aplicação (contínua). **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 21-45. 1988.

GUEDES, Emanuel Guedson Ferrerira. **O conceito de aboutness na organização e representação do conhecimento**. 2009. 90 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

GUIM, Vera Lucia Ribeiro; FUJITA, Mariângela Spotti Lopes. As linguagens de indexação e a análise de domínio. *In*: GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil: FUNDEPE, 2015. Série: Estudos Avançados em Organização do Conhecimento, v. 3. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

GUIMARÃES, José Augusto Chaves. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 1, n. 1, p. 77-99, jan./jun. 2008.

GUIMARÃES, José Augusto Chaves. Abordagens teóricas de tratamento temático da informação: catalogação de assunto, indexação e análise documental. **Ibersid**, Zaragoza, v. 3, p. 105-117. 2009.

GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade**. Marília: ISKO-Brasil: FUNDEPE, 2012. Série: Estudos Avançados em Organização do Conhecimento, v. 1. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil: FUNDEPE, 2015.

Série: Estudos Avançados em Organização do Conhecimento, v. 3. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

HENDERSON, Madeline M. Calvin N. Mooers, October 24, 1919 - December 1, 1994. **Journal of the American Society for Information Science**, Maryland, n. 47, n. 9, p. 659-661. 1996.

HJØRLAND, Birger. Subject (of Documents). **Knowledge Organization**, Würzburg, v. 44, n. 1, p. 55-64. 2017.

HJØRLAND, Birger. The concept of "subject" in information science. **Journal of Documentation**, London, v. 48, n. 2, p. 172-200. 1992.

HJØRLAND, Birger. Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. **Journal of the American Society for Information Science and Technology**, Maryland, v. 52, n. 9, p. 774-778, jul. 2001.

HODGE, Gail. **Systems of Knowledge Organization for digital libraries: beyond traditional authority files**. Washington, DC: The Digital Library Federation, 2000. Disponível em <<http://old.diglib.org/pubs/dlf090/dlf090.pdf>>. Acesso em: 20 out. 2021.

HUTCHINS, John W. On the problem of "aboutness" in document analysis. **Journal of Informatics**, v. 1, n. 1, p. 17-35, April. 1977.

HUTCHINS, John W. The concept of 'aboutness' in subject indexing. **Aslib Proceedings**, v. 30, n. 5, p. 172-181. 1978.

LANCASTER, Frederick Wilfrid. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

LARA, Marilda Lopes Ginez de. **Dos sistemas de classificação bibliográfica às search engines (II)**. São Paulo: Associação Paulista de Bibliotecários, 2001 (Ensaio APB, n. 91).

LITTON, Gaston. **Classificação e catálogos**. São Paulo: Mcgraw-hill, 1971.

MANN, Margaret. **Catálogo e classificação de livros**. Rio de Janeiro: Editora Fundo de Cultura, 1962.

MARCONDES, Carlos H. Organização e representação do conhecimento em ambientes digitais: as relações entre ontologia e Organização do Conhecimento. In: GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade**. Marília: ISKO-Brasil: FUNDEPE, 2012. Série: Estudos Avançados em Organização do Conhecimento, v. 1. Disponível em:

<https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

MEDEIROS, Graziela Martins de; LINDEN, Leolíbia Luana; VITAL, Luciane Paula; BRÄSCHER, Marisa. A representação de assunto no escopo da arquivologia: uma análise de artigos científicos nacionais e internacionais. *In*: GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil: FUNDEPE, 2015. Série: Estudos Avançados em Organização do Conhecimento, v. 3. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

METCALFE, John. When is a Subject not a Subject? *In*: RAWSKI, Conrad H. (ed.). **Towards a theory of Librarianship**. New York: Scarecrow Press, 1973. p. 303-338.

MIKSA, Francis Louis. The legacy of the library catalogue for the present. **Library Trends**, Champaign, v. 61, n. 1, p. 7-34. 2012.

MILANI, Suellen de Oliveira; GUIMARÃES, José Augusto Chaves. Problemas relacionados a biases em Sistemas de Organização do Conhecimento: perspectivas para a representação de assunto. *In*: PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves (orgs). **Memória, tecnologia e cultura na organização do conhecimento**. Recife: Ed. UFPE, 2017. Série: Estudos Avançados em Organização do Conhecimento, v. 4. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

MIRANDA, Marcos Luiz Cavalcanti de; SILVA, Fábio Gomes da. O Islamismo na CDD e CDU: religião e cultura periféricas nos esquemas de classificação bibliográfica. *In*: BARROS, Thiago Henrique Bragato; TOGNOLI, Natalia Bolfarini (orgs). **Organização do conhecimento responsável: promovendo sociedades democráticas e inclusivas**. Belém: ED. da UFPA, 2019. Série: Estudos Avançados em Organização do Conhecimento, v. 5. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

MOOERS, Calvin N. The indexing language of an information retrieval system. *In*: CHAN, Lois Mai; RICHMOND, Phyllis A.; SVENONIUS, Elaine (eds.). **Theory of subjects analysis: a sourcebook**. Colorado: Libraries Unlimited, 1985. p. 247-261.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, p. 20-32, jan. 1951. Disponível em:

<https://courses.engr.illinois.edu/cs473/fa2013/misc/zatocoding.pdf>.  
Acesso em: 22 nov. 2021.

MORAES, João Batista; GUIMARÃES, José Augusto Chaves. Análise documental de conteúdo de textos literários: em busca do diálogo entre as concepções de *aboutness/meaning* e de percurso temático/percurso figurativo. **Scire**, Zaragoza, v. 12, n. 1, en./jun. 2006.

PEIXOTO, Jorge. **Técnica bibliográfica**: subsídio para a bibliografia portuguesa. II parte – catalogação. Coimbra: Atlântida, 1962.

PENNA, Carlos Victor. **Catalogacion y clasificacion de libros**. Buenos Aires: ACME Agency, 1945.

PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves (orgs). **Memória, tecnologia e cultura na organização do conhecimento**. Recife: Ed. UFPE, 2017. Série: Estudos Avançados em Organização do Conhecimento, v. 4. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

PINHO, Fabio Assis; MILANI, Suellen de Oliveira. Metáfora e ortofemismo na representação de assunto. *In*: DOBEDEI, Vera; GUIMARÃES, José Augusto Chaves (orgs). **Complexidade e organização do conhecimento, desafios de nosso século**. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. Série: Estudos Avançados em Organização do Conhecimento, v. 2. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

REITZ, Joan M. **Online Dictionary for Library and Information Science**. Califórnia: ABC-CLIO Corporate, 2014. Disponível em: [https://www.abc-clio.com/ODLIS/odlis\\_s.aspx](https://www.abc-clio.com/ODLIS/odlis_s.aspx). Acesso em: 22 nov. 2021.

ROVIRA, Carmen. **Los epigrafs en el catalogo diccionario**. 2. ed. rev. Washington: Union Panamericana, 1966.

SALDANHA, Gustavo Silva; SILVEIRA, Naira Christofolletti. Poole, o índice e as fraturas: indexação e publicações seriadas nos Estados Unidos do oitocentos. *In*: GUIMARÃES, José Augusto Chaves; DODEBEI, Vera (orgs). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil: FUNDEPE, 2015. Série: Estudos Avançados em Organização do Conhecimento, v. 3. Disponível em: <https://isko.org.br/publicacoes/serie-estudos-avancados-em-organizacao-e-representacao-do-conhecimento/>. Acesso em: 20 out. 2021.

SHARP, Henry Alexander. **Cataloging**: a textbook for use in libraries. London: Allied Publishers Private, 1964. Disponível em: <https://archive.org/>

stream/in.ernet.dli.2015.533483/2015.533483.cataloguing-a#page/n5/mode/2up. Acesso em: 22 nov. 2021.

SIMÕES, Maria da Graça de Melo. Do conceito e dos princípios metodológicos de indexação. Subsídios para uma análise crítica. *In*: JORENTE, Maria José Vicentini; PADRÓN, Dunia Ilanes (orgs.). **Una mirada a la ciencia de la información desde los nuevos contextos paradigmáticos de la posmodernidad**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2017. p. 151-176.

TAYLOR, Arlene G. On the subject of subjects. **The Journal of Academic Librarianship**, v. 21, n. 6, p. 484-491, nov. 1995.

TAYLOR, Margaret S. **Fundamentals of practical cataloguing**. London: George Allen and Unwin, 1948.

# 3

Luís Miguel Oliveira Machado

Wilson Roberto Veronez Júnior

Daniel Martínez-Ávila

**A indeterminação  
ontológica dos conceitos:  
interpretações linguísticas  
e psicológicas**

*The ontologic  
indetermination of concepts  
linguistic and psychological  
interpretations*

DOI: 10.31560/pimentacultural/2022.95613.3



**Resumo:**

No âmbito da Organização do Conhecimento (OC), o foco ontológico encontra-se pouco presente em estudos relativos à natureza do conceito. O estudo apresenta uma análise com esse propósito, questionando possíveis modos de existência dos conceitos (como representações mentais, habilidades cognitivas ou objetos abstratos), enquadrados em quatro diferentes leituras: a linguística, a psicológica, a epistemológica e a ontológica, com enfoque nas duas primeiras. A adequação à utilização do conceito como unidade elementar dos Sistemas de Organização do Conhecimento (SOC) é analisada de acordo com as diferentes perspectivas sobre este. De entidade mental, passando para outra existente em um reino não mental, embora também não físico, seguindo para uma outra com existência linguística objetiva.

**Palavras-Chave:** Conceito; ontologia; organização do conhecimento; sistemas de organização do conhecimento.

**Abstract:**

*In the context of Knowledge Organization (KO) the ontological focus is sometimes overlooked in studies related to the nature of the concept. This study presents an analysis with this purpose, questioning possible modes of existence of concepts (such as mental representations, cognitive abilities or abstract objects), framed in four different readings: a linguistic one, the psychological one, the epistemological one, and the ontological one; and focuses on the two first ones. The suitability of using the concept as an elementary unit of Knowledge Organization Systems (KOS) is analyzed according to the different perspectives. From a mental entity, passing to another one that exists in a non-mental realm, although also non-physical, moving on to another one with an objective linguistic existence.*

**Keywords:** *Concept; ontology; knowledge organization; knowledge organization systems.*

## 1 INTRODUÇÃO

Os estudos relativos ao *conceito*, enquanto objeto de estudo em si mesmo, são de grande relevância para a Organização do Conhecimento (OC), entendida como área ou domínio disciplinar. Essa importância é enfatizada por Van den Heuvel e Smiraglia (2010) que apontam como central para a “ciência da ordem do conhecimento” o estudo desse “fenômeno inexplicável”:

Como é que a ordem do que é conhecido pode depender de um fenômeno inexplicável? E, ainda assim, é nesse contexto que o domínio da organização do conhecimento se encontra. Enquanto não efetuarmos uma exploração aprofundada e empírica relativa aos parâmetros do fenômeno central do “conceito” não compreenderemos completamente a ciência da ordem do conhecimento.” (VAN DEN HEUVEL; SMIRAGLIA, 2010, p. 54, tradução livre dos autores<sup>7</sup>).

No âmbito da OC, o foco ontológico encontra-se pouco presente em estudos relativos ao *conceito*, em que predomina uma abordagem essencialmente epistemológica, (e.g., HJØRLAND, 2009; MACHADO; MARTÍNEZ; SIMÕES, 2019; MACULAN; LIMA, 2017; MARRADI, 2012; SMIRAGLIA; VAN DEN HEUVEL, 2013). As duas abordagens ao estudo dos conceitos, a ontológica e a epistemológica, não deverão ser confundidas com as possíveis leituras para o termo *conceito*.

No que diz respeito às diferentes leituras, quatro grandes interpretações poderão ser descritas: a linguística, a psicológica, a epistemológica e a ontológica. Uma possível análise ontológica seria questionar o modo de existência dos conceitos associados a essas

7 Texto original: “how can it be that the order of that which is known is dependent upon the behavior of an unexplained phenomenon? And yet that is where the domain of knowledge organization finds itself. Until we thoroughly and empirically explore the parameters of the central phenomenon of “concept” we will not fully understand the science of the order of knowledge.”

leituras, na linha da descrição efetuada por Smith (2008, p. 84, ênfase no original, tradução livre dos autores<sup>8</sup>):

Na visão linguística, os conceitos são *termos gerais* cujos significados foram de alguma forma regulados (ou, em algumas variantes, os conceitos são esses próprios significados). Na visão psicológica, os conceitos são *entidades mentais* análogas a ideias ou crenças. Na visão epistemológica, os conceitos são *unidades de conhecimento*, tais como o conceito de gato ou de quadrado de uma criança. E na visão ontológica, os conceitos são *abstrações de tipos ou de propriedades* (ou seja, de padrões invariantes gerais) pertencentes a entidades do mundo.

Esses possíveis modos de existência dos conceitos não se encontram em uma relação unívoca de subsunção com os três possíveis estatutos ontológicos apontados por Margolis e Laurence (2019, sec. 1). Desses três estatutos – representações mentais, habilidades cognitivas ou objetos abstratos –, apenas se encontra uma correspondência direta entre as entidades associadas à leitura psicológica e às representações mentais, para as restantes variações, a relação apresenta-se mais complexa.

Pela descrição da citação de Smith, é possível conformar as “abstrações de gêneros e propriedades” tanto ao estatuto de objetos abstratos quanto ao de habilidades cognitivas, situação indicada na posição de Lowe (2006, p. 85). Para esse autor, os conceitos seriam “formas de pensar ou apreender entidades”, isto é, seriam propriedades mentais que, de uma forma geral, os sujeitos cognoscentes possuem e, de forma individual, os conceitos seriam modos particulares de um indivíduo apreender o que o rodeia: “um conceito, assim, é um modo pelo qual alguém pode estar pensando em uma entidade. Entendidos como universais, os conceitos são atributos mentais e

8 Texto original: “on the linguistic view, concepts are *general terms* whose meanings have been somehow regimented (or, as on some variants of the view, they are these meanings themselves). On the psychological view, concepts are *mental entities* analogous to ideas or beliefs. On the epistemological view, concepts are *units of knowledge*, such as your child’s concept of a cat or of a square. And on the ontological view, concepts are *abstractions of kinds or of properties* (i.e., of general invariant patterns) belonging to entities in the world.”

entendidos como particulares são modelos mentais. Os objetos que os possuem são sujeitos pensantes, isto é, *peçoas*” (LOWE, 2006, p. 85, ênfase no original, tradução livre dos autores<sup>9</sup>).

Similarmente, nas “unidades do conhecimento”, pode-se associar os dois estatutos como aparenta ser o caso da visão dos conceitos como “unidades funcionais” dos SOC, de Peponakis e outros, em que a perspectiva psicológica é também integrada:

Reconhecendo que a noção de conceito é multifacetada, estudamos os conceitos no contexto mais amplo da cognição e no contexto dos SOC com base no princípio de que ‘os conceitos devem ser considerados os blocos de construção de todas os tipos de SOC’ [7, p. 38]. Todos os tipos de SOC, verbais ou não, baseiam-se na capacidade da mente humana de criar e gerenciar categorias; os SOC do domínio BCI não são exceção. Claramente, nem todas as categorizações constituem uma classificação [15]; contudo, a categorização é uma pré-condição.” (PEPONAKIS *et al.*, 2019, p. 435–436, ênfase no original, tradução livre dos autores<sup>10</sup>).

Quanto aos “termos gerais,” estes podem ser encarados como um conjunto estruturado de sinais usados para “materializar” uma entidade não substancial como serão os *significados*: “porque não podemos perceber com os nossos sentidos o próprio significado, sendo este uma entidade abstrata, podemos nos relacionar apenas com os conjuntos de signos que o representam (isto é, com proposições escritas, faladas ou fisicamente expressas),” (ZINS, 2007, p. 487, tradução livre dos autores<sup>11</sup>). Mesmo no contexto dos vocabulários controlados,

9 Texto original: “a concept, then, is a way someone can be thinking of an entity. Understood as universals, concepts are mental attributes and understood as particulars they are mental models. The objects that possess them are thinking subjects, that is, *persons*.”

10 Texto original: “acknowledging that the notion of concept is multi-layered, we study concepts in the wider context of cognition and in the context of KOSs based on the principle that “concepts should be considered the building blocks of all forms of KOS” [7, p. 38]. All types of KOSs, verbal or not, are based on the ability of the human mind to create and manage categories; KOSs of the LIS domain are no exception. Clearly, not every categorization constitutes a classification [15]; however, categorization is a precondition.”

11 Texto original: “because we cannot perceive with our senses the meaning itself, which is an abstract entity, we can relate only to the sets of signs (i.e., written, spoken, or physically expressed propositions), which represent it.”

no qual é possível uma relação de um-para-um entre termos e significados, os termos são encarados como “etiquetas” ou “representações” dos respectivos significados (HARPRING, 2010, p. 16; PEPO-NAKIS *et al.*, 2019, p. 435). Ontologicamente, os *significados* desses termos gerais poderão ser encarados pelas três perspectivas. Embora talvez a mais óbvia seja a referida por Zins, como entidades abstratas, seguindo a tradição dos *sentidos* de Gottlob Frege (1848-1925) ou uma versão “algorítmica” mais atual:

Deve-se notar, no entanto, que algoritmos são, em algum sentido, ‘coisas’, ‘entidades’ ou ‘objetos’ e que as palavras por eles usadas são nomes para as coisas que estes ‘chamam’ ou ativam; assim, aqueles que postularam entidades platônicas para os significados podem não estar inteiramente errados, embora, como é óbvio, não tenham previsto entidades complexas ou abstratas como os algoritmos. ... Esta abordagem ‘procedural’ ou ‘algorítmica’ do significado tem algo de comportamental ou experimental” (GOGUEN, 1974, p. 546, tradução livre dos autores<sup>12</sup>).

A abordagem processual de Goguen leva também a um entendimento dos significados como habilidades. Perspectiva igualmente presente na abordagem ao *significado* de Thornley e Gibb, a qual foi influenciada pela posição tardia de Ludwig Wittgenstein (1889-1951)<sup>13</sup>, enfatizando a relevância do “contexto sobre o conteúdo para o estudo do significado” (THORNLEY & GIBB, 2009, p. 134). Nessa abordagem, os autores aliam o uso contextual do *significado*

12 Texto original: “it should be noted, however, that algorithms are “things”, “entities”, or “objects”, in some sense, and that words are then names for the things they “call” or activate; thus those who postulated Platonic entities for meanings may not have been entirely wrong, although they did not of course envisage any such complex or abstract entities. ... This “procedural” or “algorithmic” approach to meaning has something of a behavioral or experimental tinge to it.

13 A distinção entre as duas posições de Wittgenstein, a primeira expressa no “Tractatus Logico-Philosophicus” (1922) e a segunda na obra “Investigações Filosóficas” (1953), é comumente efetuada com as designações, respetivamente, “Wittgenstein inicial” (*early Wittgenstein*) e “Wittgenstein tardio” (*later Wittgenstein*). Na abordagem inicial, o filósofo segue a linha da *linguagem ideal* procurando uma “aplicação da lógica moderna à metafísica,” na segunda, Wittgenstein “critica toda a filosofia tradicional”, incluindo a sua própria fase inicial, enveredando pela *filosofia da linguagem comum*, desenvolvendo as teorias dos “jogos de linguagem” e da “semelhança de família” (BILETZKI & MATAR, 2020).

à experiência subjetiva do entendimento desse mesmo *significado*, implicando alguma forma de representação mental:

Neste documento, o significado é entendido como uma relação complexa e frequentemente conflituaosa entre o subjetivo e o objetivo (Thornley, 2005; Thornley e Gibb, 2007). A relação entre conteúdo e contexto existe no âmbito da relação geral subjetivo/objetivo. O conteúdo é normalmente do que se trata o significado no sentido daquilo a que se refere. Esse conteúdo pode ser entendido tanto como um objeto externo quanto como a experiência subjetiva de compreender o significado da palavra que descreve o objeto. Assim, argumenta-se que, em alguns casos, o significado é melhor compreendido procurando a sua referência ou conteúdo, mas esse processo só pode ser entendido dentro do contexto mais amplo de seu uso e propósito (THORNLEY; GIBB, 2009, p. 134–135, tradução livre dos autores<sup>14</sup>).

A dificuldade em individualizar um estatuto ontológico para cada uma das quatro leituras, à exceção da perspectiva psicológica, aliada à presença desta, explícita ou implícita, em várias abordagens de diferentes autores, será indicativa da sua centralidade nesta questão. Razão pela qual se considera relevante aprofundar um pouco mais esta.

## 2 REPRESENTAÇÕES MENTAIS

Defensores da leitura psicológica, Laurence e Margolis descrevem os conceitos como “representações mentais particulares subproposicionais”:

14 Texto original: “in this paper meaning is understood as a complex and often conflicting relationship between the subjective and the objective (Thornley, 2005; Thornley and Gibb, 2007). The relationship between content and context exists within the general subjective/objective relationship. Content is normally what meaning is about in the sense of that to which it refers. This content can be understood as both an external object and also the subjective experience of understanding the meaning of the word that describes the object. Thus it is argued that in some cases meaning is best understood by looking for its reference or content but that this process can only be understood within the broader context of its use and purpose.”

As representações no nível de pensamentos completos – isto é, aqueles que podem expressar proposições inteiras – são grandes demais para serem conceitos. Consequentemente, tomaremos os conceitos como representações mentais subproposicionais. ... De acordo com praticamente todas as discussões de conceitos em psicologia, assumiremos que os conceitos são pormenores mentais.” (LAURENCE; MARGOLIS, 1999, p. 4-5, ênfase no original, tradução livre dos autores<sup>15</sup>).

Para esses autores, os outros dois estatutos ontológicos, os conceitos como habilidades ou como objetos abstratos, são compatíveis com a sua própria perspectiva. Enquanto habilidades, os autores rejeitam liminarmente que estas possam ser comportamentais atendendo às limitações que estes implicam: “o behaviorismo nos limita a disposições causais, mas mesmo as tarefas cognitivas simples tendem a exigir processos psicológicos extensos – cadeias complexas em que um estado mental causa um outro” (MARGOLIS; LAURENCE, 2007, p. 591 [nota 21], tradução livre dos autores<sup>16</sup>). Enquanto habilidades psicológicas, não só estas seriam compatíveis com a visão dos autores como estariam dependentes de processos e representações mentais subjacentes (LAURENCE; MARGOLIS, 1999, p. 5 [nota 3]).

Em relação à perspectiva dos conceitos como objetos abstratos, é necessário, primeiramente, clarificar a diferença entre esta e a defendida por Margolis e Laurence. Na “visão psicológica”, defendida pelos autores (MARGOLIS; LAURENCE, 2007, p. 563, 565, tradução livre dos autores<sup>17</sup>): “conceitos são representações mentais; eles são os

15 Texto original: “representations at the level of complete thoughts—that is, ones that may express whole propositions—are too big to be concepts. Accordingly, we will take concepts to be subpropositional mental representations. ... In accordance with virtually all discussions of concepts in psychology, we will assume that concepts are mental particulars.”

16 Texto original: “behaviorism limits us to causal dispositions, yet even simple cognitive tasks tend to require extensive psychological processes—complex chains where one mental state causes another.”

17 Texto original: “concepts are mental representations; they are the constituents of propositional attitudes such as beliefs and desires. ... Concepts are abstract objects; they are the constituents of Fregean propositions.”

constituintes de atitudes proposicionais tais como crenças e desejos”; enquanto na “visão semântica”, baseada nos *sentidos fregeanos*, “conceitos são objetos abstratos; eles são os constituintes de proposições Fregeanas”. O ponto que os autores entendem ser distintivo entre ambas se encontra na admissão ou não da existência de conceitos particulares (*token concepts*), além da sua “forma partilhada” (*type concepts*):

Seguimos Sutton ao tomar como ponto central o facto de, em tais pontos de vista [onde os conceitos são objetos abstratos], os conceitos individuais não admitirem uma distinção de tipo-token. Em contraste, a Visão Psicológica abraça totalmente a ideia de que os conceitos têm tokens (tokens de conceitos exemplares são tokens de representações mentais).” (MARGOLIS; LAURENCE, 2007, p. 587 [nota 7], tradução livre dos autores<sup>18</sup>).

Admitindo a distinção, Laurence e Margolis (1999) argumentam ser possível sustentar a visão dos conceitos serem representações mentais particulares e, simultaneamente, ser possível partilhar o “mesmo conceito”:

Quando alguém diz que duas pessoas têm o mesmo conceito, não há necessidade de supor que ela esteja dizendo que ambas possuem o mesmo token de conceito. Faria tanto sentido dizer que duas pessoas não podem preferir a mesma sentença porque ambas não podem produzir o mesmo token de sentença. Claramente, o que importa para poder pronunciar a mesma frase, ou manter o mesmo conceito, é poder ter tokens do mesmo tipo” (MARGOLIS; LAURENCE, 1999, p. 7, tradução livre dos autores<sup>19</sup>).

18 Texto original: “we follow Sutton in taking the central point to be that on such views [where concepts are abstract objects] individual concepts don’t admit of a type-token distinction. In contrast, the Psychological View fully embraces the idea that concepts have tokens (token concepts are token mental representations).” O termo “token” apresenta dificuldades na tradução para a língua portuguesa, embora de forma aproximada se possa usar termos como “particular” ou “exemplar,” entendeu-se ser preferível manter o termo original.

19 Texto original: “when someone says that two people have the same concept, there is no need to suppose that she is saying that they both possess the same token concept. It would make as much sense to say that two people cannot utter the same sentence because they cannot both produce the same token sentence. Clearly what matters for being able to utter the same sentence, or entertain the same concept, is being able to have tokens of the same type.”

Uma vantagem, de acordo com os autores citados, dessa perspectiva está nela continuar consistente, com ou sem os *sentidos fregeanos*. A razão para tal está na afirmação que as representações mentais *têm* significados e não que estas são significados. Deixando em aberto a questão desses significados serem entidades, Margolis e Laurence tornam a sua “visão psicológica” compatível com as duas alternativas: com a reivindicação destes serem entidades, nomeadamente *sentidos fregeanos*; como com a posição que argumenta não existirem tais entidades (MARGOLIS; LAURENCE, 2007, p. 590 [nota 14]). Os *sentidos*, independentemente do seu estatuto ontológico, são usados pelos autores para auxiliar a explicar a tipificação das representações mentais e proporcionar parte das condições necessárias à individualização de conceitos:

Neste modelo, as crenças e outras atitudes proposicionais envolveriam tokens de representações mentais que têm outras representações—conceitos—como seus constituintes. Os sentidos seriam os valores semânticos destas representações. Ou seja, além de ter objetos e propriedades mundanas como seus referentes, as representações mentais (como as palavras, no relato original de Frege) também *teriam* sentidos. Desta forma, os sentidos ajudam a tipificar a representação mental; eles fornecem parte das condições para individualizar conceitos” (MARGOLIS; LAURENCE, 1999, p. 8, ênfase adicionada, tradução livre dos autores<sup>20</sup>).

A compatibilidade referida, entre a visão psicológica e os objetos abstratos da visão semântica, resulta na “visão mista dos conceitos” que procura incorporar os pontos fortes das duas tradições:

A Visão Mista visa combinar as forças explicativas das duas tradições dominantes no estudo dos conceitos. Seguindo a Visão Psicológica, ela abraça de forma plena as representações

20 Texto original: “On this model, beliefs and other propositional attitudes would involve token mental representations that have other representations—concepts—as their constituents. Senses would come in as the semantic values of these representations. That is, in addition to having worldly objects and properties as their referents, mental representations (like words, on Frege’s original account) would have senses too. In this way, senses help to type mental representation; they provide part of the conditions for individuating concepts.”

mentais. As representações mentais explicam a produtividade do pensamento e o facto de que os processos mentais podem ser, de uma só vez, físicos e lógicos. Ao mesmo tempo, seguindo a Visão Semântica, o apelo aos sentidos fornece uma explicação para o facto de duas expressões (incluindo duas expressões mentais) poderem ter a mesma referência, mas terem um significado cognitivo diferente. A sugestão é das representações mentais poderem apresentar um referente de diferentes formas em virtude da expressão de sentidos diferentes. Os sentidos continuam a ser entidades intermediárias, situando-se entre expressões e referências, só que agora as expressões em questão ocorrem num sistema interno de representação” (MARGOLIS; LAURENCE, 2007, p. 569, tradução livre dos autores<sup>21</sup>).

Se bem que Margolis e Laurence considerem a “visão mista” coerente, não a defendem. Para os autores, a visão psicológica, além de explicar a produtividade do pensamento humano e os aspectos físicos e lógicos dos processos mentais, não necessita recorrer a entidades abstratas, como os *sentidos fregeanos*, para explicar como diferentes expressões com o mesmo referente podem ter distintos significados cognitivos ou “modos de apresentação,” como diria Frege (1892/1948, p. 210). Esses *modos* são identificados com “propriedades mentais” e não com objetos abstratos de um “terceiro reino”, como os *sentidos fregeanos*:

A Visão Psicológica tem várias opções promissoras para lidar com este problema [do modo de apresentação]. O que estes partilham é a ideia de que os modos de apresentação devem ser identificados com propriedades de representações mentais. Ao psicologizar os modos de apresentação, já não há um enigma sobre como são apreendidos ou porque não podem ser

21 Texto original: “the Mixed View aims to combine the explanatory strengths of the two dominant traditions in the study of concepts. Following the Psychological View, it wholeheartedly embraces mental representations. Mental representations explain the productivity of thought and the fact that mental processes can be, at once, physical and rational. At the same time, following the Semantic View, the appeal to senses provides an explanation for the fact that two expressions (including two mental expressions) can have the same referent yet differ in cognitive significance. The suggestion is that mental representations can present a referent in different ways in virtue of expressing different senses. Senses continue to be intermediary entities, standing between expressions and referents, only now the expressions in question occur in an internal system of representation.”

apreendidos de diferentes maneiras. Isto é pela simples razão de já não existir uma lacuna entre as mentes e os modos de representação. Os modos de apresentação estão diretamente incorporados nas nossas mentes e no seu funcionamento” (MARGOLIS; LAURENCE, 2007, p. 581, tradução livre dos autores<sup>22</sup>).

A potencial solução para a apreensão de diferentes *modos de apresentação* ou *sentidos* para o mesmo referente, apresentada por esta visão psicológica dos conceitos, acresce às vantagens processuais enumeradas. Apesar desse potencial explicativo, se o posicionamento relativo aos conceitos não for operacional, a colocação destes na mente não será adequada ao seu uso como unidades elementares dos SOC. Quanto aos abstratos “conceitos-tipo partilháveis”, dentro da visão psicológica, estes não deixam de ser considerados mentais, como é possível verificar na exposição de Davis (2003 p. 312, 314, 317, ênfase no original, tradução livre dos autores<sup>23</sup>):

Os pensamentos neste sentido [no sentido objetivo] são *eventos*, especificamente, *eventos mentais*. Os pensamentos são o tipo de eventos que ocorrem às pessoas ou outros seres inteligentes, e qualquer ser em que um pensamento ocorre ipso facto tem uma mente. Para que um pensamento ocorra é necessário a ocorrência de um ato de pensar. Os pensamentos neste sentido são, no entanto, *eventos-tipo* e não *eventos-token*. Os pensamentos são de fato abstratos, mas apenas porque são *eventos-tipo*. Além disso, o pensamento de que o céu é azul é um tipo de evento mental. ... Usaremos os termos *ideia* e *conceito* para significar *pensamentos* ou *partes cognitivas de pensamentos*”.

22 Texto original: “the Psychological View has several promising options for dealing with this [mode of presentation] problem. What these share is the idea that modes of presentation are to be identified with properties of mental representations. By psychologizing modes of presentation, there is no longer a puzzle about how they are grasped or why they can't be grasped in different ways. This is for the simple reason that there is no longer a gap between minds and modes of representations. Modes of presentation are directly built into our minds and how they function.”

23 Texto original: “thoughts in this sense [in the objective sense] are *events*, specifically, *mental events*. Thoughts are the sorts of events that occur to people or other intelligent beings, and any being to which a thought occurs ipso facto has a mind. For a thought to occur is for some thinking to take place. Thoughts in this sense are *event-types*, however, rather than *event-tokens*. ... Thoughts are indeed abstract, but only because they are *event-types*. Moreover, the thought that the sky is blue is a type of mental event. ... We will use the terms *idea* and *concept* to mean *thoughts* or *cognitive parts of thoughts*.”

O caráter subjetivo, inerente à representação mental dos conceitos, não se adequa à objetividade pretendida para as unidades dos SOC. Em contrapartida, na tese de Frege, essa característica objetiva é reclamada:

O próprio Frege descarta as representações mentais, alegando que são muito subjetivas. Ele argumenta que a comunicação requer significados compartilhados, mas que as pessoas não podem literalmente ter as mesmas representações mentais, uma vez que as representações mentais são componentes de nossas experiências subjetivas individuais. Os sentidos, por outro lado, são objetivos e eminentemente compartilháveis” (MARGOLIS; LAURENCE, 2007, p. 566, tradução livre dos autores<sup>24</sup>).

Na próxima seção, essa potencial ligação será explorada, procurando-se clarificar a noção dos abstratos *sentidos fregeanos*.

### 3 ABSTRAÇÕES NÃO MENTAIS

A importância da tese de Frege para a OC, especificamente no que concerne aos conceitos como unidade do conhecimento, é explicitada por Dahlberg (1992 p. 65–66, ênfase no original, tradução livre dos autores<sup>25</sup>):

24 Texto original: “Frege himself dismisses mental representations, claiming they are too subjective. He argues that communication requires shared meanings but that people can’t literally have the same mental representations since mental representations are components of our individual subjective experiences. Senses, on the other hand, are objective and eminently shareable.”

25 Texto original: “now, what indeed is a concept? The German philosophers Immanuel Kant (3) and Gottlieb Frege (4) inferred already the generation of concepts by predication. Our assumption that a concept is created by predicating about an object of concern, a so called *referent* is rooted in the writings of these philosophers. ... What we need to mention here, however, is the understanding of a concept as a *knowledge unit*. In the preceding section we stated that a concept is generated by predications. One cannot predicate a true fact without expressing one’s knowledge of something. Thus every predication yields a *knowledge element* and the necessary sum total of predications can be synthesized into the corresponding *knowledge unit*.”

Agora, o que é de facto um conceito? Os filósofos alemães Immanuel Kant (3) e Gottlieb Frege (4) inferiram já a geração de conceitos por predicção. A nossa suposição de que um conceito é criado pela predicção de um objeto de interesse, o chamado *referente*, está enraizado nos escritos destes filósofos. ... O que precisamos de mencionar aqui, contudo, é a compreensão de um conceito como uma *unidade de conhecimento*. Na secção anterior, afirmámos que um conceito é gerado por predições. Não se pode predicar um facto verdadeiro sem se expressar o conhecimento de algo. Assim, cada predição produz um *elemento de conhecimento* e a soma total das predições necessárias pode ser sintetizada na *unidade de conhecimento* correspondente.

Outro autor da área da OC, Frické, não limita a relação da tese de Frege à fundamentação para o seu entendimento do que seria um conceito. O autor faz corresponder a sua posição relativamente aos conceitos com os objetos abstratos existentes em um “terceiro reino fregeano”:

Então, o que é um conceito? Neste contexto, a palavra ‘conceito’ é usada de forma bastante semelhante à do discurso comum, equivalendo aproximadamente à de ‘noção geral’ ou de ‘ideia geral’ ou mesmo de ‘significado’. Muitos descrevem conceitos como sendo mentais ou construções mentais; contudo, consideramo-los como abstrações ou objetos abstratos (no padrão Fregeano do terceiro reino) (FRICKÉ, 2012, p. 33, tradução livre dos autores<sup>26</sup>).

Frické relaciona esse “terceiro reino” de Frege ao “mundo 3” do “conhecimento objetivo” de Karl Popper (1902-1994):

Frege invocou um ‘terceiro reino’. E esse terceiro reino consistia em abstrações (que não eram ideias mentais ou psicológicas). A visão de Frege foi mais tarde retomada por outros; por exemplo, o filósofo do século XX Karl Popper distinguiu entre três mundos:

26 Texto original: “so, what is a concept? In this context, the word ‘concept’ gets used in pretty well the same way as in ordinary speech and life, and that amounts roughly to ‘general notion’ or ‘general idea’ or even ‘meaning’. Many describe concepts as being mental or mental constructions; however, we regard them as abstractions or abstract objects (in the standard Fregean third realm).”

o mundo físico, o mundo mental, e o mundo dos conteúdos abstratos” (FRICKÉ, 2012, p. 30, tradução livre dos autores<sup>27</sup>).

A ligação é assumida pelo próprio Popper (1994, p. 106, tradução livre dos autores<sup>28</sup>): “o meu terceiro mundo assemelha-se mais ao universo do conteúdo objetivo do pensamento de Frege”. Frege inclui no seu “universo de conteúdo objetivo” os *sentidos* e os *pensamentos*, distinguindo-os de um outro construído – as *concepções*, estas sim entendidas como mentais e pessoais:

A concepção é subjetiva: a concepção de um homem não é a de outro. ... Isso constitui uma distinção essencial entre a concepção e o sentido do signo, que pode ser propriedade comum de muitos e, portanto, não é uma parte ou um modo da mente individual. Pois dificilmente se pode negar que a humanidade possui um estoque comum de pensamentos que são transmitidos de uma geração para outra. ... Até agora consideramos o sentido e os referentes apenas de tais expressões, palavras ou sinais que chamamos de nomes próprios. Agora indagamos sobre o sentido e o referente de uma sentença declarativa inteira. Tal frase contém um pensamento.<sup>5</sup> [nota 5: por pensamento entendo não o desempenho subjetivo do pensamento, mas seu conteúdo objetivo, que pode ser propriedade comum de vários pensadores” (FREGE, 1948, p. 212, 214 nota incluída, tradução livre dos autores<sup>29</sup>).

27 Texto original: “Frege invoked a ‘third realm’. And that third realm consisted of abstractions (which were not mental or psychological ideas). Frege’s view was later taken up by others; for example, the twentieth century philosopher Karl Popper distinguished between three worlds: the physical world, the mental world, and the world of abstract contents.”

28 Texto original: “my third world resembles most closely the universe of Frege’s objective contents of thought.”

29 Texto original: “The conception is subjective: One man’s conception is not that of another. ... This constitutes an essential distinction between the conception and the sign’s sense, which may be the common property of many and therefore is not a part or a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another. ... So far we have considered the sense and referents only of such expressions, words, or signs as we have called proper names. We now inquire concerning the sense and referent of an entire declarative sentence. Such a sentence contains a thought.<sup>5</sup> [note 5: by a thought I understand not the subjective performance of thinking but its objective content, which is capable of being the common property of several thinkers.”

Na terminologia de Frege, não será o *sentido* e sim a *conceção* que mais se aproxima do *conceito* da visão psicológica. Sendo a subjetividade mental restrita às *concepções*, a objetividade dos objetos físicos, por seu turno, não poderia ser idêntica à objetividade que o filósofo atribui aos *sentidos* e *pensamentos*. Na tese de Frege, esses *sentidos* e *pensamentos* assemelham-se, por um lado, às *concepções* pelo fato de não serem perceptíveis aos cinco sentidos e, por outro, aos objetos físicos pela sua natureza intersubjetiva e por não necessitarem de um portador (BONEVAC, 2017, pt. 13:25). Similarmente à distinção *type-token*, relativa aos conceitos da visão psicológica, apenas as manifestações de *sentidos* e *pensamentos* fregeanos necessitariam de um portador. Entretanto, esses *sentidos* e *pensamentos* distinguem-se tanto das *concepções* quanto dos objetos físicos no modo como as pessoas os podem possuir. Se o ato de obtenção, por parte de um ser humano, de um objeto físico pode ser entendido como sendo extrínseco e o de uma *concepção* intrínseco, a apreensão (*grasp*) de um *sentido* é suposto ser um meio-termo entre os dois, embora com maiores similaridades com a via intrínseca (BONEVAC, 2017, pt. 33:20).

A dificuldade está na clarificação deste ato de apreensão dos *sentidos*: “mas em que consiste exatamente a apreensão? Claramente, a apreensão é uma metáfora para uma relação cognitiva que precisa de ser explicada” (MARGOLIS; LAURENCE, 2007, p. 580, tradução livre dos autores<sup>30</sup>). Frege não só não clarifica essa questão como parece não lhe atribuir grande importância: “Frege não aborda diretamente esta questão, nem parece estar preocupado com a pobreza fenomenológica da apreensão” (SCHWEIZER, 1991, p. 275, tradução livre dos autores<sup>31</sup>). Não fosse a atribuição, da parte de Frege, de uma natureza não mental aos *sentidos* e *pensamentos*, a solução dos “eventos-tipo mentais” de Davis poderia ser uma resposta:

30 Texto original: “but what exactly does grasping consist in? Clearly, grasping is a metaphor for a cognitive relation that needs to be explicated.”

31 Texto original: “Frege does not directly address this issue, nor does he seem to be concerned with the phenomenological poverty of grasping.”

Para que você tenha o pensamento de que o céu é azul é necessário que esse evento-tipo mental lhe ocorra. Daí que esse evento-tipo ocorra a um sujeito sempre e onde quer que este esteja a pensar que o céu é azul. É isso que é 'apreender' ou 'ter' um pensamento. Duas pessoas 'partilham' um pensamento se o mesmo ocorrer a ambas. Um pensamento não tem localização espaçotemporal em si, embora todos os seus tokens tenham, assim como um tipo de sentença não tem localização espaço-temporal, embora todos os seus tokens tenham. ... Acreditar, desejar, pretender, ou esperar que o céu seja azul envolve mais do que a ocorrência do pensamento de que o céu é azul, mas não podemos, portanto, inferir que pensar o pensamento consiste em mais do que a ocorrência do pensamento, ou que o pensamento não é mental. ... Nem a visão da relação objetiva implica que pensar envolve *ver* pensamentos em qualquer sentido, ou ter algo na mente "*direccionado*" ao pensamento (FREGE 1918: 26). Isso tornaria o pensamento misterioso" (DAVIS, 2003, p. 315, ênfase no original, tradução livre dos autores<sup>32</sup>).

Contudo, para Frege, o *pensamento* é o significado ou conteúdo não mental de um estado psicológico e não o estado em si: "vale a pena notar que Frege usa o termo 'pensamento' para significar proposições, por isso para Frege os pensamentos não são estados psicológicos, mas sim significados de estados psicológicos" (Margolis & Laurence, 2019, sec. 1.3, tradução livre dos autores<sup>33</sup>). A procura de uma resposta não psicológica torna a situação mais complexa. Não para o próprio Frege, dado o seu aparente conformismo com esta, mas

32 Texto original: "for you to think the thought that the sky is blue is for that mental event-type to occur to you. Hence that type occurs to a subject whenever and wherever the subject is thinking the thought that the sky is blue. That is what it is to "grasp" or "have" a thought. Two people "share" a thought if it occurs to both of them. A thought has no spatio-temporal location itself, although all of its tokens do, just as a sentence type has no spatio-temporal location, although all of its tokens do. ... Believing, desiring, intending, or hoping that the sky is blue all involve more than the occurrence of the thought that the sky is blue, but we cannot therefore infer that thinking the thought consists in more than the occurrence of the thought, or that the thought is not mental. ... Nor does the objective relation view entail that thinking involves *viewing* thoughts in any sense, or having something in the mind "*aimed at*" the thought (FREGE 1918: 26). That would make thinking mysterious."

33 Texto original: "it is worth noting that Frege uses the term 'thought' to stand for propositions, so for Frege thoughts are not psychological states at all but rather the meanings of psychological states."

para os seus seguidores. Schweizer, por exemplo, sugere uma solução, algo rebuscada, que envolve uma analogia com o fenômeno da “visão cega”<sup>34</sup>. De forma similar ao que ocorre na “visão cega,” apreender um sentido (*grasping a sense*), segundo Schweizer, corresponderia a um fenômeno perceptivo, no qual as manifestações da relação entre uma “entidade semântica externa” e o genuíno “objeto da consciência” correspondente não incluem um vasto leque de ideias conscientes:

Para tanto, será desenvolvida e explorada uma forma de visão ligeiramente modificada, na qual o objeto físico da visão não produz nenhuma apresentação consciente. Esta forma ampliada de visão, aqui apelidada de ‘ver\*’, se mostrará aplicável tanto aos autômatos quanto aos seres humanos, e como o ver\* reflete de perto a fenomenologia real do entendimento humano, é assim possível restaurar a analogia básica entre visão e cognição. ... Assim, se apreender é interpretado simetricamente com o ver\*, então o realismo semântico de Frege torna-se compatível com grande parte da fenomenologia real da aquisição e uso da linguagem” (SCHWEIZER, 1991, p. 275–276, 277, tradução livre dos autores<sup>35</sup>).

A viabilidade desta, ou quaisquer outras possíveis respostas, não será aqui discutida. Embora esta seja relevante para os proponentes da equiparação dos conceitos aos *sentidos fregeanos*, para a presente investigação, interessa o estatuto ontológico destes e a sua ligação com o que Frege entende por *conceito*. Isto porque, para o filósofo, nem as *concepções* nem os *sentidos* são *conceitos*. Na sua teoria, Frege atribui maior proximidade ontológica entre *sentidos* e *conceitos* que entre estes

34 “A Visão cega é a capacidade de reconhecer objetos num ambiente mesmo sem ter a consciência de os conseguir ver. O efeito ocorre em cegueiras corticais, onde o cérebro consegue processar informações que os olhos recebem. O cego é capaz, se for treinado, de reconhecer cores e expressões faciais.” Disponível em: <https://icnagency.com/neuro-cards/a-visao-cega>. Acesso em: 30 de out. 2021.

35 Texto original: “to this end, a slightly modified form of vision will be developed and explored, in which the physical object of sight does not produce any conscious presentations. This extended form of vision, herein dubbed ‘seeing\*’, will be shown to be applicable both to automata and to human beings, and since seeing\* closely reflects the actual phenomenology [sic] of human understanding, it is thereby possible to restore the basic analogy between sight and cognition. ... Thus if grasping is interpreted symmetrically with seeing\*, then Frege’s semantic realism becomes compatible with much of the actual phenomenology of language acquisition and use.”

e as *concepções*, atendendo à natureza mental, isto é, subjetiva, destas últimas ao contrário da objetividade conferida aos outros dois.

## 4 OBJETIVIDADE LINGUÍSTICA

Tal como os *sentidos*, os *conceitos*, para Frege, são ontologicamente objetivos. Mas, mais uma vez, a objetividade dos *conceitos* não é a mesma da dos *objetos físicos* e, dentro da teoria de Frege, nem da dos *sentidos*. A distinção implica uma subordinação da análise ontológica à linguística:

Uma característica distintiva desta visão é que as categorias ontológicas estão implícitas e são determinadas pelas categorias linguístico-lógicas. ... Como princípio metafísico e metodológico básico, é independente dos detalhes da lógica de Frege e da sua ontologia específica de objetos e funções. Por razões profundas que dizem respeito ao essencial da sua lógica, Frege considerava as categorias lógicas de nomes e predicados assimetricamente: embora ambas as categorias sejam essenciais, há um sentido em que os nomes formam a categoria básica e os predicados são concebidos em termos de nomes; além disso, os termos funcionais dividem-se em níveis, enquanto a categoria de nomes é homogênea - todos os nomes são tratados da mesma forma; não há distinção lógica entre, por exemplo, concreto e abstrato ou entre universal e particular. Os critérios para as categorias são puramente lógicos e, nessa perspectiva, estas distinções não importam” (BAR-ELLI, 2015, p. 292–293, tradução livre dos autores <sup>36</sup>).

<sup>36</sup> Texto original: “a distinctive characteristic of this view is that the ontological categories are implied and determined by the linguistic-logical categories. ... As a basic metaphysical and methodological principle it is independent of the details of Frege’s logic and of his specific ontology of objects and functions. For deep reasons that pertain to the essentials of his logic Frege regarded the logical categories of names and predicates asymmetrically: though both categories are essential, there is a sense in which names form the basic category and predicates are conceived in terms of names; moreover, functional terms divide into levels, while the category of names is homogenous – all names are treated alike; there is no logical distinction between e.g. concrete and abstract, between universal and particular. The criteria for the categories are purely logical, and from this perspective these distinctions don’t matter.”

Frege segue uma abordagem lógico-linguística, descartando as distinções ontológicas tradicionais, tais como particular-universal e concreto-abstrato (BAR-ELLI, 2015, p. 288), ao estabelecer a sua teoria de referência (*bedeutung*) e sentido (*sinn*). Nesta, as duas categorias essenciais, “objetos” e “funções,” pertencem ao “reino” das referências, atribuindo aos *sentidos* um estatuto ontológico *sui generis* fora desse “reino” (BAR-ELLI, 2015, p. 288). A distinção entre os dois “reinos” é essencial para entender a diferença entre *sentido* e *conceito* em Frege.

Os *conceitos fregeanos* são entendidos como referências de predicados gramaticais: “o conceito (como eu entendo a palavra) é predicativo.<sup>1</sup> [nota 1: é, de facto, a referência de um predicado gramatical” (Frege, 1951, p. 169 nota incluída, tradução livre dos autores<sup>37</sup>). Atendendo ao seu caráter predicativo, Frege considera-os funções: “vemos assim quão estreitamente aquilo a que se chama um conceito em lógica está ligado àquilo a que chamamos uma função. De fato, podemos dizer imediatamente: um conceito é uma função cujo valor é sempre um valor-verdade” (Frege, 1997, p. 139, tradução livre dos autores<sup>38</sup>). Esta particularidade, de ter sempre um dos dois possíveis “valores-verdade” (*a verdade* ou *o falso*), permitiria determinar de forma lógica os “objetos” que “caem dentro” de um conceito, isto é, a sua extensão:

Frege (1892) fez uma distinção entre conceito e objeto, definindo um conceito como uma função que possui um valor de verdade. O valor de um objeto para um argumento pode ser um dos objetos abstratos ‘o Verdadeiro’ ou ‘o Falso’. Segundo Frege, o conceito de ser humano é entendido como uma função que tem ‘o Verdadeiro’ como valor para qualquer argumento que seja humano, e ‘o Falso’ como valor para todo o resto. Na terminologia de Frege, um objeto para o qual um conceito tem valor ‘o

37 Texto original: “the concept (as I understand the word) is predicative.<sup>1</sup> [note 1: is, in fact, the reference of a grammatical predicate.]”

38 Texto original: “we thus see how closely that which is called a concept in logic is connected with what we call a function. Indeed, we may say at once: a concept is a function whose value is always a truth-value.”

Verdadeiro' diz-se que 'cai sob' o respetivo conceito" (FRIEDMAN; SMIRAGLIA, 2013, p. 23, tradução livre dos autores<sup>39</sup>).

É conveniente referir que os *conceitos*, para Frege, são apenas um caso particular de função, existindo outras como, por exemplo, as *propriedades* e as *relações*. Em comum, todas estas funções têm a "necessidade lógica" de um "objeto auto-subsistente" (da categoria "objetos") para ficarem completas, ou "saturadas," na terminologia fregeana (BAR-ELLI, 2015, p. 292). Tendo em mente que a ontologia em Frege é subordinada às exigências da lógica e de uma teoria do significado da linguagem, as categorias *objetos* e *funções* são "apenas" referentes e, como tal, insuficientes para os intentos do filósofo. É nesse contexto que uma terceira categoria, os *sentidos*, é adicionada, de forma a explicar os diferentes conteúdos cognitivos, ou "modos de apresentação," que os referentes podem apresentar em proposições:

Um resultado importante foi uma ontologia consistida por objetos e funções categoricamente distintos, mas ambos reais e objetivos. Vimos também que ele chegou à conclusão de que os sentidos - maneiras como as coisas nos são dadas como referências de termos em proposições - são, portanto, necessários tanto na lógica como numa teoria do significado. Assim, a realidade, segundo ele, deve consistir não apenas nas referências de termos de proposições verdadeiras, mas também em seus sentidos" (BAR-ELLI, 2015, p. 295, tradução livre dos autores<sup>40</sup>).

Em termos linguísticos, na tese de Frege, os *nomes próprios* estão associados à categoria "objetos" e os *nomes comuns* às funções

39 Texto original: "Frege (1892) made a distinction between concept and object, by defining a concept as a function that has a truth-value. An object's value for an argument can be either of the abstract objects "the True" or "the False." According to Frege, the concept of being human is understood as a function that has "the True" as value for any argument that is human, and "the False" as the value for everything else. In Frege's terminology, an object for which a concept has value "the True" is said to "fall under" the concept."

40 Texto original: "an important outcome was that ontology consists of objects and functions, which are categorically distinct, but are both real and objective. We have also seen that he came to the conclusion that senses - ways things are given to us as the references of terms in propositions - are thus required in logic and in a theory of meaning. So reality, according to him, must consist not only of the references of terms of true propositions, but also of their senses."

específicas a que o filósofo chama de “conceitos”. Assim, na linguagem lógica de aplicação à ciência de Frege, além dos *sentidos* que os *nomes* deverão ter (para serem “legitimamente” *nomes*), deverão ter também referentes (*objetos* para os nomes próprios e *conceitos* para os comuns). A importância da relação referencial leva Frege a preferir usar a designação “nome conceitual” em vez de “nome comum”. O uso desses nomes comuns deverá ser entendido como estando a nomear um conceito e não uma pluralidade de objetos:

A palavra “nome comum” leva erroneamente a supor que o nome comum se relaciona com os objetos, no essencial, do mesmo modo que o faz o nome próprio, apenas com a diferença de que este só nomeia um único objeto enquanto que aquele se aplica em geral a diversos objetos. Isto, porém, é falso, e por essa razão prefiro dizer “termo conceitual” (*Begriffswort*) em lugar de “nome comum” (*Gemeinname*). Um nome próprio deve ter pelo menos um sentido (na acepção em que entendo essa palavra), senão ele será apenas uma mera sequência vazia de sons, e seria ilegítimo chamá-lo de nome. Mas para que tenha um uso em ciência deve-se exigir também que ele tenha uma referência, que designe ou nomeie um objeto. ... Também o termo conceitual deve ter um sentido e, para que tenha um uso científico, deve ter uma referência; esta, porém, não é nem um objeto, nem uma pluralidade de objetos, mas um conceito (FREGE, 2009, p. 168, ênfase no original).

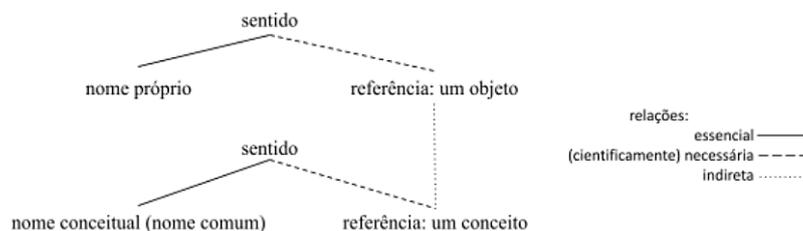
Fulcral na relação referencial, em ambos os casos (nomes próprios e “conceituais”), é o papel de mediação do *sentido* entre *nomes* e respectivos *referentes*. Mediação que, no contexto da linguagem lógica desenvolvida por Frege, deverá ser claramente determinada e objetiva:

O nome próprio se relaciona, mediante o sentido, e só mediante o sentido, com o objeto. ... Um termo conceitual pode, do ponto de vista lógico, ser absolutamente impecável sem que haja um objeto com o qual ele se relacione mediante seu sentido e sua referência (o próprio conceito). Esta relação [do termo conceitual] com um objeto é, como se vê, uma relação mais indireta e inessencial, de modo que parece pouco conveniente classificar os termos conceituais segundo caíam

sob o conceito correspondente, um, vários ou nenhum objeto. A lógica deve exigir, tanto do nome próprio como do termo conceitual, que a transição do nome para o sentido e a do sentido para a referência sejam determinadas sem deixar margem a dúvidas. Do contrário, não mais se poderia falar de uma referência (Frege, 2009, p. 168–169).

Atendendo ao descrito relativamente às relações entre *nomes*, *sentidos* e *referentes* na teoria de Frege, estas são de distintas naturezas. A relação entre *nomes* e *sentidos* poderá ser considerada essencial devido à necessidade intrínseca dos primeiros para a sua legitimação. Na relação entre os *sentidos* e os *referentes*, o filósofo apresenta-a como necessária para que os *nomes* tenham utilidade científica, sendo estes essenciais na sua abordagem lógico-linguística. A ligação entre *nomes* e *referentes* mediada pelos *sentidos* pode, assim, ser encarada como a união das duas relações parcelares (ver Figura 1). Por fim, a relação “indireta e inessencial”, entre um *nome conceitual* e um *objeto*, pode também ser entendida da mesma forma, como um percurso que passa, necessariamente, pelos *sentidos* e *conceitos* respetivos (ver Figura 1).

**Figura 1 – Esquema interpretativo das relações entre ‘nomes’, ‘sentidos’ e ‘referentes’ na teoria de Frege**



Fonte: Elaborado pelos autores, 2021.

Apesar da importância da relação de mediação dos *sentidos*, mais uma vez, a minguada informação que Frege fornece a respeito da sua exata atuação leva à procura de possíveis soluções exteriores à teoria do filósofo. Uma interpretação comum é efetuada usando a “teoria das descrições” de Russell:

Infelizmente, Frege não nos diz muito sobre como exatamente estes objetos abstratos escolhem ou apresentam as suas referências. Exatamente o que é que faz sentido uma “forma de determinar” ou “modo de apresentar” uma referência? Na esteira da teoria das descrições de Russell, um sentido Fregeano é frequentemente interpretado como um conjunto de informações descritivas ou critérios que escolhem a sua referência em virtude da simples referência satisfazendo ou encaixando essa informação descritiva” (KLEMENT, c2021, sec. 4, tradução livre dos autores<sup>41</sup>).

Será esta *função predicativa e descritiva* de um referente a influência fregeana que Dahlberg refere. A descrição da autora relativamente à formação de um conceito assim o indicia:

Cada afirmação verdadeira sobre um determinado item de referência fornece um elemento de conhecimento sobre ele juntamente com uma característica de seu conceito. A soma de afirmações necessárias sobre tal item de referência forma o conjunto das características de seu conceito apresentando distintamente o conteúdo do mesmo” (DAHLBERG, 2009, p. 171, tradução livre dos autores<sup>42</sup>).

A natureza definível dos conceitos, explicitada por Dahlberg, também pode ser encontrada na perspectiva de Frege:

Central, portanto, na teoria de conceitos de Frege não é que os conceitos sejam funções cujos valores são sempre valores de verdade - por mais importante que esta tese seja - mas a sua afirmação universal de que todos os conceitos, para serem conceitos, devem ser fechados: compostos de conjuntos definitivos de características que determinam a classificação exata

41 Texto original: “unfortunately, Frege does not tell us very much about exactly how these abstract objects pick out or present their references. Exactly what is it that makes a sense a ‘way of determining’ or ‘mode of presenting’ a reference? In the wake of Russell’s theory of descriptions, a Fregean sense is often interpreted as a set of descriptive information or criteria that picks out its reference in virtue of the reference alone satisfying or fitting that descriptive information.”

42 Texto original: “each true statement about a certain item of reference delivers a knowledge element about this together with a characteristic of its concept. The sum of necessary statements about such an item of reference forms the whole of characteristics of its concept, it presents distinctly the contents of it.”

dos objetos e o uso correto da linguagem (WEITZ, 1988, p. 198, tradução livre dos autores<sup>43</sup>).

Ambos, Dahlberg e Frege, atribuem um caráter “fechado” aos conceitos através de um conjunto completo e definitivo de características que os claramente definem. Apesar da proximidade na posição de Dahlberg, há um alargamento da restrita função predicativa, que Frege atribuía aos conceitos, ao incluir nestes o papel desempenhado pelos *sentidos fregeanos*. Enquanto Frege entendia que o *conceito* tinha uma existência categorial de “referente” separada da dos *sentidos*, na teoria de Dahlberg, os dois podem ser entendidos como componentes do *conceito* como uma única entidade.

A definição dessa entidade: “um *conceito* é uma unidade de conhecimento que integra as características de um referente por um termo ou um nome” (DAHLBERG, 1978, p. 144, ênfase no original, tradução livre dos autores<sup>44</sup>) apresenta, de forma simples, os três componentes desta: o referente, suas características e respetiva forma verbal. A “simplicidade” da definição encerra uma entidade complexa, não só pela abrangência dos seus componentes como pelo caráter instrumental que Dahlberg lhe confere. Para a autora, o *conceito* como unidade do conhecimento de um SOC ganha uma objetividade “concreta” com o propósito de “materializar” o conhecimento a ser representado. Ideia explicitada em um dos fundamentos da teoria da classificação defendida por Dahlberg (1976 p. 89, tradução livre dos autores<sup>45</sup>): “o reconhecimento do conceito como o elemento material dos sistemas de classificação”.

43 Texto original: “central, then, in Frege’s theory of concepts is not that concepts are functions whose values are always truth values - as important as this thesis is - but his universal claim that all concepts, to be concepts, must be closed: composed of definitive sets of characteristics that determine exact classification of objects and correct use of language.”

44 Texto original: “a *concept* is a knowledge unit comprising the characteristics of a referent by a term or a name.”

45 Texto original: “the recognition of the concept as the material element of classification systems”.

## 5 NOTA CONCLUSIVA

O estudo apresentou uma abordagem a possíveis modos de existência do *conceito* nas quais este foi, de certa forma, tornando-se mais “concreto”. De entidade mental, passando para outra existente em um reino não mental, embora também não físico, seguindo para uma outra com existência linguística objetiva e, por fim, para a sua “concretização material” como unidade elementar de um SOC. Essa complexa “unidade do conhecimento” exige um estudo dedicado que se pretende efetuar no futuro.

## 6 REFERÊNCIAS

BAR-ELLI, Gilead. On the Ontological Status of Senses (Sinne) in Frege. **Revista Portuguesa de Filosofia**, v. 71, n. 2/3, p. 287–305, 2015. DOI 10/gmvf62.

BILETZKI, Anat; MATAR, Anat. Ludwig Wittgenstein. (Edward N. Zalta, ed.) *In: The Stanford Encyclopedia of Philosophy*. Spring 2020 Stanford: Stanford University, 2020. Disponível em: <https://plato.stanford.edu/entries/wittgenstein/>. Acesso em: 12 abr. 2021.

BONEVAC, Daniel. **Frege on Thought**. *The Analytic Tradition*, Spring 2017. [S. l.: s. n.], 2 mar. 2017. Disponível em: <https://youtu.be/3MWb0GXF5Z0>. Acesso em: 13 abr. 2021.

DAHLBERG, Ingetraut. Classification theory, yesterday and today. **International Classification**, v. 3, n. 2, p. 85–90, 1976.

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory for INTERCONCEPT. **International Classification**, v. 5, n. 3, p. 143–151, 1978. DOI 10.5771/0943-7444-1978-3-142.

DAHLBERG, Ingetraut. Knowledge organization and terminology: philosophical and linguistic bases. **International Classification**, v. 19, n. 2, p. 65–71, 1992. DOI 10.5771/0943-7444-1992-2-65.

DAHLBERG, Ingetraut. Concepts and terms: ISKO's major challenge. **Knowledge Organization**, vol. 36, no. 2/3, p. 169–177, 2009. DOI 10.5771/0943-7444-2009-2-3-169.

DAVIS, Wayne A. **Meaning, expression, and thought**. Cambridge: Cambridge University, 2003. p. 293–550.

FREGE, Gottlob. Sense and Reference. **The Philosophical Review**, v. 57, n. 3, p. 209, May 1948.

FREGE, Gottlob. On Concept and Object. trans. by P. T. Geach; Max Black. **Mind**, vol. 60, no. 238, p. 168–180, 1951. DOI 10/fgh5zn.

FREGE, Gottlob. Function and Concept. *In*: MELLOR, D. H.; OLIVER, Alex (eds.). **Properties**. Oxford readings in philosophy. Oxford: Oxford University, 1997. p. 130–149.

FREGE, Gottlob. Digressões sobre o Sentido e a Referência. **Lógica e Filosofia da Linguagem**. Trad. Paulo Alcoforado. 2a ed. São Paulo: Cultrix, 2009. p. 159–170.

FRICKÉ, Martin. **Logic and the Organization of Information**. New York: Springer, 2012.

FRIEDMAN, Alon; SMIRAGLIA, Richard P. Nodes and arcs: concept map, semiotics, and knowledge organization. **Journal of Documentation**, v. 69, n. 1, p. 27–48, 2013. DOI 10/f4psd3.

GOGUEN, Joseph A. Concept representation in natural and artificial languages: Axioms, extensions and applications for fuzzy sets. **International Journal of Man-Machine Studies**, v. 6, n. 5, p. 513–561, Sep. 1974. DOI 10/fr78sg.

HARPRING, Patricia. What Are Controlled Vocabularies? **Introduction to controlled vocabularies: Terminology for art, architecture, and other cultural works**. Los Angeles: Getty Research Institute, 2010. p. 12–26.

HJØRLAND, Birger. Concept theory. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1519–1536, 2009. DOI 10.1002/asi.21082.

KLEMENT, Kevin C. Frege, Gottlob. (James Fieser & Bradley Dowden, eds.) *In*: Internet Encyclopedia of Philosophy. [S. l.: s. n.], c2021. Disponível em: <https://iep.utm.edu/frege/>. Acesso em: 26 mar. 2021.

LAURENCE, Stephen; MARGOLIS, Eric. Concepts and Cognitive Science. *In*: MARGOLIS, Eric; LAURENCE, Stephen (eds.). **Concepts: Core readings**. Cambridge: MIT Press, 1999. p. 3–81.

LOWE, E. Jonathan. **The four-category ontology: a metaphysical foundation for natural science**. Oxford: Clarendon Press, 2006.

MACHADO, Luís Miguel Oliveira; MARTÍNEZ ÁVILA, Daniel; SIMÕES, Maria da Graça. Concept theory in library and information science: an epistemological analysis. **Journal of Documentation**, v. 75, n. 4, p. 876–891, 2019. DOI 10.1108/JD-11-2018-0195.

MACULAN, Benildes Coura Moreira dos Santos; LIMA, Gercina Ângela Borém de Oliveira. Buscando uma definição para o conceito de “conceito”. **Perspectivas em Ciência da Informação**, v. 22, n. 2, p. 54–87, 2017. DOI 10/gmvf65.

MARGOLIS, Eric; LAURENCE, Stephen. The Ontology of Concepts: Abstract Objects or Mental Representations? **Noûs**, v. 41, n. 4, p. 561–593, 2007. DOI 10/c22c65.

MARGOLIS, Eric; LAURENCE, Stephen. Concepts. (Edward N. Zalta, ed.) *In*: The Stanford Encyclopedia of Philosophy. Summer 2019 Stanford: Stanford University, 2019. Disponível em: <https://plato.stanford.edu/entries/concepts/>. Acesso em: 18 nov. 2021.

MARRADI, Alberto. The Concept of Concept: Concepts and terms. **Knowledge Organization**, v. 39, n. 1, p. 29–55, 2012. DOI 10.5771/0943-7444-2012-1-29.

PEPONAKIS, Manolis; MASTORA, Anna; KAPIDAKIS, Sarantos; DOERR, Martin. Expressiveness and machine processability of Knowledge Organization Systems (KOS): an analysis of concepts and relations. **International Journal on Digital Libraries**, v. 20, n. 4, p. 433–452, 2019. DOI 10/gmvf63.

POPPER, Karl R. **Objective knowledge**: an evolutionary approach. 1979. 8th Rev. Oxford: Clarendon Press, 1994.

SCHWEIZER, Paul. Blind grasping and Fregean senses. **Philosophical Studies**, v. 62, n. 3, p. 263–287, 1991. DOI 10/fnfqg6.

SMIRAGLIA, Richard P.; VAN DEN HEUVEL, Charles. Classifications and concepts: towards an elementary theory of knowledge interaction. **Journal of Documentation**, v. 69, n. 3, p. 360–383, 2013. DOI 10.1108/JD-07-2012-0092.

SMITH, Barry. New Desiderata for Biomedical Terminologies. *In*: MUNN, Katherine; SMITH, Barry (eds.). **Applied Ontology**: an Introduction. Heusenstamm: Ontos Verlag, 2008. p. 83–107.

THORNLEY, Clare; GIBB, Forbes. Meaning in philosophy and meaning in information retrieval (IR). **Journal of Documentation**, v. 65, n. 1, p. 133–150, 2009. DOI 10/dvfcq6.

VAN DEN HEUVEL, Charles; SMIRAGLIA, Richard P. Concepts as particles: Metaphors for the universe of knowledge. *In*: GNOLI, Claudio; MAZZOCCHI,

Fluvio (eds.). **Paradigms and conceptual systems in knowledge organization**: 11th ISKO conference proceedings, 23-26 February 2010, Rome. Würzburg: Ergon, 2010. p. 50–56.

WEITZ, Morris. **Theories of Concepts**: a history of the major philosophical tradition. London: Routledge, 1988.

ZINS, Chaim. Conceptual Approaches for Defining Data, Information, and Knowledge. **Journal of the American Society for Information Science and Technology**, v. 58, n. 4, p. 479–493, 2007. DOI <https://doi.org/10.1002/asi.20508>.



# 4

Mario Barité

Las garantías más allá  
del horizonte de la organización  
del conocimiento:  
proyecciones posibles

*Warrants beyond the horizon  
of knowledge organization:  
possible projections*

DOI: 10.31560/pimentacultural/2022.95613.4

### Resumen:

La organización del conocimiento (OC) guarda una referencia temática estable y directa con la bibliotecología y la ciencia de la información (BCI) y se inserta en su tradición científica y de ejercicio profesional. También se reconoce al mismo tiempo, que la OC es un dominio de composición interdisciplinaria que ha evolucionado por el aporte de áreas como lingüística, terminología, recuperación de la información y ciencias cognitivas entre otras. No se ha estudiado con igual profundidad la batería de aportes que la OC puede ofrecer a otros dominios. A partir de una revisión de la literatura pertinente y una síntesis crítica de la misma, en este trabajo se actualiza la descripción del área de garantías (literaria, de usuario, cultural y otras) como formas de validación de terminología. Se analiza el potencial de utilización en ámbitos externos a la OC, tomando como estudio preliminar de caso a la lexicografía. Entre las conclusiones destaca la necesidad de promover el conocimiento acumulado sobre las garantías en publicaciones y congresos de áreas que podrían estar interesadas en integrarlo, tales como lexicografía, terminología e ingeniería del lenguaje.

**Palavras-Chave:** Garantías. Organización del conocimiento. Terminología. Lexicografía.

**Abstract:**

*Knowledge organization (KO) has a direct and stable thematic relationship with library and information science (LIS). Therefore, it is inserted in their scientific and professional practice tradition. At the same time, it is also recognized that KO is an interdisciplinary domain that has evolved through the contribution of linguistics, terminology, information retrieval, and cognitive sciences, among others. The contributions that KO can offer to other domains have not been studied to the same extent. Based on a review of the relevant literature and a critical synthesis, this paper updates the description of warrants (literary, user, cultural, and others) as forms of terminology validation. The potential of use in areas beyond KO is analyzed, taking as a preliminary case study the field of lexicography. Among the conclusions, it is highlighted that it is necessary to promote the accumulated knowledge on warrants collected in publications and congresses of areas interested in integrating warrants, for example, lexicography, terminology, and language engineering.*

**Keywords:** Warrants; Knowledge organization, Terminology; Lexicography.

## 1 INTRODUCCIÓN

La organización del conocimiento (OC) mantiene una conexión temática estable y directa con la Ciencia de la información (CI) y se inserta en su tradición científica y de ejercicio profesional. Es lo que establece Smiraglia cuando afirma que la OC *“is critical for the proper functioning of the science of information”, debido a que “without that which is learned in knowledge organization, information retrieval cannot work”* (SMIRAGLIA, 2014, p. 3). En la misma línea Hjørland señala que la OC es la disciplina central de la CI (HJØRLAND, 2008, p. 86).

Esa centralidad se caracteriza, en palabras de Lara, por “la noción de información y sus relaciones con el usuario” (LARA, 2007, p. 104), lo que se expresa en aspectos concretos y significativos del procesamiento de información que están a cargo de la OC: la organización, el proceso y la recuperación de documentos y de información a través – he aquí el elemento principal- de llaves temáticas. Dicho de otra manera, lo que importa en nuestro campo es la utilidad que puede brindar para distintos propósitos, la agrupación de documentos y referencias por temas, así como el modo en que las búsquedas por temas pueden resolver distintos problemas de información de las personas.

Se reconoce asimismo que la OC es un dominio de composición interdisciplinaria que ha evolucionado con el aporte de diversas áreas que la fundamentan y la enriquecen: *“mathematics, systems theory, psychology, science in general, semantics, grammar problems, online retrieval systems and technologies, terminology, translation”* (DAHLBERG, 2006, p. 15), así como *“computer science, linguistics, natural language processing, theory of knowledge, theory of social organization (...) cognition and communication”* (HJØRLAND, 2008, p. 98). Esta diversidad de fluencias se manifiesta en la terminología del área. Por ese motivo es frecuente encontrar en la literatura específica de OC términos de informática (base de datos, banco de datos), lingüística (sintaxis, lenguaje,

vocabulário), recuperación de información (acceso, frecuencia, ocurrencia, revocación), clasificación de las ciencias (género, especie, árbol de Porfirio, clasificación dicotómica), filosofía clásica (categoría, accidente, materia, propiedades) y matemáticas (algoritmo, álgebra booleana).

Sin embargo, no se ha estudiado con igual profundidad la batería de contribuciones que la OC puede ofrecer a otras áreas del conocimiento desde perspectivas teóricas, metodológicas, o aplicativas. El objetivo de este trabajo es explorar los aportes que en concreto se pueden hacer desde una conceptualización propia de la OC, nacida y desarrollada en su seno, y que parece haber alcanzado un punto de madurez suficiente para extender su potencial de aplicación a otros dominios del saber: la noción de garantía. Para ello se realiza en primer lugar, una visita de su concepto. En segundo término se establece el estado de situación de las diferentes garantías en la OC. A continuación se analiza el potencial de aplicación en el dominio de la lexicografía, aprovechando su caudal teórico y práctico. Finalmente, por tratarse de un trabajo exploratorio, se establecen conclusiones preliminares.

## 2 NOCIÓN DE GARANTÍA

Hulme (1911) inauguró la historia del concepto, al acuñar la expresión 'garantía literaria' (*literary warrant*) por la cual estableció que solo un volumen de libros (o por extensión, cualquier otro tipo de documentos) que trate específicamente sobre un tópico determinado, puede justificar la creación de un símbolo representativo de su contenido para fines de clasificación y recuperación en bibliotecas. En su expresión actual, este símbolo puede tratarse de un código numérico, alfabético o mixto propio de un sistema de clasificación, una expresión del lenguaje natural (palabra clave, *tag*) o un término normalizado (descriptor, encabezamiento de materia, etiqueta) que se integra a un

sistema de organización del conocimiento. En ese documento iniciático, Hulme introduce la idea de garantía como “*test of validity*” (HULME, 1911, p. 448) para la inclusión o exclusión de tópicos (que él llamaba clases) en un sistema de clasificación.

Hay que esperar hasta 1986 para que Beghtol, hablando en su caso de “*semantic validity*”, realizara un estudio pormenorizado de antecedentes de cuatro garantías (literaria, científico-filosófica, educativa y cultural), al tiempo de realizar algunas apreciaciones sobre la que mencionó como garantía académica. En ese trabajo, establece la siguiente definición que expande y profundiza los elementos que caracterizan a una garantía:

The warrant of a classification system can be thought of as the authority a classificationist invokes first to justify and subsequently to verify decisions about what classes/concepts to include in the system, in what order classes/concepts should appear in the schedules, what unit classes/concepts are divided into, how far subdivision should proceed, how much and where synthesis is available, whether citation orders are static or variable and similar questions (BEGHTOL, 1986, p. 110).

Desde este enfoque, se convalida la idea de que las garantías son una expresión de autoridad para justificar la toma de decisiones, y además se extiende su justificación no solo a los términos candidatos a integrar un sistema de clasificación, sino además a las relaciones que se establecen entre ellos, los criterios de ordenación de términos, la manera en que se subdividen y la profundidad de estas subdivisiones. De este modo, cada término, cada relación y en suma todo el sistema nocional que sostiene la estructura de un sistema de organización del conocimiento, se organiza en función de la utilización consistente de garantías.

Los autores que han avanzado en diversas direcciones en el estudio de las garantías, encontraron en esta definición un acuerdo básico, un punto de partida común para tomar decisiones de índole teórica y metodológica (COCHRANE, 1993; DUFF, 1998; BEGHTOL, 2002a,

2202b, CAMPBELL, 2008; BARITÉ, 2011, 2018, 2019; BULLARD, 2017; LEE, 2017). Por ese motivo, Bullard afirma con convicción que “*warrant is a common thread across a wide variety of systems ranging from traditional library classification to in-application menus and categories for web-based collections*” porque “*all designers of textual organizing schemas must look to some source for terminology*” (BULLARD, 2017, p. 76).

Una garantía afianza en primer lugar, la pertinencia y la adecuación de un término para representar un concepto, de modo tal que ese término pueda ser utilizado para la clasificación o para la indización de datos, documentos o información. Dicho así, parecería que la cuestión de las garantías se dirime término a término. Sin embargo, conviene que prevalezca una visión de conjunto, porque el diseño de un sistema de organización del conocimiento es, antes que nada, un proyecto intelectual. Esto quiere decir que toda garantía debe ofrecer terminología que favorezca una representación temática consistente y actualizada del contenido de los documentos, así como debe presentar un mapa integral de un dominio, que contenga elementos de su desarrollo, su evolución y su proyección.

Para fundar mejor la necesidad de una perspectiva epistemológica integrada, Olson (2004) rastreó el fondo de la noción de garantía en la obra clasificatoria del conocimiento de Francis Bacon, y sus reflejos en los sistemas bibliotecológicos de clasificación. En ese trabajo señala que Bacon creó una estructura de conocimiento basada en una episteme común, sobre la que los clasificacionistas han ido montando sus propios diseños de sistemas desde el siglo diecinueve hasta el presente. Y concluye: “*perhaps we should also follow his epistemological warrant and let our classifications not only reflect knowledge, but also have a role in directing the creation of new knowledge*” (OLSON, 2004, p. 4).

De algún modo, los clasificacionistas alimentaron, a lo largo de un siglo y medio, una visión epistemológica presuntamente común, bajo la idea de que los preceptos de las ciencias son objetivos, neutrales y

de alcance universal. De forma más soterrada quizás también fue instalándose el pensamiento de que la justificación de la terminología de un sistema de organización del conocimiento debería tomarse en forma íntegra, ya del lenguaje formal de las ciencias y de la educación de las ciencias (BLISS, 1929), ya de las evidencias terminológicas que ofrece la documentación (HULME, 1911). Ese esquema tradicional, basado en la pretendida –o al menos, esperable- objetividad y neutralidad del lenguaje de las ciencias y la tecnología se fue quebrando poco a poco, al proponerse nuevas garantías basadas en los valores y las creencias culturales de una comunidad (garantía cultural), en el lenguaje elegido por los usuarios de un sistema de información para buscar información (garantía de uso y garantía de usuario), en la terminología interna de una organización (garantía organizacional), o incluso en las denominaciones y los alcances variables que un mismo concepto (por ejemplo, uno del área de la geología) recibe en diferentes regiones, en base a, siguiendo el ejemplo de la geología, las características peculiares de los suelos, o las particularidades lingüísticas de cada lugar.

En cualquier caso, no se refuta la idea de que las garantías constituyen criterios necesarios de selección de términos para su inclusión en los sistemas de organización del conocimiento, con el objetivo principal de favorecer las búsquedas y las recuperaciones basadas en tópicos o asuntos.

### 3 ESTADO DE SITUACIÓN DE LAS GARANTÍAS EN OC

La elección de una garantía adecuada a las finalidades que se persiguen con el diseño de un sistema de organización del conocimiento, debería expresarse en un mayor rigor metodológico en la elección, depuración y sistematización de la terminología. También debería

proveer términos representativos del estado del arte de un dominio, en sus consensos, en sus disensos, y en la presentación equilibrada de diferentes corrientes o teorías. Debería por tanto, evitar la selección de términos ambiguos, demasiado genéricos, demasiado específicos, u obsoletos (salvo que se pretenda atender colecciones documentales de espectro diacrónico amplio).

Sin embargo, la relación directa entre la aplicación rigurosa de una garantía y la calidad del producto final parece ponerse en entredicho por el hecho de que se han propuesto una veintena de garantías en la literatura, más de la mitad de ellas en los últimos veinticinco años. Ello estaría sugiriendo cierto grado de insatisfacción en los equipos de clasificacionistas, con relación a las soluciones aportadas por las garantías tanto clásicas como recientes. Las causas de esa insatisfacción pueden ser, en principio, las siguientes:

- la insuficiencia de cada garantía en particular, para decidir en todas y cada una de las instancias en las que se requiere seleccionar terminología para representar conocimiento y recuperar información;
- la carencia de soluciones universales;
- la comprobación de que la terminología de un dominio cualquiera puede ser organizada de diferentes maneras, todas ellas válidas para determinados usuarios o contextos, conforme a diversas perspectivas epistemológicas o prácticas.

A continuación se presenta un cuadro actualizado con las garantías identificadas en la literatura de OC, y que cuentan con una base mínima de documentos que la referencian. El cuadro presenta a las garantías ordenadas en orden cronológico de aparición, comenzando por las más antiguas. Se ofrecen los nombres de las garantías en inglés y en español, los autores que las propusieron y el año de acuñación en cada caso. Se incluye además una columna de comentarios.

Cuadro 1 – Tipo de garantías

Nombre en inglés	Nombre en español	Autor y año	Comentario
Usage /Use	Uso	Cutter, 1876	Antecedente de garantía de uso
Literary warrant	Garantía literaria	Hulme, 1911	
Scientific /philosophical and educational warrant (consensus)	Garantía científico / filosófica y educativa, y consenso	Bliss, 1929	Antecedente de garantía académica
Cultural warrant	Garantía cultural	Lee, 1976	
User warrant	Garantía de usuario	Lancaster, 1977	
Logical warrant	Garantía lógica	Fraser, 1978	
Concrete institutional warrant	Garantía institucional concreta	Coates, 1978	
Enquiry warrant	Garantía de consulta	CRG, 1984, citado por Beghtol 1986	
Semantic warrant	Garantía semántica	Beghtol, 1986	Nombre genérico que comprende las garantías literaria, cultural, científico / filosófica, educativa y cultural
Education/academic warrant	Garantía educativa o académica	Beghtol, 1986	
Gender warrant	Garantía de género	Olson and Ward, 1998	Quizás un tipo de garantía cultural
Phenomenological warrant	Garantía fenomenológica	Ward, 2000	
Structural warrant	Garantía estructural	Svenonius, 2000	
Use warrant	Garantía de uso	Svenonius, 2000	
Ethical warrant	Garantía ética	Beghtol, 2002b	Relacionada con la garantía cultural
Academic warrant (also named scholarly warrant)	Garantía académica	Sachs and Smiraglia, 2004	Similar a la garantía científico /filosofía y relacionada con la educativa
Organizational warrant	Garantía organizacional	National Information Standards Organization, 2005	

Autopoietic warrant	Garantía autopoietica	Mai, 2011	Basada en Rafferty and Hilderley, 2007. Quizás un tipo de garantía de usuario
Textual warrant	Garantía textual	Tennis, Thornton and Filer, 2012	
Market warrant	Garantía de mercado	Martínez Ávila, 2012	
Indigenous warrant	Garantía indigenista	Doyle, 2013	Un tipo de garantía cultural
Genre warrant	Garantía de géneros	Andersen, 2015	Un tipo de garantía cultural
Epistemic warrant	Garantía epistémica	Budd and Martínez Ávila, 2016	
Policy warrant (corresponding to policy based indexing)	Garantía en políticas	Hjørland 2017	¿Un tipo de garantía cultural?

Fuente: Barité (2018), modificada con datos de Martínez-Ávila y Budd (2017)

No todos los autores utilizaron explícitamente la palabra 'garantía' aunque quedó claro en los textos respectivos que, en sustancia, se estaba hablando de la autoridad invocada para representar conocimiento a través de descriptores, epígrafes, palabras clave, números de clasificación, taxones u otros símbolos. Las garantías se distinguen, como es de presumir, por la apelación a distintas fuentes de autoridad para recoger términos.

La literatura generada por estas garantías es bastante desigual. La garantía literaria es la que más atención ha recibido, por ser la más antigua y consecuentemente, establecer un marco de referencia. Cuenta con un cuerpo más o menos consolidado de conocimiento acumulado en torno a la garantía literaria: artículos canónicos (RODRÍGUEZ, 1984; BEGHTOL, 1986), estudios abarcativos y tesis de posgrado (BARITÉ, 2011; 2018), producción regular en los últimos años (HOWARTH; JANSEN, 2014; BULLARD, 2017), así como un análisis recurrente de la aplicación de la garantía literaria a la Clasificación de la Biblioteca del

Congreso (por ejemplo, HALLOWS, 2015) o a la Clasificación Decimal Dewey (por ejemplo, VIZINE-GOETZ and BEALL, 2004).

Un segundo grupo se forma con el volumen de documentos que se ha ido generando en torno a la garantía cultural (LEE, 1976; BEGHTOL, 2002a), la garantía de usuario (LANCASTER, 1977; HJØRLAND, 2013); la garantía académica o científica (BLISS, 1929; SACHS; SMIRAGLIA, 2004) y la garantía organizacional (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). A un segundo grupo de garantías se les ha dedicado un conjunto limitado de trabajos, en los cuales se encuentran habitualmente introducciones genéricas y aproximaciones generales a las características, los propósitos y las utilidades que puede ofrecer cada garantía. Es el caso de la garantía cultural, la garantía de usuario, la garantía organizacional, la garantía académica, y más recientemente la garantía indígena o indigenista. Un tercer grupo reúne a las garantías que han sido propuestas y/o mencionadas puntualmente, y que no tienen un desarrollo ulterior significativo, como en la garantía de mercado, la estructural o la autopoietica. Estos dos últimos grupos de garantías no cuentan todavía con suficiente análisis crítico, y necesitan del mismo.

El resto de las garantías han sido solo esporádicamente tratadas, en general por sus autores, y no cuentan todavía con una expresión teórica, metodológica y crítica de envergadura. No obstante, se entiende que son opciones abiertas, con una base asegurada de fiabilidad, que pueden ser aplicadas en contextos particulares con éxito. Por ejemplo, la garantía institucional y la garantía organizacional son dos formas de justificación que toman como sustrato la existencia de formas de expresión propias de instituciones y empresas. En algunos casos, solo son útiles para la comunicación interna entre funcionarios; en otros, permiten establecer índices de productos o de servicios; y aún en otros es posible que afiancen el conocimiento social o en el mercado, de determinados conceptos que son impulsados desde la misión y la visión particular de cada organismo, sea público o privado.

Se ha discutido si las garantías pueden utilizarse en forma combinada, o si es preferible mantener una garantía única para toda la terminología seleccionada en el proceso de construcción de un sistema de organización del conocimiento. Svanberg establece que *“the diverging warrants may be used to complement each other. Besides complementing each other, warrants may be contrary to each other”* (SVANBERG, 1996). Aitchison, Gilchrist and Bawden (2000) aportan una visión inclusiva, cuando dicen la justificación de terminología puede provenir tanto de la selección de fuentes de referencia en la literatura corriente como del análisis de las búsquedas en el sistema de información, y el conocimiento aportado por usuarios e indizadores. En la misma dirección, Tennis establece que hoy día *“warrant is based on literature, users, scholarly opinion and is culturally biased”* (TENNIS, 2005, p. 85).

Para Huvila las garantías funcionan como herramientas analíticas y como un soporte para el diseño, incluso cuando se aplican dos o más que pueden dar resultados opuestos (HUVILA, 2006, p.78). En cualquier caso, no parece haber razón para excluir la combinación entre garantías, siempre y cuando exista la necesidad. De hecho, explica Bullard *“the various warrants available to classification designers represent contradictory positions in classification theory yet they compete and are combined by classification designers in daily practice”* (BULLARD, 2017, p. 77). Finalmente, Lee genera ideas para entender la naturaleza del conflicto entre varias garantías semánticas, y ofrece algunas alternativas de negociación de su uso y combinación, en el marco de los procesos de evaluación de sistemas de organización del conocimiento (LEE, 2017).

## 4 APLICACIÓN DE GARANTÍAS FUERA DE LA OC

En este apartado se analiza el potencial de utilización de las garantías en ámbitos externos a la OC, con la finalidad de explorar su posible extrapolación, desde perspectivas teóricas y metodológicas.

No se trata de que la OC tenga algo que enseñar a otras disciplinas, sino de explorar con especialistas de esas disciplinas, las posibilidades de intersección y utilización de ideas, técnicas y procedimientos, con la perspectiva de generar espacios de trabajo interdisciplinario, en los que el sentido del tráfico sea bidireccional.

En principio, la exploración se concentra en posibles puntos de encuentro con la lexicografía, dedicada a la producción de diccionarios generales de las lenguas y a repertorios similares.

Se reconoce en la actualidad que la lexicografía (generalmente considerada una subdisciplina de la lingüística), tiene al menos dos vertientes: una práctica, dedicada a compilar, escribir o editar diccionarios; otra teórica o metalexigráfica, dirigida al estudio de los fundamentos del análisis y la descripción del vocabulario de una lengua particular, así como de la estructura del sistema semántico de relaciones que se producen entre las palabras diccionarizadas, y el proceso de evaluación de diccionarios (ANGLADA ARBOIX, 1991; BERGENHOLTZ y GOUWS, 2012). Algunos autores dudan de la claridad de esta división de la disciplina, y trasladan la duda a la definición misma del campo, porque *“the exact scientific meaning of the word lexicography is continuously changing and can only be determined on the basis of an up-to-date knowledge of the theoretical and practical development of this branch of human activity”* (TARP, 2010, 454).

Una cuestión crucial a la que se enfrentan los lexicógrafos de todas las lenguas es ¿qué palabras se deben elegir para incluir en un diccionario de la lengua, de qué manera y con cuáles argumentos?

La rica tradición lexicográfica existente en la producción de diccionarios de la lengua castellana o de sus regionalismos, lo que se expresa principalmente (pero no exclusivamente) en la producción de un diccionario 'oficial' de la lengua, publicado en sucesivas ediciones a lo largo de más de trescientos años, ha permitido consolidar los criterios de

selección de sus lemas. También ha permitido la incorporación –restringida pero continua- de voces de especialidad, procedentes de distintas áreas del saber. Por ese motivo, la terminología, con una historia mucho más reciente desde su constitución formal como dominio a principios del siglo veinte, ha tomado enseñanza de los métodos de la lexicografía para la elaboración de diccionarios y glosarios especializados.

Un criterio histórico para fundamentar la inclusión de una palabra o una voz especializada en un diccionario de la lengua general, es la documentación, entendida como el rastreo de voces y expresiones en un conjunto predefinido de fuentes (llamado *corpora*). Ese conjunto puede formarse con obras literarias, filosóficas, históricas o científicas.

La documentación ofrece la seguridad de que la palabra existe o ha sido usada por un autor reconocido. Por otra parte, la voz puede ser datada considerando la fecha de publicación de las obras en las que comparece. La desventaja más notoria de este método es que limita las posibilidades de que se haga visible el lenguaje popular, coloquial o festivo, y puede concentrar en forma excesiva registros cultos o poco usuales. En su forma clásica, el argumento principal a favor de la documentación radicaba en el respaldo de las opiniones autorizadas, que por este motivo pasaban a llamarse autoridades. Por esa razón, no extraña que la primera edición del diccionario de la Real Academia Española, publicado entre 1726 y 1739, se llame Diccionario de Autoridades. La *corpora* "constituía para los académicos un repertorio de escritores y obras que se consideraban modélicos y que, por lo tanto debían ser el fundamento del repertorio lexicográfico de la academia" (FREIXAS ALÁS, 2003, p. 177).

Con el paso del tiempo, otras opiniones se han hecho escuchar. Al discutir los criterios para incorporar neologismos a los diccionarios, Bernal, Freixa y Torner examinan con cierto detalle las siguientes cuatro alternativas de justificación: de frecuencia, formales, semánticas y lexicográficas, conforme a la síntesis que se realiza a continuación.

1) Criterio de frecuencia, que se subdivide a su vez en cinco:

1.1 Frecuencia absoluta de uso. “A mayor frecuencia de un neologismo, mayores son sus posibilidades de estabilización en la lengua y, por consiguiente, mayor es su grado de diccionabilidad” (BERNAL, FREIXA y TORNER, 2020, p. 595).

1.2. Extensión de uso. “Los neologismos son más diccionabilizables si se emplean de forma igualmente frecuente en textos de diferente tipología” (BERNAL, FREIXA y TORNER, 2020, p. 596).

1.3. Estabilidad en el uso. El uso frecuente sostenido en el tiempo constituye un índice de estabilización que actúa a favor de la diccionabilidad (BERNAL, FREIXA y TORNER, 2020, p. 596).

1.4. Cambio longitudinal en la frecuencia. Si el uso de una palabra crece con el paso del tiempo no solo muestra una tendencia a su estabilización; en muchos casos pasa “de tener un carácter meramente testimonial a una elevada frecuencia de uso. El cambio longitudinal en la frecuencia indica que un vocablo existente ha pasado a ser un vocablo necesario” (BERNAL, FREIXA y TORNER, 2020, p. 596).

1.5 Extensión (geográfica) en el uso. “Los neologismos que se emplean en dialectos particulares difícilmente pasan a formar parte de la lengua general, por lo que se ha considerado que los utilizados en distintas variedades geolectales son mejores candidatos a ser recogidos en diccionarios de lengua”. (BERNAL, FREIXA y TORNER, 2020, p. 596).

2) Criterios formales. “A mayor conformidad con las reglas de formación de la lengua, mayores las posibilidades de que la nueva unidad léxica se consolide en el uso y, por lo tanto, deba ser repertoriada en las obras lexicográficas” (BERNAL, FREIXA y TORNER, 2020, p.597). Los criterios formales permiten establecer los alcances de los procesos derivativos, que producen familias de palabras o términos, a partir de una raíz común.

3) Criterios semánticos. Bernal, Freixa y Torner (2020) mencionan dos criterios de este orden, la impredictibilidad semántica y la necesidad denominativa. Explican que cuanto menos predecible sea el significado de una unidad léxica mayor será su grado de dictionariabilidad. Lo mismo acontece, según estas autoras, con derivados que nacen de una base que también es neológica. En cuanto a la necesidad denominativa plantea las dificultades de implantación de neologismos cuando ya existe una alternativa en la lengua, lo que puede provocar la coexistencia de sinónimos.

4) Criterios documentales. En este caso, Bernal, Freixa y Torner (2020) abogan por un tipo específico de documentación, aquel que se apoya en la inclusión previa de una palabra o un término en otros diccionarios, preferentemente de prestigio, para establecer la posible dictionariabilidad de la expresión candidata. Todos los criterios mencionados no son de expresión lineal; por lo contrario, admiten en la lexicografía matices, excepciones o contextos que pueden favorecer o desaconsejar algunos de ellos.

A partir de la enumeración realizada, puede establecerse una relación bastante cercana entre los criterios lexicográficos de selección de lemas y aquellos proporcionados por las garantías en los procesos de decisión sobre términos a incorporar a los sistemas de organización del conocimiento. Así se expresa en el cuadro siguiente.

**Cuadro 2 – Correlación criterios lexicográficos - garantías**

<b>Criterios lexicográficos</b>	<b>Tipos de garantía</b>
Documentación	Garantía literaria
Frecuencia	Garantía de uso, garantía de usuario
Criterios formales	¿Garantía formal?
Criterios semánticos	Garantías semánticas
Criterios documentales	Garantía literaria

Fuente: El autor

La documentación y la garantía literaria suelen ser los dos criterios tradicionales utilizados en la lexicografía y la organización del conocimiento respectivamente. El fundamento último se encuentra en el hecho de que toda expresión que representa a un concepto que es reconocido como valioso o necesario, en virtud de sus evidencias o de la autoridad personal que los respalda, se documenta en alguna parte (libros, artículos, ponencias a congresos, monografías y otras expresiones, principalmente textuales).

Más recientemente, los criterios vinculados al uso han comenzado a prevalecer y se han valorizado a partir de las posibilidades que brinda la tecnología y el desarrollo de algoritmos, en el manejo y la gestión de grandes cantidades de datos.

Los criterios formales procedentes de la lexicografía constituyen un punto de apoyo importante en la organización del conocimiento, en lo relativo a la configuración gráfica de los términos, aunque esa vertiente se canaliza a través de los estándares de construcción de vocabularios controlados (por ejemplo, NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). De todos modos, no sería descabellado aventurar la posibilidad de desarrollar una garantía formal, que permita fundamentar la inclusión de series derivativas de términos en sistemas de organización del conocimiento.

Cabe considerar que al menos algunas de las garantías no mencionadas en el cuadro 2, podrían utilizarse también por parte de los lexicógrafos en circunstancias puntuales. Es el caso de la garantía ética o la garantía cultural, que puede orientar la selección de los lemas más apropiados en determinadas circunstancias. Esto implicaría tomar partido por expresiones políticamente convenientes por sobre mecanismos meramente descriptivos de la lengua, lo que puede abrir un frente amplio para el debate.

Lo que puede extraerse de esta primera aproximación al tema estudiado es que el conocimiento acumulado sobre garantías podría ser aprovechado por la lexicografía, para favorecer diversas instancias de justificación en el proceso de diccionarización de lemas. Por su parte, la organización del conocimiento podría aprovecharse de los criterios lexicográficos que traen tras de sí una tradición de varios siglos de desarrollo, para complementar, fortalecer o incluso poner en cuestión las formas de aplicación de alguna de sus garantías.

## 5 CONCLUSIONES

Las garantías son vistas en la actualidad como un componente ineludible en el proceso de construcción, evaluación y revisión de sistemas de organización del conocimiento, en la medida en que su adecuada comprensión y aplicación deberían asegurar una terminología consistente, actualizada y ajustada a los propósitos de los diseñadores de sistemas, y a las necesidades de información temática de los usuarios.

Luego de un siglo y medio de evolución como concepto, como herramienta analítica y como soporte para el diseño de sistemas, su eficacia y su utilidad parecen no necesitar ya el respaldo de nuevas argumentaciones en OC. Sí quedan aún muchas puertas abiertas para la investigación, sobre todo en materia metodológica. En efecto, ¿cómo orientar la elección de una garantía y no de otra?; ¿cómo decidir cuál es la garantía más apropiada a cada sistema?; ¿cómo se evalúa el 'rendimiento' de una garantía en función de principios como los de consistencia, actualidad, exhaustividad, adecuación temática, adecuación lingüística y otros indicadores que pueden proponerse?; ¿qué beneficios se pueden obtener de los avances tecnológicos para fortalecer la calidad de aplicación de las garantías?; o ¿cómo se vinculan las garantías con la indización por lenguaje natural? Todas ellas son preguntas

que no tienen todavía respuestas totalmente satisfactorias. Se trata de espacios ciegos que pueden aprovecharse para la investigación, en un ámbito en el cual han sido constantes tanto la aparición de nuevos tipos de sistemas de organización del conocimiento (ontologías, taxonomías web, folksonomías y clasificaciones sociales entre otros), como de innovaciones tecnológicas que han impactado en las costumbres y en los hábitos en relación con la búsqueda y el acceso a la información.

No obstante esas cuestiones pendientes, las teorías y las metodologías garantistas exhiben ya un estado de desarrollo que habilita a favorecer espacios de diálogo con otros dominios del conocimiento, con vistas a la articulación con teorías y métodos similares o afines. En este trabajo en particular se exploraron las posibilidades que se ofrecen a la lexicografía, en lo relativo a la fundamentación de las palabras y los términos candidatos a transformarse en lemas o voces especializadas de referencia en diccionarios.

Hay otras áreas a las que se puede extender el uso de garantías, pues reforzarían la utilización de criterios lingüísticos y/o terminológicos consistentes, y la presentación homogénea de datos. Entre ellas está la asignación de etiquetas temáticas en sitios y portales web, o la elaboración de índices temáticos. En internet se multiplican espacios dedicados a culturas de grupos minoritarios, terminologías alternativas propuestas por pensadores o movimientos sociales -en especial los contraculturales-, o incluso por los proveedores de nuevas tecnologías. En contraposición suelen faltar glosarios ordenados y actualizados con voces nuevas, que estén debidamente respaldadas por renovadas figuras de autoridad.

Quizás lo que más haya que valorar es que el tópico de las garantías mantiene vigencia debido a la importancia que se asigna hoy a la recuperación temática en Internet, en bases y bancos de datos y en otras fuentes y sistemas de información, vinculados con la ciencia, el comercio, el gobierno electrónico, la cultura y las industrias

del entretenimiento. Su proyección y utilidad tienen, por consiguiente, un amplio horizonte para afianzarse.

Por lo expuesto, se destaca la necesidad de promover la difusión del conocimiento acumulado sobre las garantías dentro de la OC, en publicaciones y congresos de áreas tales como la lexicografía, la terminología y las diversas ingenierías del lenguaje, en procesos comunes de exploración.

## 6 REFERENCIAS

AITCHISON, Jean, GILCHRIST, Alan and BADWEN, David (2000). **Thesaurus construction and use: a practical manual**. London: ASLIB, 2000.

ANDERSEN, Jack. Re-describing knowledge organization: A genre and activity-based view. *In*: ANDERSEN, Jack. (Ed.). **Genre theory in information studies**. Bingley: Emerald, 2015. p. 13-42.

ANGLADA ARBOIX, Emilia. Lexicografía, metalexigrafía, diccionario, discurso. **Sintagma**, n.3, p. 5-11, 1991.

BARITÉ, Mario. Towards a general concept of warrants: first notes. **Knowledge Organization**, v. 46, n. 8, p. 647-655, 2019.

BARITÉ, Mario. Literary warrant. **Knowledge Organization**, v. 45, n. 6, p. 517-536, 2018.

BARITÉ, Mario. **La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento: modelo y aplicación**. Tesis de doctorado. Granada: Universidad de Granada, 2011.

BEGHTOL, Clare. Universal concepts, cultural warrant and cultural hospitality. **Advances in Knowledge Organization**, v. 8, p. 45-49, 2002a.

BEGHTOL, Clare. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation**, v. 58, n. 5, p. 507-532, 2002b

BEGHTOL, Clare. Semantic validity: concepts of warrant in bibliographic classification systems. **Library Resources & Technical Services**, v. 30, n. 2, p. 109-123, 1986.

BERGENHOLTZ, Henning; GOUWS, Rufus H. What is lexicography? **Lexikos**, n. 22: p. 31-42, 2012.

BERNAL, Elisenda; FREIXA, Judit; TORNER, Sergi. Criterios para la diccionari-  
zación de neologismos: De la teoría a la práctica. **Revista Signos: Estudios de  
Lingüística**, v. 53, n. 104, p. 592-618, 2020. Disponível em: <https://www.scielo.cl/pdf/signos/v53n104/0718-0934-signos-53-104-592.pdf>. Acesso el 15 nov. 2021.

BLISS, Henry Evelyn. **The organization of knowledge and the system of the sciences**. New York: Holt, 1929.

BUDD, John M.; MARTÍNEZ ÁVILA, Daniel. Epistemic warrant for categoriza-  
tional activities in knowledge organization. **Advances in Knowledge Organi-  
zation**, v. 15, p. 142-145, 2016.

BULLARD, Julia. Warrant as a means to study classification system design. **Journal of Documentation**, v. 73, n. 1, p. 75-90, 2017.

CAMPBELL, D. Grant. Derrida, logocentrism and the concept of warrant on the  
semantic web. **Advances in Knowledge Organization**, v. 11, p. 222-228, 2008.

COATES, Eric J. The Broad System of Ordering (BSO). *In*: TAYLOR, P.J., (Ed.),  
**New trends in documentation and information: proceedings of the 39th  
FID Congress, University of Edinburgh, 25-28 September 1978**. London:  
Aslib, 1978, p.260.

COCHRANE, Pauline. Warrant for concepts in classification schemes. **Advan-  
ces in Classification Research**, v. 4, p. 35-46, 1993.

CUTTER, Charles A. **Rules for a dictionary catalog**. Washington, D.C.:  
Government Printing Office, 1876.

DAHLBERG, Ingetraut. Knowledge organization: A new science? **Knowledge  
Organization**, v. 33, n. 1, p. 11-19, 2006.

DOYLE, Ann Mary. **Naming, claiming, and (re)creating: Indigenous know-  
ledge organization at the cultural interface: doctoral dissertation**. 2013.  
Disponível em: <https://circle.ubc.ca/handle/2429/44262>.

DUFF, Wendy M. Harnessing the power of warrant. **American Archivist**, v. 61,  
n. 1, p. 88- 105, 1998.

FRASER, Walter J. Literary, user and logical warrants as indexing constraints.  
*In*: **The Information age in perspective: proceedings of the ASIS Annual  
Meeting 1978**. White Plains, Knowledge Industry Publications, New York,  
1978, p. 130-132.

FREIXAS ALÁS, Margarita. **Las autoridades en el primer diccionario de la Real Academia Española. Tesis doctoral dirigida por José Manuel Blecua Perdices.** Bellaterra: Universitat Autònoma de Barcelona. 2003.

HALLOWS, Kristen M. It's all enumerative: reconsidering Library of Congress Classification in U.S. law libraries. **Law Library Journal**, v.106, n. 1, p. 85-99, 2015.

HJØRLAND, Birger. What is knowledge organization (KO)? **Knowledge Organization**, v. 35, n. 2/3, p. 86-101, 2008.

HJØRLAND, Birger. Theories of knowledge organization—theories of knowledge. **Knowledge Organization**, v. 40, n. 3, p. 169-181, 2013.

HJØRLAND, Birger. Subject (of documents). **Knowledge Organization**, v. 44, n.1, p. 55-64, 2017.

HOWARTH, Lynne C.; JANSEN, Eva Hourihan. Towards a typology of warrant for 21st century knowledge organization systems. **Advances in Knowledge Organization**, v. 14, p. 216-221, 2014.

HULME, Edward Wyndham. Principles of book classification: chapter III: On the definition of class headings, and the natural limit to the extension of book classification. **Library Association Record**, n. 13, p. 444-449, 1911.

HUVILA, Isto. **The Ecology of information work: a case study of bridging archaeological work and virtual reality based knowledge organisation.** Åbo: Åbo Akademi University Press, 2006. Disponible en: <https://oa.doria.fi/bitstream/handle/10024/4153/TMPobjres.83.pdf?sequence=1> Acceso el 15 nov. 2021.

LANCASTER, Frederick Wilfrid. Vocabulary control in information retrieval systems. **Advances in Librarianship**, v. 7, p. 1-40, 1977.

LARA, Marilda Lopes Ginez de. Ciencias del lenguaje, Terminología y Ciencia de la Información: relaciones interdisciplinarias y transdisciplinariedad. En: RODRÍGUEZ BLANCO, Blanca; ALVITE DÍEZ, María Luisa (eds.). **La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico = Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge: Actas del VIII Congreso ISKO-España, León, 18, 19 y 20 de Abril de 2007.** León: Universidad de León, Secretariado de Publicaciones, 2007, p. 101-110.

LEE, Joel M. E. Wyndham Hulme: A reconsideration. In: RAYWARD, W.B. (Ed.). **The variety of librarianship: essays in honour of John Wallace Metcalfe.** Sydney: LAA, 1976.

LEE, Wan Chen. **Conflicts of semantic warrants in cataloging practices.** Presented at the North American Symposium on Knowledge Organization, Urbana-Champaign, Illinois, 2017. Disponível em: [http://www.iskocus.org/NASKO2017papers/NASKO2017\\_paper\\_25.pdf](http://www.iskocus.org/NASKO2017papers/NASKO2017_paper_25.pdf). Acesso el 15 nov. 2021.

MAI, Jens-Erik. Folksonomies and the new order: authority in the digital disorder. **Knowledge Organization**, v. 38, n. 2, p. 114-122, 2011.

MARTÍNEZ-ÁVILA, Daniel. **DDC-BISAC switching as a new case of reader-interest classification: tesis de doctorado.** Getafe: Universidad Carlos III de Madrid, 2012.

MARTÍNEZ-ÁVILA, Daniel; BUDD, John M. Epistemic warrant for categorizational activities and the development of controlled vocabularies. **Journal of Documentation**, v. 73, n. 4, 2017

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies: An American National Standard Developed by the National Information Standards Organization: ANSI/NISO Z39.19-2005 (r2010).** Bethesda: NISO Press, 2005. Disponível em: [http://www.niso.org/apps/group\\_public/download.php/12591/z39-19-2005r2010.pdf](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf). Acesso el 15 nov. 2021.

OLSON, Hope A. **Bacon, warrant and classification.** ASIS&T SIG-CR workshop, november 13, 2004.

OLSON, Hope A. and WARD, Dennis B. Charting a journey across knowledge domains: feminism in the Decimal Dewey Classification. **Advances in knowledge organization**, v. 6, p. 238-244, 1998.

RAFFERTY, Pauline; HIDDENLEY, Rob. Flickr and democratic indexing: dialogic approaches to indexing. **Aslib Proceedings**, v. 59, p. 397-410, 2007.

RODRÍGUEZ, Robert D. Hulme's concept of literary warrant. **Cataloging & Classification Quarterly**, v. 5, n. 1, p. 17-26, 1984.

SACHS, Moshe Y.; SMIRAGLIA, Richard P. From Encyclopedism to domain-based ontology of knowledge management: the evolution of the Sachs Classification (SC). **Advances in Knowledge Organization**, v. 9, p. 167-172, 2004.

SMIRAGLIA, Richard P. **The elements of knowledge organization.** NY: Springer, 2014.

SVANBERG, Magdalena. Classification, warrants and principles. **Swedish Library Research**, v. 2, n. 3, p. 66-75, 1996.

SVENONIUS, Elaine. **The intellectual foundation of information organization.** Cambridge MA: MIT Press, 2000.

TARP, Sven. Reflections on the academic status of lexicography. **Lexikos**, n. 20, 450-465, 2010.

TENNIS, Joseph T. Experientialist epistemology and classification theory". **Knowledge Organization**, v. 32, n. 2: p. 79-92, 2005.

TENNIS, Joseph T., THORNTON, K.; FILER, A. Some temporal aspects of indexing and classification: toward a metrics for measuring scheme change." In **Proceedings of the North American Symposium on knowledge organization**, 2012. Disponível em: <https://ssrn.com/abstract=2881557> Acesso el 15 nov. 2021.

VIZINE-GOETZ, Dianne; BEALL, Julianne. Using literary warrant to define a version of the DDC for automated classification services. **Advances in Knowledge Organization**, v. 9, p. 147-152, 2004.

WARD, Martin. Phenomenological warrant: the case for working from the user's viewpoint. **Managing Information**, v.7, n. 9, p. 68-71, 2000.

# 5

Isadora Victorino Evangelista

José Augusto Chaves Guimarães

**Aspectos éticos em organização  
e representação do conhecimento:  
uma análise de sua presença  
na literatura internacional  
da ciência da informação**

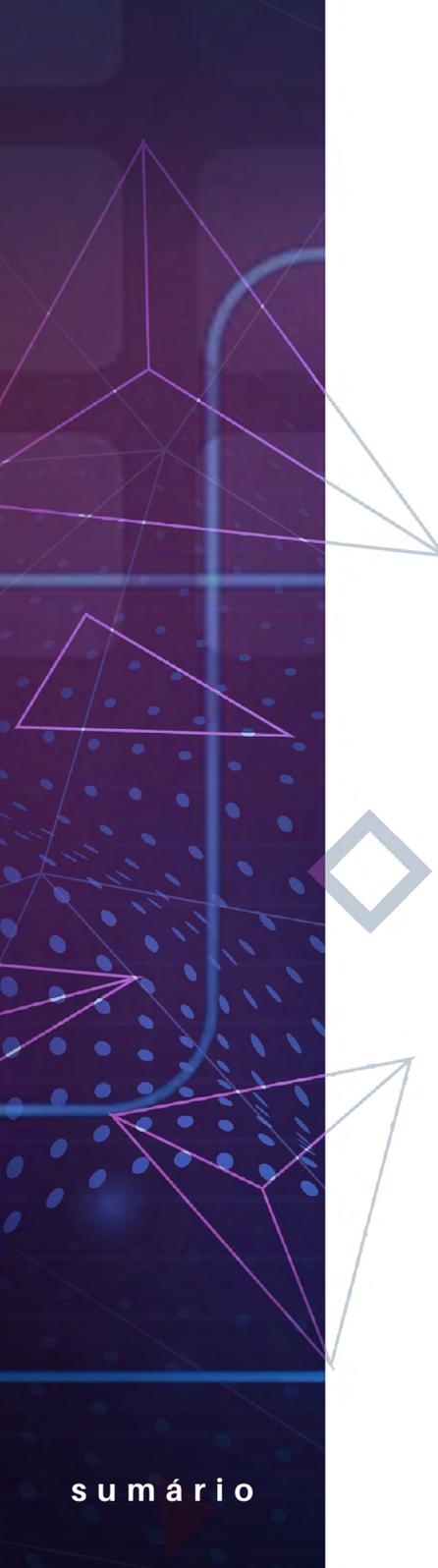
*Ethical aspects  
in knowledge organization:  
an analysis about the theme  
in the international literature  
in information science*

DOI: 10.31560/pimentacultural/2022.95613.5

#### Resumo:

No âmbito da Ciência da Informação, as atividades relacionadas à organização e representação do conhecimento são responsáveis pelo uso de processos e instrumentos que atuam como mediadores entre o documento e a necessidade informacional dos usuários. Essas atividades não são neutras, uma vez que são elaboradas por profissionais que possuem preconceitos, visões de mundo e tendenciosidades construídos ao longo de sua vida. É a partir dessa premissa que uma dimensão ética é identificada nesse âmbito, justificada pelo fato de que os profissionais envolvidos nesses processos podem negar ou dar acesso ao conhecimento, atuando de modo cooperativo ou excludente. Dessa forma, o objetivo do presente capítulo é apresentar uma análise das questões éticas especificamente presentes na organização e representação do conhecimento (ORC), em aspectos como preconceitos e tendenciosidades encontrados em sistemas de organização do conhecimento, o impacto do multiculturalismo em tais instrumentos, e o perigo das marginalizações – por exemplo, de mulheres, negros e homoafetivos – em tais contextos, entre outros aspectos. A partir de uma revisão da literatura internacional da área nessa temática específica, realizada diacronicamente a partir da década de 1970, analisam-se e discutem-se os outros aspectos éticos que ali se fazem presentes até que se chegue aos impactos éticos sofridos por essa área no atual contexto de pluralidade do ambiente digital.

**Palavras-Chave:** Ética. Ciência da Informação. Organização e Representação do Conhecimento.



**Abstract:**

*In the field of Information Science, the activities associated with the knowledge organization are responsible for the use of processes and instruments that act as mediators between the document and the informational need of users. These activities are not neutral, as they are developed by professionals who have prejudices, worldviews and tendencies built up throughout their lives. It is from this premise that an ethical dimension is identified, by the fact that the professionals involved in these processes can deny or give access to knowledge. In this way, the objective of this chapter is to present an analysis of the ethical issues specifically presented in the knowledge organization field, in aspects such as prejudices and tendencies found in knowledge organization systems, the impact of multiculturalism on instruments, and the danger of marginalizations – for example, of women, blacks and homosexuals – in such contexts, among other aspects. Based on a review of the international literature in the area on this specific theme, carried out diachronically from the 1970s onwards, other ethical aspects are analyzed and discussed, that reach the ethical impacts in the current context of plurality of the digital environment.*

**Keywords:** *Ethics. Information Science. Knowledge Organization and Representation.*

## 1 INTRODUÇÃO

A Ciência da Informação, enquanto área que se preocupa com as questões relacionadas ao processamento, ao comportamento e à otimização do uso da informação (BORKO, 1968), possui, em seu cerne, um conjunto de atividades que envolvem processos, produtos e instrumentos de natureza mediadora. Essas atividades, de forma mais ampla, estabelecem a conexão entre um conhecimento produzido e, socializado e registrado e o seu uso e apropriação possibilitando a contínua geração de novo conhecimento, em um movimento de natureza helicoidal (GUIMARÃES, 2006).

Essas atividades, relativas à organização e representação do conhecimento, não são neutras, estando passíveis de influência de visões de mundo, preconceitos e tendenciosidades dos profissionais que as realizam, dos idealizadores dos instrumentos e mesmo dos produtores do conhecimento. Nesse sentido, uma dimensão ética é identificada, visto que os profissionais envolvidos, com o seu “poder de nomear” (OLSON, 2002), acabam podendo dar ou negar acesso ao conhecimento, atuando de modo cooperativo ou excludente (SOUZA, 2002).

Desse modo, pretende-se, no presente capítulo, apresentar uma análise das questões éticas especificamente presentes na organização e representação do conhecimento (ORC), em aspectos como os preconceitos e tendenciosidades encontrados em sistemas de organização do conhecimento, o impacto do multiculturalismo em tais instrumentos, e o perigo das marginalizações – por exemplo de mulheres, negros e homoafetivos – em tais contextos, entre outros. Nesse sentido, e a partir de uma revisão da literatura internacional da área nessa temática específica, realizada diacronicamente a partir da década de 1970, analisam-se e discutem-se os aspectos éticos que ali se fazem presentes até que se chegue aos impactos sofridos por essa área no atual contexto de pluralidade do ambiente digital.

## 2 A QUESTÃO ÉTICA NOS PROCESSOS, PRODUTOS E INSTRUMENTOS DE ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

O primeiro trabalho que menciona essas questões no universo da ORC é a obra seminal *Prejudices and antipatie: a tract of Library of Congress Subjects Headings concerning people*, publicada por Sanford Berman, em 1971. Na obra, o autor busca ressaltar todos os preconceitos encontrados na referida linguagem, amplamente utilizada em várias partes do mundo. O livro é dividido em seis capítulos: “Chauvinismo, ‘Síndrome de Bwana’ e o terceiro mundo”; “Raças, nacionalidades, fés e grupos étnicos”; “Políticas, paz, trabalho, aplicação da lei etc.”; “Homem/mulher/sexo”; “Crianças, adolescentes, ‘idiotas’ e o ‘subterrâneo’ e, por último, “Faça você mesmo”. Cada um dos capítulos fornece exemplos claros de discriminações sofridas por esses grupos na linguagem documentária e uma sugestão de como esses preconceitos poderiam ser evitados caso houvesse uma transformação da linguagem.

Após a publicação do referido livro e pelo fato de a terminologia de representação da informação passar sucessivamente por revisões, alguns cabeçalhos foram alterados – uma vez que os valores da sociedade se alteraram ao longo do tempo. É importante ressaltar, no entanto, o quanto as linguagens eram produzidas sem uma preocupação de natureza ética e como refletiam uma visão de mundo na qual a norma era o homem, branco e cristão (BERMAN, 1971).

No mesmo ano, A. C. Foskett (1971) publica um artigo semelhante, denominado “*Misogynists all: a study in critical classification*”, em que analisa como as classificações podem se demonstrar preconceituosas e tendenciosas, principalmente no que tange aos homossexuais e mulheres. Uma das críticas do autor é que, embora as classificações devam

refletir uma verdade “externa e eterna”, o que realmente ocorre é que elas estão longe de ser objetivas, refletindo os preconceitos do período em que são elaboradas e das pessoas que as elaboram.

Em 1984, Foskett se aprofunda nessas questões. Para o autor, foram vistas mudanças maiores nas ferramentas e produtos que utilizamos do que na maneira e teoria de como classificamos. O autor demonstra, por meio de exemplos da CDD – 19ª ed., que, na classe “Costumes Gerais (394)”, podem ser classificados documentos que tratem de assuntos como palavrões, suicídio e canibalismo; na classe “Controle de população (304.66)”, podem ser listadas práticas como genocídio ou infanticídio; na classe “Ética das relações sociais (177)”, aparecem assuntos como amor e amizade, de maneira conjunta com assuntos completamente diferentes, como escravidão e práticas discriminatórias, profanidade, suicídio, homicídio, genocídio etc.; entre outros.

No ano de 1981, Anthony Judge apresenta uma série de vieses (ou *biases*, como o autor utiliza) que são identificadas no momento da construção de um tesouro. Essas tendências são classificadas pelo autor como: a relação estática que alguns tesouros estabelecem, sem considerar a dinamicidade da linguagem; o fato de os tesouros serem elaborados de acordo com o modo de pensar e a linguagem ocidental; o padrão conservador das ferramentas; a rigidez das estruturas na construção de tesouros; a disfuncionalidade desses instrumentos, elaborados sem considerar ou ressaltar as relações entre fenômenos; o viés demonstrado pela hierarquia de classes; e, por fim, a polarização encontrada nas representações dicotômicas.

Em 1991, Hazel K. Bell elenca cinco possíveis fontes de tendenciosidade no momento da indexação: atitudes autoritárias, quando o indexador reforça opiniões “fortes” expressas pelo autor; indexadores intrusivos, quando o profissional reforça seus próprios preconceitos na representação do documento; imposição da classificação, quando a linguagem demonstra-se inflexível, com cabeçalhos de assunto que não

correspondem da maneira correta (ou ideal) às suas subdivisões; limitações linguísticas pelo fato de as modificações linguísticas ocorrerem mais lentamente do que as alterações sociais (atualmente, há problemas de identificação de minorias na linguagem, subdivisões que podem ofender essas minorias e ainda o sexismo patriarcal); e, por último, a supressão de assuntos, que também pode ser considerada um fator de censura.

Ingetraut Dahlberg, fundadora da International Society for Knowledge Organization – ISKO, elabora, em 1992, um editorial de fascículo da revista Knowledge Organization em memória a Rangathan. Neste, a autora pondera sobre algumas implicações éticas na organização do conhecimento, como, por exemplo, quando se utiliza um sistema de classificação como base para a criação de outro sem a devida permissão do autor; quando documentos são indexados ou classificados de maneira negligente e sem o zelo apropriado; ou, ainda, quando um sistema de informação é comprado por um valor considerável, mas não cumpre com o que se compromete.

Cinco anos após o editorial de Dahlberg, Michele Hudon (1997) reforça a preocupação ética no âmbito da indexação, como já ressaltado por outros autores anteriormente. Em seu artigo, a autora demonstra atenção em relação aos tesouros multilíngues. Para a autora, com o avanço das tecnologias, a linguagem tornou-se uma barreira ainda mais crítica e uma forma de evitar dificuldades é promover uma linguagem controlada que abarque diferentes idiomas, o que permitiria que estrangeiros pudessem acessar os dados dos sistemas de recuperação da informação. No entanto, alguns obstáculos são encontrados ao desenvolver ferramentas como esta: ampliar uma linguagem para que abarque conceitos estrangeiros pode fazê-la ficar pouco reconhecível para seus próprios falantes; transferir toda uma estrutura conceitual de uma cultura para outra, mesmo quando não pareça adequado; a tradução literal de termos de um idioma para outro sem carregar consigo os conceitos próprios. Segundo a autora, produzir um tesouro multilíngue

vai além de achar termos equivalentes em idiomas diferentes, há uma dimensão cultural no processo, o que transforma essa ferramenta em um tesouro multicultural. Nesse contexto, a experiência bilíngue do Canadá (francesa e inglesa) ultrapassa a dimensão linguística na medida em que os termos de cada uma das línguas para a representação de um dado conceito trazem consigo a visão de mundo que aquela cultura tem a respeito (HUDON, 1997).

Um tesouro multicultural é um instrumento que oferece a terminologia e os conceitos de determinados idiomas e toda a estrutura semântica de um tesouro – relações de equivalência, hierárquicas e por afinidade, de maneira fidedigna. Dessa forma, independentemente do idioma da pesquisa, as informações semânticas recuperadas serão as mesmas (HUDON, 1997).

No que concerne à representação da comunidade homoafetiva, Campbell (2000) declara que esse grupo é geralmente marginalizado em ferramentas consideradas universais, pois os termos utilizados usualmente são arcaicos ou inválidos, não satisfazendo às necessidades informacionais desses usuários. Outro problema ressaltado pelo autor é o fato de o processo de indexação ser um ato subjetivo, que varia entre indexadores, com diferentes perspectivas.

Para compreender melhor essas questões, o autor analisa os conceitos de *aboutness* – o principal assunto de um documento – e *meanings* – os usos específicos de determinado documento –, relacionando-os com teorias *queer*. De acordo com Campbell (2000), esses conceitos são especialmente frágeis na literatura homoafetiva, pois são documentos que, em geral, recebem diversos tipos de interpretação e, embora os assuntos permaneçam relativamente estáveis, seus diversos usos se alteram de acordo com a comunidade ou o período.

Beghtol (2002) também reflete sobre as questões culturais na elaboração de sistemas de representação, principalmente relacionadas

aos aspectos éticos no desenvolvimento desses sistemas diante do avanço das novas tecnologias. Para a autora, o ponto essencial reside nos conceitos de garantia e hospitalidade cultural, em que diferentes culturas podem estar inseridas em uma mesma sociedade e elas devem ser representadas pelos sistemas de maneira igualitária. A autora ressalta que a própria cultura reside na informação, visto que as necessidades informacionais daquele grupo de pessoas, como as informações são utilizadas, o que eles valorizam e escolhem perpetuar, podem auxiliar na definição dos aspectos culturais.

A autora propõe o conceito de “hospitalidade cultural”, relacionando-o com o conceito de garantia cultural. Para Beghtol (2002), a hospitalidade cultural assegura que as necessidades individuais de usuários de comunidades específicas que se utilizam do sistema sejam representadas no sistema, permitindo tanto a escolha individual do usuário quanto a representação coletiva de uma comunidade muitas vezes marginalizada pelos sistemas de organização do conhecimento.

A preocupação com os sistemas de organização do conhecimento, especificamente as linguagens de indexação, também é reforçada por Olson (2002). A autora deixa explícita a importância do nomear inerente à indexação e como, muitas vezes, esse ato demonstra-se preconceituoso e enviesado, principalmente quando representando assuntos relacionados a mulheres, latinos e homossexuais. Para a autora, o ato de nomear é uma necessidade básica da comunicação e, como bibliotecários, nós decidimos o que deve ser representado e o que não deve ser nomeado, afetando o acesso e o uso da informação.

Para Olson, os problemas relacionados a essa prática possuem dois aspectos principais: o ato de nomear reflete valores sociais e não há neutralidade nesse processo ou em suas ferramentas, que são geralmente construídas a partir de visões patriarcais. A escolha dos conceitos que serão incluídos no vocabulário controlado demonstra suas limitações, o que é incluído na linguagem e o que

não é permitido, muitas vezes desconsiderando variáveis individuais dos documentos e impondo uma linguagem universal que pode não ser a utilizada pelo usuário (OLSON, 2002).

A respeito das classificações, Olson (2002) afirma que, apesar de elas terem a pretensão de serem universais, são, na verdade, culturalmente específicas, refletindo o *mainstream* da sociedade em que são elaboradas, e diferentes paradigmas culturais, conforme a evolução do tempo podem afetar as ferramentas de representação. O elemento da globalização contribuiu muito para que esses instrumentos se demonstrassem imperialistas e culturalmente homogêneos. Um exemplo desse fator é a Classificação Decimal de Dewey (CDD), que é uma ferramenta produzida no contexto anglo-saxão, em uma estrutura hierárquica do geral para o mais específico – o que também pode ser entendido como do mais para o menos importante, ou estabelecendo a ordem de todo-parte –, mas que foi traduzida para mais de 30 idiomas e utilizada por culturas completamente diferentes (OLSON, 2002).

No mesmo ano, García-Gutierrez (2002) remete ao conceito de “cultura de fronteira” para nos propor uma “ética transcultural de mediação”. De acordo com o autor, vivemos em uma época de permanente mutação, com as posições se alterando constantemente, em que se torna complexo definir onde são os limites e fronteiras dos territórios e, portanto, sistemas fechados e rígidos não podem ser aceitos nesse paradigma. Para tanto, o autor propõe a ideia de “epistemografia interativa”, em que redes semânticas se configuram de maneira aberta e horizontal, diferentemente de uma posição vertical e hierárquica, permitindo, assim, que a diversidade não seja reduzida<sup>46</sup>.

<sup>46</sup> Vale ressaltar que o ano de 2002 foi especialmente importante para a discussão das questões éticas em ORC, pois além da obra seminal de Olson (2002), merece especial destaque a realização da 7<sup>a</sup>. Conferência Internacional da ISKO, em Granada, Espanha, cujo tema foi relativo aos desafios na representação do conhecimento para o século XXI, mais especificamente no que tange às possibilidades – e necessidades – de integração ao longo das fronteiras (geográficas, culturais etc.).

No que concerne às teorias de classificação, Mai (2004) ressalta a importância de um novo paradigma, focado nos estudos de usuários e nas estruturas dos domínios por meio de uma visão construtivista. Dessa maneira, de acordo com o autor, o ideal no momento de criar uma classificação é estudar o universo da unidade de informação que fará uso dessa ferramenta, estudando o discurso daquela comunidade e suas principais atividades e conhecendo os termos utilizados pelos usuários para que os sistemas de organização do conhecimento correspondam a suas necessidades (MAI, 2004).

Ainda relacionado ao âmbito das linguagens de indexação e classificação, Van der Walt (2004) salienta adversidades que podemos encontrar diante da norma básica da Organização do Conhecimento, que é dar acesso à informação. Dentre elas: diferentes usuários possuem necessidades informacionais diversas e, portanto, um termo de indexação pode não servir a todos; o processo de indexação deve ser subjetivo, os assuntos em potencial do documento devem ser reconhecidos; e nós ainda não temos uma teoria sedimentada sobre a relação entre documentos.

Esses fatores implicam problemas éticos, como uma má indexação, termos não correspondentes – de maneira a atribuir uma maior “popularidade” ao documento – e uma representação tendenciosa ou preconceituosa. Para refletir sobre esses problemas, o autor traz alguns questionamentos que merecem maior aprofundamento, como estabelecer quais são os conflitos éticos em ORC, quais são as diretrizes éticas que devem conduzir essa área, quais são as responsabilidades sociais da área, além de reflexões sobre o aspecto pedagógico das unidades informacionais e o acesso livre à informação (VAN DER WALT, 2004).

Feinberg (2007) demonstra grande preocupação com as *biases* implícitas aos sistemas de organização do conhecimento considerado universal. A autora afirma que, mesmo em abordagens como a análise de domínio – proposta na Ciência da Informação por Hjørland e

Albrechtsen (1995) –, é ainda provável que ocorram tendenciosidades e que, possivelmente, a melhor opção não seja estabelecer um único domínio, mas ter em mente que múltiplos domínios coexistem entre si e não apenas diferentes perspectivas em um mesmo domínio.

Para a autora, caso não seja viável eliminar todos os tipos de vieses, o ideal seria ser mais responsável sobre eles, explicitar sua presença e defender a perspectiva que o sistema está adotando. Desse modo, para elaborar um sistema de organização do conhecimento é necessário “tomar partido” sobre como aquele domínio do conhecimento está sendo interpretado e, para tanto, é necessário que esteja claro como a informação é avaliada, incluída, excluída e mantida pela organização. Mesmo que ocorra algum tipo de bias, será de uma maneira responsável, com as perspectivas do sistema e da unidade claras. Deve-se, ainda, estar ciente e ser responsável pelas possíveis consequências que esse posicionamento pode acarretar (FEINBERG, 2007).

Refletindo sobre a teoria crítica racial como abordagem para a análise da CDD, Furner (2007) atenta para a legitimação da ideologia dominante por esquemas de classificação e que o baixo nível de usuários negros em unidades informacionais pode se dever ao fato de eles não se sentirem representados por esses esquemas, ademais a essas unidades não se encontrarem predominantemente em regiões pobres e suburbanas. Por essas circunstâncias, a teoria crítica racial seria de grande valia para bibliotecários: pelo comprometimento social dessas instituições e pela potencial aplicabilidade dessa teoria nesses espaços, de maneira a evitar ao máximo o racismo institucionalizado.

Assim, ao fazer referência a um serviço justo de biblioteca, Furner (2007) cita alguns objetivos – ou valores éticos – que devem ser perseguidos pela equipe da unidade. Por exemplo: a satisfação do usuário, o amparo aos usuários, a equidade de oportunidades, a divulgação dos serviços oferecidos, a liberdade de escolha, as recomendações de qualidade, a responsabilidade na prestação de contas, a

eficiência, a diversidade, conformidade com as normas e padrões e a busca pela estabilidade, evitando-se a opressão, de tal modo que uma biblioteca justa possa ser aquela em que os recursos são utilizados de maneira ativa, de forma a dar suporte às mudanças sociais e reconhecendo os direitos dos grupos oprimidos.

A partir da primeira conferência em Ética na Organização da Informação, realizada na University of Wisconsin-Milwaukee em maio de 2009, a questão ética na Organização e Representação do Conhecimento assumiu um espaço investigativo institucionalizado. A conferência foi publicada na revista *Cataloging and Classification Quarterly* referindo-se a temáticas como a abordagem retórica na tecnologia da informação (BADE, 2009); a falta de reflexão das necessidades dos usuários na prática da catalogação (HOFFMAN, 2009); os aspectos éticos da classificação na segurança nacional (UNSWORTH, 2009); a análise da tabela auxiliar 5 da CDD sobre etnias e grupos nacionais e sua relação com os padrões de representação dessa linguagem (BEALL, 2009); e a análise dos padrões de descrição dos recursos e suas implicações éticas (SMIRAGLIA, 2009).

Em 2012, ocorreu a segunda conferência em Ética na Organização da Informação e, dessa vez, igualmente em Milwaukee, os artigos geraram um número especial da revista *Knowledge Organization*. Esses artigos versam sobre: aspectos relativos à globalização e aos efeitos na relação entre bibliotecários e editores (MARTÍNEZ-ÁVILA; KIPP; OLSON, 2012), a folksonomia de nomenclaturas sexuais e a liberdade de classificação nesse universo (KEILTY, 2012); as implicações éticas de dar nome aos metadados sobre objetos visuais (SEEMAN, 2012); as consequências éticas da digitalização de documentos de arquivo e sua representação (ZHANG, 2012), a representação das ideias e visões de mundo dos autores dos documentos no momento da descrição arquivística (GILLILAND, 2012); a responsabilidade dos bibliotecários no desenvolvimento de coleções e suas atitudes diante

de livros considerados “ruins” perante normas éticas (HOMAN, 2012), a apresentação de um modelo para o desenvolvimento do trabalho do bibliotecário enquanto um ofício majoritariamente intelectual (COPE, 2012); a análise de linguagens documentárias brasileiras e a forma como representam termos homoafetivos (PINHO; GUIMARÃES, 2012); o uso de social tagging como alternativa à utilização de linguagens disciplinares como a da Biblioteca do Congresso (ADLER, 2012); a análise das responsabilidades morais das instâncias de desenvolvimento e uso dos sistemas de organização do conhecimento (FOX; REECE, 2012); e a análise dos conceitos de disfemismo, eufemismo e ortofemismo e as discussões em indexação (MILANI; PINHO, 2012).

Baseando-se na filosofia dos movimentos de Artes do século XX, na teoria crítica da organização do conhecimento e na filosofia do budismo, Tennis (2013) propõe uma organização do conhecimento eticamente comprometida, destacando o trabalho de catalogação e classificação como de natureza não apenas técnica, mas artística, de tal modo que a industrialização e a produção em massa podem afetar sua realização. Indo além, o autor ressalta a necessidade de decisão responsável durante a indexação ou a classificação, ciente de que a ausência desse cuidado pode gerar um ato de violência contra o usuário. Para tanto, evoca trabalhos de Slavoj Žižek e sua sistematização de violência objetiva – aquela presente em linguagens de grupos específicos e em consequência de sistemas políticos e econômicos –, o que se encontra presente nas atividades de indexação, razão pela qual a prática da indexação deve estar comprometida com ações corretas e responsáveis (TENNIS, 2013).

Nessa seara, tem lugar a intenção do indexador, que pode escolher entre agir de maneira a beneficiar ou a prejudicar outros. Para que se possa medir as intencionalidades, o autor propõe cinco níveis: um ato praticado sem intenção nenhuma de prejudicar alguém – como pisar acidentalmente em um inseto; agir de maneira incorreta mesmo

sabendo os possíveis danos, mas quando está fora do controle de si, como bêbado ou apaixonado; realizar um ato prejudicial, mas quando se está enganado sobre a parte prejudicada; agir de maneira incorreta mesmo consciente sobre o que está fazendo e de maneira premeditada; e o último nível é semelhante ao primeiro, mas, nesse caso, o autor da ação não reconhece que está agindo de modo danoso.

Refletindo sobre o conceito de *harm* – ou dano – na organização do conhecimento, Adler e Tennis (2013) ponderam sobre como a linguagem utilizada de forma errônea pode ser considerada um ato de violência e, por conseguinte, as classificações estão sempre suscetíveis a causar algum dano. Como no trabalho citado anteriormente, os autores vão além da perspectiva de “poder do acesso” que as ferramentas – e os bibliotecários – possuem, pois esses instrumentos podem ser empregados de maneira violenta e autoritária, uma vez que eles impõem e controlam a partir de suas utilizações. Esse poder é disciplinado e quem o exerce determina a posição dos que serão classificados ou transformados em categorias. Assim, para evitar esse tipo de violência, os autores elaboraram uma taxonomia do dano, centrando-se em três questões relativas ao momento da classificação: o que acontece, quem participa e quem e como pode ser afetado.

Para responder à primeira questão, os autores relembram algumas ações que podem acontecer no momento da classificação e que tendem a ser danosas, como excepcionalismo, omissão, terminologia enviesada, estrutura do padrão inapropriada, entre outras. Pensando sobre a salubridade dessas ações, reforça-se que os classificadores possuem a responsabilidade de causar o menor dano possível, e, para tanto, deve-se recorrer a atitudes sábias e conscientes. Outro componente considerado essencial pelos autores é a intencionalidade desses atos – o que já havia sido alertado por Tennis (2013). Por último, devem-se considerar as implicâncias desses atos, pois, ao invés de darem acesso aos recursos, esses sistemas centrados na objetividade e padronização acabam por impedir o uso desses documentos (ADLER; TENNIS, 2013).

Sobre a participação dos envolvidos, os autores questionam sobre os níveis de cooperação desses participantes. Quem produz esses sistemas possui um poder elevado em relação àqueles que escolhem, dentre os termos determinados, quais utilizar; este segundo grupo possui um poder maior que um terceiro grupo, daqueles que apenas importam os dados desses registros sem modificá-los. Em qualquer um desses níveis é possível uma decisão ética e crítica em relação a essas ferramentas (ADLER; TENNIS, 2013).

Sobre as possíveis lesões causadas por essas questões, os autores reforçam que ainda que a principal lesão seja a falta de acesso à informação, esses danos se estendem à propagação de discursos dominantes e premissas discriminatórias. Como exemplo desses danos, os autores citam o uso do termo “parafilia” pela Biblioteca do Congresso, de forma a representar “desvios sexuais” e o termo “eugenia”, utilizado pela CDD até a década de 1950 e depois excluído, mesmo sendo o termo utilizado pela literatura, possivelmente pelo fato de representar algo que “negativo” para a sociedade (ADLER; TENNIS, 2013).

Em 2015, teve lugar a terceira edição da conferência sobre Ética em Organização da Informação. Nesta, discorreu-se sobre: o desenvolvimento de uma taxonomia das críticas à organização do conhecimento (GROSS, 2015); a cobertura de artigos relativos à temática em ética em organização do conhecimento em duas bases de dados importantes da área – *Web of Science* e *Library and Information Science and Technology Abstracts* (MARTÍNEZ-ÁVILA; GUIMARÃES; PINHO; FOX, 2015); a análise das temáticas dos livros banidos pela American Library Association (KIPP; BEAK; GRAF, 2015); a representatividade do público infantil em sistemas de organização do conhecimento (BEAK, 2015); a representação de músicas tradicionais em sistemas de organização do conhecimento (WEISSENBERGER, 2015); a análise do RDA e de uma faceta do modelo conceitual FRBR focado no bibliocentrismo da representação descritiva (SMIRAGLIA, 2015); o estudo sobre o conceito de

cultura e como ele pode ser levado em consideração no momento da construção de sistemas de classificação para diferentes grupos (LEE, 2015); o feito das práticas de classificações sobre dietas no desenvolvimento de manuais nacionais para a saúde e o entendimento sobre esse assunto pela população (McTAVISH, 2015); os aspectos éticos envolvidos em práticas acadêmicas como plágio, coautoria ou citações (ORAVEC, 2015); a análise de conceitos bakhtinianos em relação aos profissionais da organização do conhecimento e sua responsabilidade em mediar discursos no momento da representação (ARBOIT; GUIMARÃES, 2015); a análise de valores como representação fidedigna, acessibilidade e controle em esquema de metadados arquivístico no âmbito artístico (McQUEEN, 2015); a proposta de um modelo analítico para a avaliação de práticas de classificação (McCOURRY, 2015); a análise de valores éticos para processos de descrição e classificação arquivística por meio de códigos de ética da área (SILVA; GUIMARÃES; TOGNOLI, 2015); a proposta de recomendações para um código de ética em catalogação (SHOEMAKER, 2015); e o estudo sobre a classificação utilizada por serviços como a Netflix para recomendações de serviços e sua falta de neutralidade (LAWRENCE, 2015).

### 3 CENÁRIO BRASILEIRO NO ÂMBITO ÉTICO EM ORGANIZAÇÃO DO CONHECIMENTO

A temática de pesquisa relativa aos aspectos éticos em organização do conhecimento teve forte repercussão na Ciência da Informação brasileira, em especial pelo forte componente social que a integra. Assim, no ano 2000, José Augusto Chaves Guimarães apresentou um conjunto de compromissos éticos do profissional da informação, divididos em cinco instâncias: com o usuário, com a organização, com a informação, com a profissão e consigo mesmo. O primeiro compromisso considera que o usuário não é meramente o destinatário de uma

entrega, mas aquele que irá se apropriar da informação e produzir um novo conhecimento; a organização ou unidade de informação constitui-se como compromisso do profissional, pois é ali que ele encontra os meios para a realização da profissão; em relação à profissão, seu compromisso ético deve estar expresso para que haja "(...) aspirações conjuntas e práticas integradas" (GUIMARÃES, 2000, p. 66); por último, há o compromisso consigo mesmo, do profissional como indivíduo que possui preferências, ideias e perspectivas de autorrealização.

Em 2006, Guimarães propões uma série de valores ou premissas éticas, dessa vez voltadas às atividades de organização e representação do conhecimento. Para o autor, faltava uma sistematização desses compromissos éticos, que eram subjacentes a esses processos. Dessa forma, o autor chama a atenção para fatores como precisão, tempo, especificidade, garantia literária, revocação, preconceitos nas linguagens documentárias, entre outros, que até então eram visualizados como meros aspectos, mas que, na verdade, implicam responsabilidade do profissional que os executa.

O autor identifica, a partir de uma análise no periódico *Ethics and Information Technology*, um grupo de cinco valores éticos que merecem destaque nas atividades de organização do conhecimento: privacidade, propriedade intelectual, compromisso, segurança e liberdade. São reconhecidos ainda um grupo de valores diretamente ligados ao desenvolvimento tecnológico, como diversidade na representação, desvios ou representação insuficiente e má-fé na representação.

Considerando o exposto pela literatura pesquisada, Guimarães (2006) formula então um grupo de problemas concretos que podem ocorrer por conta do descumprimento desses valores. Estes são: os preconceitos que residem nos sistemas de organização do conhecimento, as categorias dicotômicas estabelecidas por essas linguagens, suas visões de mundo específicas, problemas relacionados à precisão

dos termos utilizados, a polissemia dos termos e ainda a questão do politicamente correto na representação.

Também em 2006, Fábio Pinho elabora sua dissertação fazendo uma triangulação entre os conceitos de epistemografia interativa, de García Gutiérrez, o de multilinguismo, de Hudon, e as concepções de garantia e hospitalidade cultural preconizadas por Beghtol. Para o autor, há aspectos convergentes entre os discursos desses autores, como, por exemplo, a ideia de epistemografia interativa difundida por García Gutiérrez como um grupo de diretrizes para a elaboração de representações em meio a um universo de diferentes culturas e discursos por meio de uma linguagem documentária comprometida com a representação da diversidade e utilizando-se, principalmente, de relações associativas, para que nenhum assunto seja priorizado. Nessa perspectiva, encontra-se uma maneira de se efetivar esse instrumento por meio dos estudos de Hudon a respeito da elaboração de um tesouro multilíngue enquanto ferramenta que respeita e pode ser inserida em diferentes sociedades e culturas (PINHO, 2006).

García Gutiérrez e Hudon ainda trazem conceitos como “purificação conceitual” e “imperialismo linguístico”, respectivamente, para refletir sobre a predominância de determinados idiomas nos documentos científicos, o que afeta seu acesso e uso por nações não falantes de idiomas dominantes, muitas vezes sendo essas caracterizadas como subdesenvolvidas. Esses conceitos vão ainda ao encontro do que Beghtol afirma ser necessário: uma integração entre o conhecimento e a ciência mediante um sistema de representação efetivo e agregador. A autora ainda evoca o conceito de “ética transcultural”, de García Gutiérrez, no sentido de que, para um sistema possuir garantia cultural, sem que outras culturas se sintam excluídas, é necessário que haja um elemento de união entre essas diferentes culturas (PINHO, 2006).

No tocante aos problemas e valores éticos em OC, Guimaraes *et al.* (2008) analisaram um grupo de seis revistas internacionais

importantes na área — *Journal of the American Society for Information Science and Technology* — JASIST, *Journal of Documentation, Knowledge Organization, Cataloging & Classification Quarterly, The Indexer e Ethics and Information Technology*. Para recuperar os artigos relacionados à temática, foram estabelecidos dois domínios conceituais: Ética (por meio dos termos *ethics, ethical, ethos, deontology, value(s), conduct e moral*) e Organização do Conhecimento (por meio dos termos *knowledge organization, knowledge representation, indexing, classification, subject cataloguing, subject analysis, content analysis, abstracting, thesaurus(i) e subject headings*), compreendendo como período de busca os anos de 1995 a 2004.

Desse modo, chegaram a um grupo de valores intrinsecamente relacionados à ORC. Este pode ser dividido em três dimensões: valores superiores que devem guiar as atividades profissionais – respeito à privacidade, direitos autorais, acessibilidade, liberdade, segurança, equidade, diversidade e minimização de riscos; valores reconhecidos enquanto requisitos profissionais – competência, eficiência, flexibilidade, confiabilidade, reconhecimento profissional, atualidade, autonomia, consciência de poder e cooperação; e, por fim, valores antes vistos apenas como medidas, mas que foram reconhecidos enquanto parte do universo axiológico da OC – precisão, revocação, garantia cultural, exaustividade, consistência, usabilidade e hospitalidade (GUIMARÃES *et al.*, 2008).

Por meio do *corpus*, chegou-se ainda a um grupo de problemas éticos associados à OC. Estes estão divididos em: problemas não específicos à OC, mas onipresentes na atualidade – segregação digital, pornografia, lixo eletrônico, profissionais substituídos pela tecnologia e violência; e problemas diretamente relacionados às atividades de OC – vigilância, censura, falta de garantia cultural, negligência, informação direcionada, ineficiência profissional, má representação, racismo, ambiguidade, marginalização, imparcialidade ou crença na neutralidade,

idiosincrasia, inacessibilidade informacional, terminologia tendenciosa e traduções inadequadas (GUIMARÃES *et al.*, 2008).

No que tange ao universo feminino, Milani (2010) buscou analisar, nas ferramentas de indexação, como esse grupo é representado. Por meio dos termos feminilidade, feminina(s), feminino(s), feminismo(s), feminista(s), materna(o), maternal, mulher(es) e seus respectivos equivalentes em inglês, a autora realizou uma busca nas linguagens do Vocabulário Controlado da Universidade de São Paulo, Vocabulário Controlado Básico do Senado Federal, Terminologia de Assuntos da Fundação Biblioteca Nacional e Classificação Decimal de Direito para analisar como esses termos eram representados. Na referida obra, a autora estabelece uma distinção conceitual entre os vieses negativos (*biases*) e os vieses sem essa conotação (*slants*). Por meio dessa análise, a autora identificou uma série de *biases* – tendenciosidades – nessas linguagens, como, por exemplo: a violência conjugal vista como uma questão apenas de saúde pública, negligenciando os aspectos sociológicos e criminais; a mulher enquanto um tipo específico e o homem como geral (mulheres delinquentes, mulheres casadas, mulheres na ciência etc.); qualificadores de gênero sem a devida necessidade (educação feminina); o sufixo *-ismo* em lesbianismo, que caracteriza ser uma doença, ao invés de condição afetiva; entre outras (MILANI, 2010).

Buscando aprofundar-se na temática relacionada às linguagens de indexação no que tange aos grupos marginalizados, Pinho e Guimarães (2012) realizaram uma análise da representação dos homossexuais masculinos em linguagens brasileiras. O ponto de partida dos autores é que as linguagens podem estar carregadas de eufemismos preconceituosos, visto que, muitas vezes, principalmente no universo da homossexualidade, elas representam conceitos-tabus e são consideradas impróprias. Esses eufemismos nem sempre estão claros, o que pode revelar uma tentativa de “abafar” o termo ou atribuir-lhe uma perspectiva de politicamente correto. Como exemplo, o autor cita

que a expressão “garoto sensível e educado”, dita para referir-se a um homossexual, pode, além de mascarar o que a pessoa realmente quer dizer, ser ofensiva ou irônica para a pessoa referenciada.

Como universo de pesquisa, Pinho e Guimarães (2012) selecionaram três das mais importantes linguagens controladas do país, a saber: Vocabulário Controlado Básico do Senado Federal, o Vocabulário Controlado da Universidade de São Paulo e a Terminologia de Assuntos da Fundação Biblioteca Nacional. Para realizar a análise, os autores selecionaram um grupo de termos oriundos de artigos publicados em revistas científicas de destaque na temática homossexual – *Journal of Homosexuality*, *Sexualities* e *Journal of Gay & Lesbian Mental Health* – e os comparou com os termos utilizados pelas linguagens documentárias.

Como resultados, os autores puderam perceber que grande parte dos termos utilizados pela comunidade representada – aqueles advindos das revistas da área – não se encontra refletido nas linguagens documentárias analisadas. Uma das justificativas para esse fato é que esses termos geralmente são tratados de maneira periférica, sem uma efetiva reflexão sobre sua representatividade. Para Pinho e Guimarães (2012), visto que é impossível representar todo o conhecimento humano, deve-se, pelo menos, ter em mente a sua pluralidade e, para tanto, os estudos éticos tornam-se especialmente necessários.

Novamente no âmbito axiológico da área, utilizando-se metodologia semelhante a pesquisas anteriores, Guimarães, Milani e Evangelista (2015) expandem o universo de busca para os anos de 1990-1994 e 2005-2012 e incluem também a série *Advances in Knowledge Organization* e os *Proceedings* da *North American Society for Knowledge Organization*, de maneira a ampliar os resultados de Guimarães *et al.* (2008) e atualizar esse grupo de valores e problemas. Nessa nova pesquisa, foram incluídos valores como respeito ao propósito específico de cada linguagem de indexação, multiculturalismo, comprometimento com as bases filosóficas da organização da informação e postura mediadora

na indexação. Em relação aos problemas, foi possível identificar ainda *biases* na representação de assunto, falta de comprometimento com a catalogação descritiva, distorção da informação, categorias remanescentes gerando empecilhos ao fluxo informacional.

De maneira a abordar em dois valores específicos, que até então eram vistos apenas como medidas, Evangelista, Simões e Guimarães (2016) realizaram um estudo analisando a exaustividade e a especificidade enquanto valores éticos da organização do conhecimento. Para tanto, foram recuperados artigos das bases de dados B-On, EBSCO e *Search Direct*, da Elsevier, visto que esse estudo foi realizado durante um estágio de pesquisa na Universidade de Coimbra e estas serem as bases mais relevantes ao nível português. Dessa forma, identificaram-se variáveis que vêm à tona nesses processos: o número de conceitos extraídos de um documento, a forma e o nível de especificidade que são representados são diretamente proporcionais à informação que pretende representar-se; os níveis de exaustividade e de especificidade concorrem para o impacto na recuperação informacional; elevados níveis de exaustividade concorrem a uma maior atribuição de termos a um documento, que se fará de acordo com o serviço e o usuário; um elevado nível de especificidade concorre para uma maior precisão do documento; a especificidade pode contribuir para a elaboração de uma rede semântica consistente no que diz respeito a relações hierárquicas; um conjunto significativo de termos específicos atribuídos a um documento expressa de forma mais precisa os conceitos; a articulação entre a exaustividade e a especificidade concorre para uma maior recuperação informacional; a exaustividade corresponde a um nível elevado de revocação; a especificidade corresponde a níveis elevados de precisão; a exaustividade traz regularidade e densidade à indexação; e, por fim, as relações semânticas entre termos-conceitos são identificadas de forma mais precisa a partir de meios intelectuais (EVANGELISTA, SIMÕES, GUIMARÃES; 2016; EVANGELISTA, 2016).

De maneira a identificar esses valores éticos, mas no âmbito arquivístico, Silva, Tognoli e Guimarães (2017) realizaram estudo relativo às atividades de descrição e classificação consideradas nucleares para a representação arquivística. Para constatar esses valores, foram recuperados 16 códigos de ética da profissão que compreendiam dez países e uma associação internacional, que, após uma primeira análise – relativa à presença ou não dos termos classificação e representação –, chegou-se a um *corpus* de análise de dez artigos. Foi, assim, possível identificar, referente ao processo de classificação, valores como: acesso à informação; respeito à proveniência; respeito à ordem original; preservação do valor arquivístico do documento; preservação; conservação; confiabilidade; autenticidade; custódia; e imparcialidade. Na segunda categoria de análise, o processo de descrição, chegou-se a valores como: acesso à informação; respeito à proveniência; respeito à ordem original; preservação; confiabilidade; custódia; e imparcialidade (SILVA; TOGNOLI; GUIMARÃES, 2016).

A preocupação de a linguagem refletir sua comunidade usuária, principalmente no que tange aos usuários homossexuais masculinos, também é objeto de reflexão de Guimarães, Nascimento e Pinho (2017). Os autores abordaram essa realidade na região do Cariri, no interior do Ceará, que, embora seja conhecida por seus hábitos religiosos, possui minoria que foge à normativa heterossexual. De forma a representar essa comunidade até então marginalizada, os termos utilizados por essa população foram categorizados, chegando-se a um total de 815 termos metafóricos, que podem ser divididos em três categorias temáticas: nomes femininos; partes do corpo, geralmente fazendo referência a órgãos sexuais; e ações geralmente relativas a ações sexuais e traços de personalidade. Em aspectos linguísticos, os autores perceberam uma forte influência de termos indígenas e estrangeiros e identificaram que as figuras de linguagem constituem uma forma de essa comunidade construir sua identidade, utilizando-se, muitas vezes, de metáforas, o que é um grande desafio para a ORC (GUIMARÃES; NASCIMENTO; PINHO, 2017).

Posteriormente, os autores, em coautoria com Campbell e Martínez-Ávila, relacionaram esses resultados com a questão do “poder de se autoneomear”, em alusão ao termo evocado por Olson (2002) “poder de nomear”. Para os autores, esses padrões terminológicos utilizados por essas comunidades reforçam três aspectos: o poder de afirmar sua identidade, o poder de ironizar ou realizar uma crítica social e ainda o poder que evoca a autoproteção dessa população. Esses fatores devem ser considerados no momento da construção de uma linguagem documental que abarque regiões como a estudada (CAMPBELL *et al.*, 2017).

Por meio de uma reflexão da área, Guimarães (2017) propõe que a perspectiva ética nos processos, produtos e instrumentos em Organização do Conhecimento deve sempre pautar-se nos eixos espaço e tempo, sem os quais ela se torna descontextualizada. Para o autor, esse campo é afetado por influências geográficas e diacrônicas, socialmente dinâmico, o que reflete na construção terminológica das ferramentas inerentes a esse campo científico, o que pressupõe uma variedade cultural e linguística. Nesse sentido, Guimarães afirma que é necessário encontrar um equilíbrio entre o respeito por especificidades locais e a necessidade de um acesso global à informação. De acordo com o autor, o reconhecimento desses eixos tempo e espaço demonstram sua importância quando consideramos que: evita-se preconceitos presentes e disseminados até então; reconhece os diferentes grupos e comunidades usuárias das unidades informacionais; promove uma ética intercultural, evitando danos no futuro.

## 4 CONSIDERAÇÕES FINAIS

Por meio da revisão de literatura exposta neste capítulo, apresentada diacronicamente de modo a melhor situar como a temática foi se desenvolvendo ao longo do tempo, foi possível perceber que,

embora a temática relativa aos aspectos éticos Organização e Representação do Conhecimento tenha dado os primeiros passos há cerca de cinquenta anos, ainda há um caminho a ser percorrido que o tempo todo é atualizado face aos novos contextos tecnológicos e sociais encontrados atualmente.

Ainda que as pesquisas descritas não sejam classificadas de maneira explícita pelos autores como investigações na dimensão ética da Ciência da Informação, as preocupações foco das pesquisas analisadas mostram o inconformismo dos autores com preconceitos evidenciados nessas ferramentas, ainda que de maneira velada, e como esses fatores podem afetar o acesso e recuperação da informação de grupos marginalizados.

Considerando a importância de estudos nesse sentido e que a mudança nessas ferramentas só ocorre com dados científicos, recomenda-se fortemente que essas pesquisas continuem a ocorrer e que, mais do que isso, que esses estudos sejam sistematizados e reunidos a partir de estudos como análise de domínio ou de comunidades epistêmicas, de modo a que essas informações sejam evidenciadas e utilizadas como fonte para novos estudos, buscando sempre a retroalimentação da ciência.

## 5 REFERÊNCIAS

ADLER, Melissa A. Disciplining knowledge at the Library of Congress. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 370-376, 2012.

ADLER, Melissa; TENNIS, Joseph T. Toward a taxonomy of harm in knowledge organization systems. **Knowledge Organization**, Würzburg, v. 40, n. 4, p. 266-272, 2013.

ARBOIT, Aline Elis; GUIMARÃES, José Augusto Chaves. The ethics of knowledge organization and representation from a Bakhtinian perspective. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 324-331, 2015.

BADE, David. Ethos, Logos, Pathos or Sender, Message, Receiver?: a problematological rhetoric for information technologies. **Cataloging and Classification Quarterly**, Londres, v. 47, n. 07, p. 612-630.

BEAK, Jihee. Where is childrens' voice in KO? **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 284-289, 2015.

BEGHTOL, C. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation**, London, v. 58, n. 5, p. 507-532, 2002.

BEALL, Julianne. Racially mixed people, DDC Table 5 Ethnic and National Groups, and MARC 21 Bibliographic Format Field 083. **Cataloging and Classification Quarterly**, Londres, v. 47, n. 07, p. 657-670.

BELL, Hazel. Bias in indexing and loaded language. **The Indexer**, Sheffield, v. 17, n. 03, 1991, p. 173-177.

BERMAN, S. **Prejudices and antipathies**: a tract of Library of Congress Subjects Headings concerning people. Metuchen: Scarecrow Press, 1971.

BORKO, Harold. Information science: what is it? **American Documentation**, Washington, v. 19, n. 1, p. 3-5, 1968.

LL, Grant. Queer theory and the creation of contextual subject access tools for gay and lesbian communities. **Knowledge Organization**, Würzburg, v. 27, n. 03, p. 122-131, 2000.

CAMPBELL, Grant *et al.* The terminological polyhedron in LGBTQ terminology: self-naming as a power to empower in Knowledge Organization. **Knowledge Organization**, Würzburg, v. 44, n. 08, p. 586-591, 2017.

COPE, Jonathan. Librarianship as intellectual craft: the ethics of classification in the realms of leisure and waged labor. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 356-362, 2012.

DAHLBERG, Ingetraut. Ethics and knowledge organization: in memory of Dr. S. R. Ranganathan in his centenary year. **International Classification**, Frankfurt, v. 19, n. 1, p. 1-2, 1992.

FOX, Melodie J.; REECE, Austin. Which Ethics? Whose Morality?: an analysis of ethical standards for information organization. **Knowledge Organization**, Würzburg, n. 39, v. 5, p. 377-383, 2012.

EVANGELISTA, Isadora Victorino. **A exaustividade e a especificidade como valores éticos no processo de indexação**: aspectos conceituais e deontológicos. 2016. 142 f. Dissertação (Mestre em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. Marília, 2016.

EVANGELISTA, Isadora Victorino; SIMÕES, Maria Graça Melo; GUIMARÃES, José Augusto Chaves. A exaustividade e a especificidade como valores éticos no processo de indexação: uma análise baseada na literatura disponibilizada em Portugal. **Páginas A&B**, Porto, v. 3, n. 5, p. 58-75, 2016.

FEINBERG, Melanie. 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"**, v. 12, n. 04, 2007.

FOSKETT, A. C. Better dead than read: further studies in critical classification. **Library Resources & Technical Services**, Illinois, v. 28, n. 04, p. 346-359, 1984.

FOSKETT, A. C. Misogynists all: a study in critical classification. **Library Resources & Technical Services**, Illinois, v. 15, n. 2, p. 117 - 121, 1971.

FURNER, Jonathan. Dewey deracialized: a critical race-theoretic perspective. **Knowledge Organization**, Würzburg, v. 34, n. 3, 2007.

GARCÍA-GUTIÉRREZ, A. 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, 2002. p. 516-522.

GILLILAND, Anne J. Contemplating co-creator rights in archival description. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 340-346, 2012.

GROSS, Tina. Naming and reframing: a taxonomy of attacks on knowledge organization. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 263-268, 2015

GUIMARÃES, José Augusto Chaves. O profissional da informação sob o prisma de sua formação. *In*: VALENTIM, M. L. P. (Ed.). **Profissionais da informação**: formação e atuação profissional. São Paulo: Polis, 2000. p. 53-70.

GUIMARÃES, José Augusto Chaves. Aspectos éticos em Organização e Representação do Conhecimento (ORC): uma reflexão preliminar. *In*: GONZÁLEZ DE GÓMEZ, Maria Nélide; ORRICO, Evelyn Goyannes Dill (Org.). **Políticas de memória e informação**: reflexos na organização do conhecimento. Natal: EDUFERN, 2006.

GUIMARÃES, José Augusto Chaves *et al.* 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, 2008. p. 340-346.

GUIMARÃES, José Augusto Chaves; MILANI, Suellen Oliveira; EVANGELISTA, Isadora Victorino. Valores éticos na organização e representação do conhecimento. **Encontros Bibli**, Florianópolis, v. 20, n. 1, p. 19-32, 2015.

GUIMARÃES, José Augusto Chaves; NASCIMENTO, Francisco Arrais; PINHO, Fabio Assis. The metaphorical dimension of LGBTQ information: challenges for its subject representation. **Informação & Sociedade**, João Pessoa, v. 27, n. 3, p. 49-57, 2017.

GUIMARÃES, José Augusto Chaves. Slanted knowledge organization as a new ethical perspective. *In*: ANDERSEN, Jack; SKOUVIG, Laura (Org.). **The organization of knowledge**: caught between global structures and local meaning. Bingley: Emerald Publishing Limited, v. 12, p. 87-102, 2017.

HOFFMAN, Gretchen. Meeting users' needs in cataloging: what is the right thing to do? **Cataloging and Classification Quarterly**, Londres, v. 47, n. 07, p. 631-641, 2009.

HOMAN, Philip A. Library catalog notes for "Bad Books": ethics vs. responsibilities. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 347-355, 2012.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information Science: domain-analysis. **Journal of the American Society for Information Science**, v.46, n.6, p.400-425, 1995.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization**, Würzburg, v. 24, n. 2, p. 84-91, 1997.

JUDGE, Anthony. Anti-developmental biases in thesaurus design. **Laetus in praesens**, S.l., 1981. Disponível em: <http://www.laetusinpraesens.org/docs/antidev.php>. Acesso em: 6 ago. 2018.

KEILTY, Patrick. Tagging and Sexual Boundaries. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 320-324, 2012.

KIPP, Margaret E. I.; BEAK, Jihee; GRAF, Ann M. Tagging of banned and challenged books. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 276-283, 2015.

LAWRENCE, Emily. Everything is a recommendation: netflix, altgenres and the construction of taste. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 358-364, 2015.

LEE, Wan-Chen. Culture and classification: an introduction to thinking about ethical issues of adopting global classification standards to local environments. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 302-307, 2015.

MAI, J.-E. Classification in context: relativity, reality, and representation. **Knowledge Organization**, Würzburg, v. 31, n. 1, p. 39-48, 2004.

MARTÍNEZ-ÁVILA, Daniel; GUIMARÃES, José Augusto Chaves; PINHO, Fabio Assis; FOX, Melodie J. The representation of ethics and knowledge organization in the WoS and LISTA databases. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 269-275, 2015.

MARTÍNEZ-ÁVILA, Daniel; KIPP, Margaret E. I.; OLSON, Hope A. DDC or BISAC: the changing balance between corporations and public institutions. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 309-319, 2012.

McCOURRY, Maurine. Domain analytic, and domain analytic-like, studies of catalog needs: addressing the ethical dilemma of catalog codes developed with inadequate knowledge of user needs. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 339-345, 2015.

McQUEEN, Kelli. Ethical issues of knowledge organization in designing a metadata schema for the Leo Kottke Archives. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 332-338, 2015.

McTAVISH, Jill. The ethics of querying and permeating canadian everyday life nutritional classification technologies and processes. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 308-315, 2015.

MILANI, Suellen Oliveira. **Estudos éticos em representação do conhecimento**: uma análise da questão feminina em linguagens documentárias brasileiras. 2010. 140 f. Dissertação (Mestre em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho. Marília: Universidade Estadual Paulista Júlio de Mesquita Filho, 2010.

MILANI, Suellen Oliveira; PINHO, Fabio Assis. Knowledge Representation and Orthophemism: a reflection aiming to a concept. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 377-393, 2012.

OLSON, H. A. **The power to name**: marginalizations and exclusions of subject representation in library catalogues. 1996. 432 f. Tese (Doutorado em Library and Information Studies) – University of Wisconsin-Madison, Madison, Wisconsin, United States, 1996.

OLSON, H. A. **The power to name**: locating the limits of subject representation in libraries. Dordrecht: Kluwer Academic Publishers, 2002.

ORAVEC, Jo Ann. The moral imagination in an Era of “Gaming Academia”: implications of emerging reputational issues in scholarly activities for knowledge organization practices. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 316-323, 2015.

PINHO, Fabio Assis. **Aspectos éticos em representação do conhecimento:** em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol. 2006. 123 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves. Male homosexuality in brazilian indexing languages: some ethical questions. **Knowledge Organization**, v. 39, n. 5, p. 363-369, 2012.

SEEMAN, Dean. Naming names: the ethics of identification in digital library metadata. **Knowledge Organization**, v. 39, n. 5, p. 325-331, 2012.

SHOEMAKER, Elizabeth. No one can whistle a symphony: seeking a catalogers' code of ethics. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 353-357, 2015.

SILVA, Andrieli Pachú; GUIMARÃES, José Augusto Chaves; TOGNOLI, Natália Bolfarini. Ethical values in archival arrangement and description: an analysis of professional codes of ethics. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 346-352, 2015.

SILVA, Andrieli Pachu da; TOGNOLI, Natalia Bolfarini; GUIMARÃES, José Augusto Chaves. Os valores éticos na organização e representação do conhecimento arquivístico. **Brazilian Journal of Information Studies: research trends**, Marília, v. 11, n. 1, p. 44-53, 2017.

SOUZA, Francisco das Chagas. Ética e deontologia: textos para profissionais atuantes em bibliotecas. Florianópolis: Editora da UFSC, 2002.

SMIRAGLIA, Richard. Bibliocentrism, cultural warrant, and the ethics of resource description: a case study. **Cataloging and Classification Quarterly**, Londres, v. 47, n. 07, p. 671-686, 2009.

SMIRAGLIA, Richard P. Bibliocentrism revisited: RDA and FRBRoo. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 296-301, 2015.

TENNIS, Joseph. Ethos and ideology of knowledge organization: toward precepts for an engaged knowledge organization. **Knowledge Organization**, Würzburg, v. 40, n. 1, p. 42-49, 2013.

UNSWORTH, Kristene. Ethical Concerns of Information Policy and Organization in National Security. **Cataloging and Classification Quarterly**, Londres, v. 47, n. 07, p. 642-656, 2009.

WALT, Martin Van der. Ethics in indexing classification. International Society for Knowledge Organization - Germany, 9., 2004, Duisburg. **Exposição oral [...]**. Proceedings... Duisburg: Universität Duisburg, 2004.

WEISSENBARGER, Lynnsey K. Traditional musics and ethical considerations of knowledge and documentation processes. **Knowledge Organization**, Würzburg, v. 42, n. 5, p. 290-295, 2015.

ZHANG, Jane. Archival context, digital content, and the ethics of digital archival representation. **Knowledge Organization**, Würzburg, v. 39, n. 5, p. 332-339, 2012.

# 6

Luciana Monteiro-Krebs

Rita do Carmo Ferreira Laipelt

Rafael Port da Rocha

**Metodologia de análise  
de *logs* na ciência da informação:  
revisão de literatura e melhores práticas**

*Logs analysis as methodology  
for studies on information science:  
literature review and best practices*

DOI: 10.31560/pimentacultural/2022.95613.6

### Resumo:

Em um cenário de crescente acesso à informação através de plataformas digitais e sistemas de informação, pesquisas envolvendo a análise de *logs* tornam-se cada vez mais relevantes. Através dos *logs*, pesquisadores e profissionais da informação podem conhecer as demandas dos usuários com quem não mantêm contato direto, em função da prestação de serviço se dar remotamente, mediada via sistema de recuperação da informação. No entanto, estudos que utilizam a análise de *logs* como método no Brasil são ainda incipientes, e descrições sistemáticas dos procedimentos metodológicos na literatura nacional sobre o tema inexistem. Para preencher essa lacuna, o objetivo deste capítulo é investigar, sistematizar e descrever as etapas metodológicas e aspectos técnicos da análise de *logs* em CI, com enfoque em sistemas de recuperação da informação. Concretiza-se esse objetivo através da sistematização da seção metodológica dos 10 artigos mais citados disponíveis na *Web Of Science* que tenham feito uso de análise de *logs*, publicados no período entre 2006 e 2016. A proposta metodológica apresenta quatro etapas principais: (i) contextualização do *log*; (ii) seleção; (iii) coleta e preparação dos dados; e (iv) análise dos dados. A etapa de seleção (ii) é ainda dividida em três subetapas, nas quais se identificam a relevância e adequabilidade das informações do *log* aos objetivos de pesquisa, o recorte de dados e a disponibilização e uso dos dados para análise. Os procedimentos são apresentados com exemplos tanto dos textos do *corpus* de pesquisa quanto de estudos mais recentes baseados na experiência dos autores do presente capítulo.

**Palavras-chave:** Análise de *logs*; metodologia da pesquisa; sistemas de recuperação da informação; estudos de usuário.

**Abstract:**

*In a scenario of increasing access to information through digital platforms and information systems, research involving log analysis becomes increasingly relevant. Through system logs, researchers and librarians can learn about the demands of users with whom they do not have direct contact, due to the fact that the informational service is provided remotely, usually mediated via an information retrieval system. However, studies that use log analysis as a method in Brazil are still incipient, and systematic descriptions of methodological procedures in the national literature on the subject do not exist. To fill this gap, the objective of this chapter is to investigate, systematize and describe the methodological steps and technical aspects of log analysis in Information Science with a focus on information retrieval systems. This objective is achieved through the systematization of the methodological section of the 10 most cited articles available on the Web Of Science that have made use of log analysis, published between 2006 and 2016. The methodological proposal has four main steps: (i) contextualization of the log; (ii) selection; (iii) data collection and preparation; and (iv) data analysis. The selection step (ii) is further divided into three sub-steps, being the relevance and suitability of the log information to the research objectives, the methodological cut and the availability of data for analysis. The procedures are presented with examples both from the texts of the research corpus and from more recent studies based on the experience of the authors of this chapter.*

**Keywords:** Log analysis; research methodology; information retrieval systems; user studies.

## 1 INTRODUÇÃO

No Brasil, cerca de 152 milhões de indivíduos são usuários da internet, o equivalente a 81% da população com 10 anos ou mais (CE-TIC, 2021). Esse índice, crescente principalmente em centros urbanos, possibilita às unidades de informação no país ampliarem o conhecimento a respeito de seus usuários, expressos através de seu comportamento online. As interações do usuário com o sistema aplicativo são gravadas virtualmente em arquivos de *log*, registrando, para cada requisição, a data, o horário, o local de acesso (obtido via número IP do computador que realizou o acesso), assim como a ação determinada pela requisição e sua situação, entre outras informações. A análise de tais dados é importante para a melhoria dos serviços disponibilizados e o atendimento às necessidades informacionais dos usuários em um cenário cada vez menos presencial de acesso à informação. No que tange a estudos de usuário na Ciência da Informação (CI), a análise dos registros de uso em plataformas digitais (registros em *logs*) oferece a possibilidade de expandir tanto a quantidade de usuários incluídos nas amostras quanto a granularidade da informação disponível.

No entanto, a literatura da CI a respeito do uso dos registros de interações dos usuários em bases de dados, sites e/ou catálogos online ainda é incipiente no Brasil. Por exemplo, uma consulta<sup>47</sup> sobre o tema na base de dados BRAPCI<sup>48</sup> em junho de 2021 retornou apenas cinco publicações. Em sua maioria, os estudos sobre análise de *logs* na CI brasileira apresentam resultados de estudos empíricos, com pouco ou nenhum aprofundamento metodológico. O trabalho de Laipelt (2015) e Monteiro-Krebs, Rocha e Ribeiro (2017) são exemplos de pesquisas

47 A expressão de busca utilizada foi "análise de *logs*", no campo Resumo.

48 BRAPCI (<http://www.brapci.ufpr.br/brapci/index.php/home>) é uma base que indexa 57 periódicos científicos brasileiros na área de Ciência da Informação com cobertura temporal desde 1972.

realizadas no Brasil que analisam interações entre sistemas de recuperação da informação e usuários. Laipelt (2015) analisou expressões de busca dos usuários do Portal LexML (Senado Federal Brasileiro) para demonstrar o potencial dos *logs* como fonte de coleta de dados para a escolha de descritores para a representação da informação. Monteiro-Krebs, Rocha e Ribeiro (2017) analisaram o uso de um sistema de recomendação para catálogos on-line de bibliotecas universitárias. Já Aires (2003) descreveu a análise de *logs* e como ferramenta para incrementar a qualidade dos resultados das máquinas de busca, porém não realizou uma análise de *logs* propriamente dita, por não ter obtido acesso aos *logs* das máquinas de busca comerciais. O que de fato foi realizado foi um estudo das interações de usuários com motores de busca a partir de relatórios de consultas redigidos pelos próprios sujeitos da pesquisa. Efetivamente, esse trabalho não poderia ser considerado uma análise de *logs*, pois o *log* propriamente dito não foi utilizado como fonte de dados.

Observa-se, portanto, uma carência em publicações que tragam orientações metodológicas para pesquisas científicas focadas na análise de *logs*, em especial com relação aos procedimentos técnicos para a seleção, a coleta e a análise destes. O presente capítulo oferece subsídios metodológicos para pesquisadores e profissionais da informação interessados em implementar a análise de *logs* através de análise sistemática das principais publicações sobre o assunto e uma proposta metodológica com base nas lições aprendidas dessas publicações.

Tendo em vista o potencial do uso de *logs* em pesquisas no campo da CI, o objetivo deste capítulo é investigar, sistematizar e descrever as etapas metodológicas e aspectos técnicos da análise de *logs* em CI com enfoque em sistemas de recuperação da informação. O procedimento metodológico é proposto com base em uma revisão de literatura, utilizando como fonte pesquisas reconhecidas pela comunidade científica que utilizaram *logs*. Desta forma, colhemos as melhores práticas em estudos da CI com alto volume de citações que utilizam a análise de

*logs* como método. Ao encontrar diferentes descrições, buscamos identificar, analisar e discutir os procedimentos apresentados pelos autores na tentativa de sistematizá-los e apresentá-los de forma convergente. Essa contribuição visa oferecer subsídios aos profissionais e pesquisadores da área fornecendo, também, exemplos práticos de aplicação da análise de *logs* em benefício das unidades de informação no país, além de identificar um panorama de como a questão dos procedimentos técnicos relacionados à análise de *logs* é abordada nas pesquisas.

Este capítulo está organizado da seguinte forma. A seção 2 caracteriza a análise de *logs* no contexto da recuperação da informação em sistemas de informação, apresentando a definição de análise de *logs* e introduzindo diferentes tipos de análise. A seção 3 apresenta a metodologia, identificando as pesquisas na área da CI que utilizam análise de *logs* e que apresentam os procedimentos técnicos utilizados com relação aos *logs*, tendo como fonte a plataforma *Web of Science*. A seção 4 apresenta etapas que caracterizam procedimentos metodológicos na análise de *logs*, delineadas a partir da análise das pesquisas identificadas na seção 3. O capítulo é encerrado com uma breve conclusão na seção 5.

## 2 ANÁLISE DE LOGS NA RECUPERAÇÃO DA INFORMAÇÃO

Trabalhos empíricos utilizando *logs* começaram sendo chamados de *web searching studies* (JANSEN; POOCH, 2000; DAVIS, 2004), *search engine transaction log studies*, *log analysis*, entre outros. Ao longo do tempo, alguns termos foram se consolidando na literatura da CI, como é o caso da análise do *log* de transações (*transaction log analysis* - TLA). “A análise do *log* de transações é o uso de dados coletados em um *log* de transações para investigar uma questão de pesquisa

específica relacionada ao usuário, ao sistema ou ao conteúdo” (SPINK; JANSEN, 2005, p. 36). As transações são todas as interações entre o usuário e o sistema, que não precisam, necessariamente, dar-se no ambiente web. Para estas últimas, Jansen, Taksa e Spink (2009) usam o termo *Web log analysis*.

A análise do log de transações é uma ampla categorização de métodos que abrange várias subcategorizações, incluindo análise de logs da Web (ou seja, análise de logs do sistema da Web), análise de blog e análise de logs de pesquisa (análise de logs de mecanismos de busca) (JANSEN; TAKSA; SPINK, 2009, p. 2).

Segundo Jansen (2006, p. 408), “[...] um *log* de transações é um registro eletrônico de interações que ocorreram durante um episódio de pesquisa entre um mecanismo de pesquisa da Web e usuários pesquisando informações nesse mecanismo de pesquisa da Web.”. Um *log* de transações, segundo Jansen (2006), registra a comunicação entre os usuários e um sistema em um arquivo. Os *logs* “[...] representam os usuários, são pegadas de informação digital” (NICHOLAS; HUNTINGTON; WATKINSON, 2005, p. 250).

Na literatura, no entanto, não há convergência para essa classificação. Diferentes termos designam técnicas operacionalmente distintas entre si, mas, em linhas gerais, contempladas pela TLA. Assim como Jansen, Taksa e Spink (2009) por vezes utilizam *log analysis* para se referir à TLA, Nicholas, Huntington e Watkinson (2005) também utilizam *log analysis* quando se referem a *deep log analysis* (DLA - análise profunda de *logs*), que é diferente de TLA, embora, ao analisar os métodos descritos em seus artigos e artigos citados por eles, percebe-se que a *deep log analysis* é, na prática, um aprimoramento da TLA. A DLA agrega, aos dados de pesquisa/navegação obtidos via *log*, informações contextuais relativas aos usuários, que podem ser demográficas ou pessoais extraídas a partir de outras técnicas de coletas de dados (questionários, entrevistas, dados de *login*/identificação etc.), com isso, tem-se a caracterização como análise “profunda” de *logs*.

Um exemplo desse tipo de estudo é o realizado por Nicholas *et al.* (2006a) que investigam o comportamento de busca de informações de quase três milhões de usuários à medida que exploram o site<sup>49</sup>, o número de visitas realizadas, bem como o tipo de itens e o conteúdo visualizado em duas bibliotecas digitais de periódicos, a EmeraldInsight e a Blackwell Synergy. O que caracteriza esse estudo como DLA é a combinação desses dados com a identificação de perfis de usuários por profissão, local de trabalho, tipo de assinatura do periódico, localização geográfica, tipo de universidade, número de itens visualizados em uma sessão etc. (NICHOLAS *et al.*, 2006a, p. 1345). Uma excelente revisão dos estudos de pesquisa na Web (*Web searching studies*) é resumida por Jansen e Pooch (2001). Para acessar um histórico mais antigo de TLA, recomendamos consultar Peters (1993).

Na busca de informação, o usuário pode utilizar filtros e categorias que determinem o conjunto de resultados que ele quer obter, para além da expressão de busca. Após realizar a busca, o usuário navega pelos resultados, selecionando as obras que decidir ler ou acessar o registro. É a partir dos registros gerados por essas interações entre os usuários e o sistema que a análise de *logs* se faz possível. As expressões de busca, em que itens dos resultados o usuário clicou, quais documentos foram visualizados e/ou baixados e todos os demais dados podem ser usados como variáveis para pesquisas no campo da Ciência da Informação.

Através de estudos com análise de *logs*, pode-se identificar quais métodos de recuperação da informação são preferidos pelos usuários de determinadas áreas do conhecimento. No ambiente de informação eletrônica, estudos de usuários de periódicos on-line apontam a preferência geral dos usuários para pesquisar por assunto em bancos de dados ao invés de navegar pelos periódicos (CHEN, 2010;

49 O número de itens ou páginas visualizadas em uma sessão é chamado de "site penetration", que traduzimos como "penetração no site" (ver seção 4.1) (NICHOLAS *et al.*, 2006a; NICHOLAS *et al.*, 2007).

MENG-XING; CHUN-XIAO, YONG, 2010). O estudo de Vakkari e Talja (2006) identifica essa tendência, principalmente para ciências naturais e medicina, em relação a outras disciplinas. Além da busca e da navegação, outros métodos também se mostram relevantes para a RI, (Recuperação da Informação) como o encadeamento de citações, que se mostrou “[...] um método de pesquisa significativamente mais relevante em economia e engenharia em comparação com humanidades e medicina” (VAKKARI; TALJA, 2006, documento eletrônico).

Apesar de haver distinção conceitual entre busca de informação e navegação, para a análise de *logs*, ambas são interessantes e podem auxiliar nos estudos que visam à melhoria dos serviços informacionais oferecidos pela instituição. Na busca de informação (ou pesquisa), podem-se analisar os termos que compõem a expressão de busca – no intuito de melhorar o vocabulário controlado e rede de remissivas, assim como identificar tendências temáticas de pesquisa. Na navegação, por outro lado, pode-se analisar a jornada do usuário, que itens ele visualizou e em quais ele clicou etc. – no intuito de melhorar a estrutura do site, formatos de conteúdo, caminhos internos.

Recentemente, outros tipos de análises que incluem informações oriundas de *logs* começaram a ganhar popularidade entre gestores de sistemas e bibliotecas, chamando a atenção de pesquisadores também. Com o advento de ferramentas de análise estatística (como AWStat, Google Analytics, entre outras), os gestores passam a ter acesso a gráficos e dashboards com resumos de tráfego para os sites que gerenciam. Essas ferramentas leem os *logs* do sistema e mostram os resultados em formas de tabelas e até gráficos, o que facilita a tomada de decisão no que tange a potenciais melhorias do sistema. No entanto, tais ferramentas são diferentes da análise de *logs* como concebida em nosso trabalho (DLA e TLA) por três motivos:

Primeiro, o pesquisador não possui acesso ao *log* em si, apenas aos resultados por meio da análise estatística da ferramenta. Isso

limita as informações que a ferramenta apresenta, ou seja, nem todas as variáveis são computadas nem são passíveis de cruzamentos. A decisão do que é ou não apresentado é feita anteriormente à consulta do pesquisador, ainda na fase de *design to dashboard*, e não há a possibilidade de o pesquisador escolher quais variáveis deseja cruzar para atender seus objetivos de pesquisa. Por exemplo, se a ferramenta determina que o IP dos acessos será apagado dos registros para preservar a privacidade dos usuários, o pesquisador não poderia se valer dessa informação (IP) para analisar a jornada de um usuário específico. Apenas dados agregados estariam disponíveis, permitindo somente estudos de conjuntos de usuários, o que é mais interessante para estudos quantitativos e menos valioso para análises qualitativas.

Segundo, em geral, essas ferramentas também não documentam exaustivamente de onde vêm os valores apresentados nas estatísticas. Assume-se que as informações provêm dos *logs*, porém não é possível encontrar nenhuma documentação que indique qual campo do *log/http* é utilizado em cada análise estatística. Nas análises TLA e DLA, esses dados são conhecidos pelos pesquisadores, até por uma questão de transparência metodológica e replicabilidade.

Terceiro, os relatórios estatísticos também reúnem dados de outras fontes, como, por exemplo, o site visitado pelo usuário antes de chegar na *home* do sistema analisado. Através de APIs, data e hora de consulta são incorporadas à análise e, em alguns casos, até expressões inseridas nos motores de busca são oferecidas nos relatórios. Todas essas características impedem o pesquisador de isolar e cruzar variáveis, além de estabelecer livremente os recortes de dados de acordo com os seus objetivos de pesquisa. As ferramentas de análise estatística, como Google Analytics e AWStat, estão limitadas ao que os relatórios que essas empresas definem previamente e, portanto, não se encaixam na definição de análise de *logs* conforme a metodologia aqui descrita.

### 3 PESQUISAS EM ANÁLISE DE LOGS NA CIÊNCIA DA INFORMAÇÃO

Várias pesquisas mencionam que seus estudos foram feitos a partir da análise de *logs*. Entretanto, essas pesquisas pouco relatam os procedimentos utilizados com relação aos *logs*. Esta seção identifica e caracteriza pesquisas que usam *logs*.

Realizou-se uma busca na bibliografia especializada em que a análise de *logs* consta como assunto principal da publicação ou foi mencionada junto aos métodos em pesquisas na Ciência da Informação. Optou-se pela busca na *Web of Science* pela abrangência da base, pelos recursos de recuperação e filtragem dos conteúdos e pela facilidade em gerar relatórios preliminares para rápida visualização de características do *corpus*. A coleta foi realizada em 20 de agosto de 2017. Foi realizada busca na plataforma *Web of Science* por trabalhos sobre análise de *logs* ou que usam análise de *logs* como metodologia, na área da Ciência da Informação<sup>50</sup>. A maior parte dos artigos encontrados são estudos aplicados.

A triagem dos textos selecionados obedeceu a dois critérios: impacto (número de citações) e adequação ao escopo desse artigo. Após a busca na base, ordenamos a lista de resultados (141 artigos) por número de citações (começando pelo mais citado). O artigo mais citado da lista contém 297 citações e o menos citado, uma citação.

A partir desse extrato, utilizamos como segundo filtro os objetivos do presente capítulo. Nosso intuito é criar um protocolo metodológico para auxiliar profissionais e pesquisadores da Ciência da

<sup>50</sup> Expressão de busca utilizada na *Web of Science*: Você pesquisou por: **Tópico:** ("log analysis" OR "análise de logs" OR TLA OR "transaction log analysis") **OR Título:** ("log analysis" OR "análise de logs" OR TLA OR "transaction log analysis") **Refinado por:** **Categorias do Web of Science:** (INFORMATION SCIENCE LIBRARY SCIENCE) **Tempo estipulado:** 2006-2016. **Índices:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI. **Resultados:** 141 (de Principal Coleção do *Web of Science*).

Informação a realizar estudos utilizando análise de *logs*. Assim, faz sentido nos guiarmos pelos materiais que trouxessem embasamento teórico sobre os métodos e, por isso, artigos exclusivamente de relatos de pesquisa foram desconsiderados. Como a maioria está relatando estudos aplicados a partir do uso da análise de *log*, selecionamos os que mais traziam definições de *log*, das variáveis que o compõem e as etapas do processo de pesquisa. O quadro abaixo mostra os artigos que serviram de fundamento para a sistematização apresentada neste capítulo. Ele fornece os dados dos artigos analisados e como eles foram utilizados nessa revisão bibliográfica.

**Quadro 1 – Artigos mais citados entre os que utilizam análise de *logs***

Título	Autor(es)	Periódico	Data	# citações	Utilizado na seção
How are we searching the World Wide Web? A comparison of nine search engine transaction logs	Jansen, B.J.; Spink, A.	INFORMATION PROCESSING & MANAGEMENT. v. 42, n. 1, p. 248-263	JAN 2006	297	4.3
Search log analysis: What it is, what's been done, how to do it	Jansen, B.J.	LIBRARY & INFORMATION SCIENCE RESEARCH, v. 28, n. 3, p. 407-432	2006	81	4.1 4.2 4.4
Defining a session on web search engines	Jansen, B.J.; Spink, A.; Blakely, C.; Koshman, S.	JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, v. 58, n. 6, p. 862-871	APR 2007	63	4.1
The information seeking behaviour of the users of digital scholarly journals	Nicholas, D.; Huntington, P.; Jamali, H.R.; Watkinson, A.	INFORMATION PROCESSING & MANAGEMENT, v. 42, n. 5, p. 1345-1365	2006a	40	4.3

E-textbook use, information seeking behaviour and its impact: Case study business and management	Nicholas, D.; Rowlands, I.; Jamali, HR.	JOURNAL OF INFORMATION SCIENCE, v. 36, n. 2, p. 263-280	APR 2010	37	4.2.1 4.4
What deep log analysis tells us about the impact of big deals: case study OhioLINK	Nicholas, D.; Huntington, P.; Jamali, HR.; Tenopir, C.	JOURNAL OF DOCUMENTATION, v. 62, n. 4, p. 482-508	2006b	32	4.4
Characterising and evaluating information seeking behaviour in a digital environment: Spotlight on the 'bouncer'	Nicholas, D.; Huntington, P.; Jamali, HR.; Dobrowolski, T.	INFORMATION PROCESSING & MANAGEMENT, v. 43, n. 4, p. 1085-1102	2007	22	4.1
Online use and information seeking behaviour: institutional and subject comparisons of UK researchers	Nicholas, D.; Clark, D.; Rowlands, I.; Jamali HR.	JOURNAL OF INFORMATION SCIENCE, v. 35, n. 6, p. 660-676	2009	18	4.3
Empirical observations on the session timeout threshold	Huynh, T.; Miller, J.	INFORMATION PROCESSING & MANAGEMENT, v. 45, n. 5, p. 513-528	2009	9	4.1
Library and information resources and users of digital resources in the humanities	Warwick, C.; Terras, M.; Galina, I.; Huntington, P.; Pappa, N.	PROGRAM-ELECTRONIC LIBRARY AND INFORMATION SYSTEMS, v. 42, n. 1, p. 5-27	2008	9	4.3

Fonte: elaborado pelos autores.

A partir dos artigos selecionados, realizamos a sistematização dos conteúdos com foco nas principais contribuições potenciais para a CI. Analisando os estudos, observou-se a relevância de estágios anteriores à coleta de dados para a análise de *logs*, conforme destacado pelos trabalhos de Nicholas *et al.* (2005), Jansen (2006), Nicholas *et al.* (2007), Jansen *et al.* (2007) e Huynh e Miller (2009). Por exemplo, Nicholas *et al.* (2005) explicam como a análise de *logs* se diferencia de outros métodos de pesquisa, ao refletir

ações dos usuários sem a interferência do pesquisador. Os demais autores apresentam definições de diversos conceitos relevantes, como, sessão, consulta, penetração no site e itens visualizados. Ao sistematizar os procedimentos metodológicos, essas informações foram incorporadas na etapa de contextualização do *log* (seção 4.1).

Também se observou como os autores explicam a etapa de seleção dos *logs* (seção 4.2), o que é destacado por Jansen (2006) e Nicholas, Rowlands e Jamali (2010). Especificamente, três subetapas da seleção foram identificadas. Primeiro, como o pesquisador pode compreender a relevância e adequabilidade das informações do *log* (seção 4.2.1). Nessa fase, a sistematização traz exemplos de linhas de *log* e informações fornecidas pelo registro. Segundo, como o recorte de dados pode ser feito dependendo dos objetivos da pesquisa (seção 4.2.2). E terceiro, a disponibilização e uso dos dados (seção 4.2.3), quando ocorre a negociação com os responsáveis pela guarda do *log* para acesso a estes. Cuidados com a privacidade dos usuários são necessários e explicados nesta seção.

Diversos trabalhos entre os mais citados apontam ainda como a etapa de coleta e preparação de dados (seção 4.3) se dá empiricamente. A sistematização traz exemplos de possibilidades de análise de expressões de busca (JANSEN; SPINK, 2006) e análises de diversas outras variáveis, especialmente aplicáveis para DLA, conforme destacado por Nicholas *et al.* (2006a; 2009) e Warwick *et al.* (2008).

Finalmente, a análise de dados (seção 4.4) é explorada. Apresenta-se como os estudos de *logs* podem contribuir para melhoria de sistemas de recuperação da informação, com exemplos de aplicabilidade baseados em Jansen (2006). Evidencia-se também a possibilidade de cruzamento de variáveis do *log* com dados de diferentes fontes, como questionário e grupo focal. Exemplos desses cruzamentos (típico de DLA) e os resultados esperados são demonstrados com citações de Nicholas, Rowlands e Jamali (2010) e Nicholas *et al.*, (2006b). As limitações dos estudos com *logs* (JANSEN, 2006) são apresentadas na conclusão.

## 4 SISTEMATIZAÇÃO DOS PROCEDIMENTOS TÉCNICOS DE ANÁLISE DE LOGS NA CIÊNCIA DA INFORMAÇÃO

A análise de *logs* não é um método simples de ser empregado, por dois motivos principais. O primeiro é a dificuldade de acesso à matéria-prima, haja vista que diversos cuidados devem ser empregados para garantir a segurança da informação e o respeito à privacidade dos usuários cujas interações estão registradas nos *logs*. A anonimização de IP é um exemplo. O segundo motivo é o conhecimento técnico que se demanda para manusear os *logs*. É necessário entender minimamente de programação para entender a linguagem e a estrutura dos arquivos que se vai receber.

Porém, informações extraídas dessa fonte são valiosas por, pelo menos, duas razões. Em primeiro lugar, a confiabilidade dos *logs*. Usar *logs* como fonte de dados para conhecer o comportamento informacional elimina qualquer eventual interferência do pesquisador sobre o usuário no momento da consulta e acesso à informação. As informações dos *logs* são fiéis aos fatos, pois provém um registro direto e imediato do que as pessoas de fato fizeram durante a interação com o sistema, “[...] não o que dizem que poderiam ou gostariam de fazer; nem o que eles foram levados a dizer, tampouco o que eles pensam que fizeram” (NICHOLAS *et al.*, 2005, p. 1445). Em segundo lugar, as pessoas têm dedicado cada vez menos tempo para participar de pesquisas e preencher questionários (LAIPELT, 2015). Portanto, através dos *logs*, é possível coletar grandes volumes de dados com riqueza de detalhes de forma quase instantânea, sem as limitações de tempo e espaço impostas pela disponibilidade dos participantes da pesquisa para questionários ou entrevistas, por exemplo.

Para facilitar essa tarefa, nessa seção apresenta-se uma sistematização dos procedimentos e aspectos técnicos utilizados em análise de *logs*, a partir do exame dos trabalhos apresentados no Quadro 1. Essa sistematização inicia pela compreensão das interações entre usuário e sistema aplicativo e o registro dessas interações em arquivos de *log*, prosseguindo por aspectos e procedimentos que envolvem seleção, coleta, preparação e análise.

Segundo Jansen (2006), a análise de *logs* envolve três etapas principais: (i) coleta: processo de recolha dos dados de interação de um determinado período em um *log* de transações; (ii) preparação: processo de limpeza e organização dos dados de *log* de transações; e (iii) análise: processo de exame dos dados com vistas a alcançar os resultados da pesquisa. Nesse estudo, adiciona-se às etapas de Jansen a contextualização do *log*, na qual os *logs* são primeiramente compreendidos pelo pesquisador no contexto das interações entre usuários e Sistemas de Recuperação da Informação (SRI), e a seleção, na qual são feitas análises prévias do escopo e das viabilidades do uso de um *log* em uma pesquisa. Essas duas questões são fundamentais para a tomada da decisão sobre o uso ou não de um *log* e para minimizar riscos que possam vir a ocorrer em passos subsequentes da pesquisa. Elas foram propostas por este estudo à medida que se observou nos artigos analisados preocupações em demonstrar *logs* e seu funcionamento, assim como as viabilidades e as possibilidades do uso desses *logs* nos estudos.

Partindo dessas etapas, a seguir apresenta-se a proposta metodológica baseada em uma sistematização da literatura sobre a análise de *logs*, considerando, além das etapas previstas por Jansen, outras questões que se consideram relevantes do ponto de vista de diversos autores que são citados ao longo do texto.

#### 4.1 CONTEXTUALIZAÇÃO DO LOG

A contextualização compreende a percepção, por parte do pesquisador que utilizará *logs* em suas análises, do contexto em que ocorrem as interações entre o sistema aplicativo, os usuários e o registro dessas interações no *log*. No contexto dos sistemas aplicativos que operam na plataforma Word Wide Web, um *log* representa o registro das interações entre usuários e sistemas que rodam em Servidores Web, que, no caso desse capítulo, são SRI. Jansen (2006, p. 419) define interação como “[...] qualquer troca específica entre o pesquisador e o sistema (ou seja, enviar uma consulta, clicar em um hiperlink etc.)”.

Ao realizar suas buscas e ao navegar pelos resultados dessas buscas, um usuário de um SRI realiza várias interações, determinadas através de requisições ao servidor. Por exemplo, ao preencher e enviar um formulário que contém uma expressão de busca, o usuário está realizando uma requisição, que terá como resposta o resultado da consulta expressa no formulário. Ao clicar em um item do resultado dessa consulta, o usuário está realizando uma outra requisição. Essa requisição terá como resposta a visualização das informações bibliográficas do item.

As interações ocorrem através do protocolo HTTP, que é um protocolo de requisição-resposta entre o cliente (usuário) e o servidor (sistema). Os *logs* são arquivos situados no servidor em que ficam registradas informações referentes às requisições. *Common Log Format* (CFL) é o formato básico para representação de arquivos de *logs*. *Logs* representados em CFL registram: o endereço **IP** do computador do usuário que realizou a requisição; a identificação do **usuário**; a **data** e o **horário** da requisição; a **requisição**; um código que identifica o **estado** da requisição (como sucesso ou erro); e o tamanho em **bytes da resposta** à requisição.

Em uma requisição, informações são passadas ao servidor. Essas informações são específicas de cada sistema aplicativo e basicamente indicam a ação solicitada ao sistema (como a de realizar uma consulta) e os parâmetros necessários para a execução dessa ação (como a expressão de uma consulta).

No protocolo HTTP, as requisições não possuem estado, isto é, para o servidor, cada requisição é independente da outra, não havendo o conhecimento de que duas ou mais requisições participam de um mesmo conjunto de interações e compartilham uma mesma situação ou estado. Entretanto, para permitir a realização de interações com estado, sessões podem ser implementadas pelos sistemas aplicativos. Segundo Jansen (2006), uma **sessão** compreende uma série de interações que ocorrem em um determinado período para atender a uma ou mais necessidades de informação.

O sistema aplicativo determina quando uma sessão é iniciada ou encerrada. Por exemplo, uma sessão pode ser iniciada por um determinado sistema aplicativo quando o usuário acessa o site ou quando este executa uma determinada função da aplicação. Já o término de uma sessão, por exemplo, pode ser determinado quando o usuário executa uma determinada função ou por situações como tempo limite ou desativação do navegador. As sessões são identificadas por um código interno, e todas as requisições que são realizadas em uma mesma sessão carregam o código dessa sessão, que também fica registrado no *log*, junto ao registro de cada requisição. Na análise de *logs*, esse código de sessão é usado para recuperar todas as requisições que ocorreram em uma mesma sessão, e o início e o fim de uma sessão são conhecidos, respectivamente, pelas datas da requisição mais antiga e da última requisição registrada na sessão.

A delimitação das sessões faz parte da estratégia que cada sistema estabelece para realizar suas interações com o usuário. Jansen (2006), considerando os *logs* analisados em sua pesquisa, caracteriza

a sessão como “episódio de pesquisa”, e afirma que “A duração da sessão é o tempo total que o usuário passou interagindo com o mecanismo de pesquisa, incluindo o tempo gasto visualizando o primeiro e os documentos subsequentes da Web, exceto o documento final.” (JANSEN, 2006, p. 419)<sup>51</sup>. No sistema da biblioteca digital utilizado por Nicholas *et al.* (2007), a sessão é identificada por um número único no *log*. Para cálculo de tempo da sessão, eles utilizam etiqueta (*tag*) de início de sessão e uma *tag* de finalização de sessão. “Se isso não estiver disponível, os pesquisadores escolhem um prazo máximo no qual um usuário pode estar inativo e considere isso como o final dessa sessão” (NICHOLAS *et al.*, 2007). Huynh e Miller (2009) trazem diversos exemplos de estudos com diferentes métricas para definir o tempo de sessão. Para contribuir com os estudos no campo, os autores apresentam um modelo matemático para estimar o tempo de sessão baseado em observações empíricas. Jansen *et al.* (2007) compararam três métodos que contemplam diferentes variáveis para estabelecer qual a melhor maneira de determinar o que é uma sessão de pesquisa. Para os autores, o método que apresenta a melhor identificação contextual, com extensão e duração da sessão, é o que utiliza as variáveis endereço IP, *cookie*, e alterações de conteúdo das consultas.

Na análise de *logs* de sistemas aplicativos de recuperação de informação, normalmente os seguintes conceitos são relevantes:

- **Consulta:** “Uma consulta é definida como uma seqüência de zero ou mais termos submetidos a um mecanismo de pesquisa” (JANSEN, 2006, p. 418, tradução nossa). A consulta é registrada no *log* como informação adicionada a uma requisição que determina a execução de uma consulta. Cada sistema aplicativo possui uma forma específica para representar uma requisição que solicita a execução de consulta.

51 A duração da sessão é tipicamente curta, com pesquisadores da Web usando entre cinco e 120 minutos (JANSEN, 2006).

- **Penetração no site:** Nicholas *et al.* (2006a) e Nicholas *et al.* (2007) utilizam esse termo (*site penetration*) para nomear a quantidade de itens ou páginas visualizadas em uma sessão. Isso pode ser obtido a partir da identificação das requisições registradas no *log* que determinam a visualização de itens.
- **Itens visualizados:** Esse conceito varia de sistema para sistema. Para Nicholas *et al.* (2007, 1088, tradução nossa), considera-se um “[...] item ‘completo’ retornado pelo servidor ao cliente em resposta a uma ação do usuário”. Segundo o autor, arquivos de *log* tradicionais diferem grandemente dos registros de biblioteca digital. Os primeiros registram imagens e documentos de texto separadamente. Já nos registros de biblioteca digital, um item completo pode ser todas as páginas, gráficos etc. de um artigo, e isso é registrado como um único item, incluindo um resumo, um artigo ou um sumário. (NICHOLAS *et al.*, 2007).

## 4.2 SELEÇÃO

A seleção compreende a definição do escopo e a identificação das viabilidades do uso dos *logs* na pesquisa. Envolve verificar se o sistema produz informações relevantes e adequadas para a realização de uma pesquisa científica, definir o recorte temporal, negociar a disponibilidade e o uso dos dados, incluindo questões com relação à privacidade dos dados.

### 4.2.1 Relevância e adequabilidade das informações do *log*

As questões de pesquisa vão nortear quais dados devem ser coletados e em que período, porém deve-se observar que cada sistema possui características específicas criadas durante o seu

desenvolvimento, o que implica padrões predefinidos pelos aplicativos de software (JANSEN, 2006, p. 412; LAIPELT, 2015). Outro motivo de variação é a “[...] técnica empregada no momento da coleta dos mesmos” (LAIPELT, 2015, p. 166). Nicholas *et al.* (2005, p. 1446) evidenciam diferentes formas de medir o uso dos documentos: o registro pelo servidor de um único item “[...] pode incluir um resumo, um artigo ou um índice”, ou pode-se considerar como um item completo todas as páginas, gráficos etc. de um artigo, se isso for assim registrado pelo servidor. Tradicionalmente, no entanto, os arquivos de *log* gravam imagens e documentos de texto separadamente.

Fica claro, portanto, que a variedade de possibilidades de informação a serem obtidas nos *logs* depende do que o servidor apresenta, e cabe ao pesquisador definir as métricas utilizadas para análise. Assim, o pesquisador precisa adequar-se ao que oferece o *log* em termos de dados para articular suas questões de pesquisa tendo em mente o que a fonte é capaz de fornecer. Isso evidencia a importância de uma etapa de planejamento antes da coleta dos dados nos *logs*.

A seleção envolve verificar se o sistema em questão contém informações relevantes para a análise e se o *log* é adequado e suficiente para a realização dessa análise. Isso pode ser realizado primeiramente através da experimentação (operação) do ambiente, pela sua exploração via interface de usuário, com a anotação das **interações relevantes ao estudo** e das informações associadas a cada interação. Com base nesse levantamento de interações, pode-se identificar **variáveis** e **indicadores** potenciais para o estudo.

O pesquisador pode navegar no sistema, realizar consultas e anotar os resultados, procurando entender que tipo de retorno o sistema oferece via pesquisa. Por exemplo, uma consulta a um catálogo de biblioteca geralmente envolve um campo de busca e, às vezes, alguns parâmetros (como os campos onde a busca será realizada). Em outros casos, esses parâmetros aparecem somente na busca

avançada. Após a realização da busca, é esperada uma lista de resultados (ordenados por algum critério, que também pode ser observado) que apresenta algumas informações básicas de cada documento, como título, autor, data etc. Este é um exemplo de interação com o sistema, do qual se pode apreender as possibilidades que o sistema oferece, o que pode ser solicitado e o que ele devolve da requisição.

Por exemplo, Monteiro-Krebs (2013) analisou o uso de sistemas de recomendação, selecionando as variáveis de uso ou não da recomendação feita pelo sistema, quantidade de documentos visualizados, tipos de recomendação usados, tempo de sessão, data, IP do usuário. Observa-se que, nesse caso, as expressões de busca não foram coletadas pois não eram relevantes para o objetivo da pesquisa. Se, por outro lado, o objetivo da pesquisa fosse examinar as tendências temáticas de um domínio, seria recomendado considerar a análise dos termos de busca, estabelecendo o ponto de corte a partir da saturação dos dados (LAIPELT; MONTEIRO-KREBS, 2021).

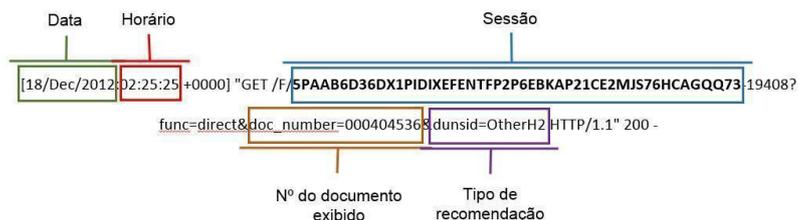
Foi o caso da pesquisa de Laipelt (2015), que coletou *logs* do Portal LexML referentes a um período de 15 dias. A partir da aplicação de um filtro, a pesquisadora selecionou os parâmetros relevantes para a sua pesquisa, que foram IP, data, horário, expressão de busca e tipo de documento (Legislação, Doutrina e Jurisprudência). Além disso, usou para seleção de seus *corpora* de estudo os critérios de frequência e funcionalidade (LAIPELT; MONTEIRO-KREBS, 2021). Como observa-se nesse caso, se o intuito da pesquisa é estudar a linguagem, as expressões de busca dos usuários são mais relevantes do que o tempo da sessão.

Após identificar as interações relevantes ao estudo, o passo seguinte é verificar como essas interações são representadas no *log*, a partir da análise de uma pequena amostra de registros. O *log* de interações de sistemas de recuperação da informação tipicamente oferece informações como quem fez a consulta (IP), o que foi procurado (expressão de busca

e filtros utilizados – se aplicável), o que obteve como resposta (lista de resultados), qual(is) documento(s) escolheu(ram)-se para ver em detalhes (item(ns) clicado(s)/visualizado(s)) e que ações realizou (apenas abriu o registro, fez *download* do documento, realizou uma reserva etc.).

A interpretação do *log* envolve a identificação do que cada registro de *log* informa, porque cada sistema possui uma estrutura específica de informações, que pode variar bastante. Cada registro de *log* é composto por informações que representam uma interação do usuário com o sistema, na forma de requisição. Os elementos que compõem uma requisição são chamados de parâmetros, e cada parâmetro pode ser utilizado como uma variável na pesquisa. Na Figura 1, apresentamos a decodificação de um registro de requisição do trabalho de Monteiro-Krebs (2013):

Figura 1 – Exemplo de registro do log decodificado



Fonte: Monteiro-Krebs (2013, p. 65) sistematizado pela autora.

A partir da decodificação representada na Figura 1, pode-se identificar a ação do usuário: ocorre no dia 18 de dezembro de 2012, às 02h25min25seg, o usuário (IP oculto), dentro da sessão de número 5PAAB6[...]QQ73, acessou o registro do documento 0004045368, que foi originado de uma recomendação do tipo OtherH2 (que significa ligação por classificação, ou seja, o sistema está recomendando outras obras sobre o mesmo assunto).

Já na Figura 2, o *log* analisado por Nicholas, Rowlands e Jamali (2009) registra informações diferentes do exemplo anterior. Os dois primeiros campos mostram data e hora de quando o arquivo foi enviado pelo servidor ao computador do usuário. O próximo campo fornece a ação (*download*) e descreve o tipo de página; aqui */browse/open.asp* refere-se à página inicial. O campo a seguir é o número do IP: “Mozilla / 5.0 + (Windows; + U; + Windows + NT + 6.0; + fr; + rv: 1.8.1.11) + Gecko/20071127 + Firefox/2.0.0.11” identifica o tipo de navegador da máquina cliente. O campo <http://www.sussex.ac.uk/library/resources/e-books.php> é o campo de referência e fornece os detalhes do site e da página da página anterior visualizada pelo usuário. Nesse caso, esta era uma página de recursos de e-book ou links na Universidade de Sussex (NICHOLAS; ROWLANDS; JAMALI, 2010).

**Figura 2 – Exemplo de registro do *log* decodificado**

2007-12-01	04:33:38	GET	/browse/open.asp	-	139.184.30.131	HTTP/1.0
Mozilla/5.0+(Windows;+U;+Windows+NT+6.0;+fr;+rv:1.8.1.11)+Gecko/20071127						
+Firefox/2.0.0.11						
<a href="http://www.sussex.ac.uk/library/resources/e-books.php">http://www.sussex.ac.uk/library/resources/e-books.php</a>						
<a href="http://www.mylibrary.com">www.mylibrary.com</a>						

Fonte: Nicholas, Rowlands e Jamali, 2010, p. 4.

Uma questão importante a ser analisada é que *logs* podem não registrar informações que permitam a identificação do encadeamento entre as interações, impossibilitando análises “relacionais”. Por exemplo, pode ser de interesse do pesquisador investigar a quantidade de itens resultantes de uma consulta que são realmente visitados pelo usuário. A análise dessa relação envolve dois tipos de requisições, que estão inter-relacionadas: a requisição da consulta e as requisições das visualizações dos itens da consulta. Essa análise somente é possível se, no registro da requisição de visualização de item, houver uma informação que permita a identificação da requisição que gerou a respectiva consulta. Por exemplo, o registro de *log* da Figura 1 contém a informação que diz que a requisição de visualização de um item foi originada de uma interação prévia de recomendação, mas não contém informação que permita

identificar a recomendação de origem. O pesquisador, portanto, precisa ter em mente o que cada *log* pode oferecer (e o que não pode) em termos informacionais, para que não incorra em conclusões imprecisas.

#### 4.2.2 Recorte dos Dados

Na etapa de seleção, são definidos os **períodos de interações** relevantes para a análise, isto é, os recortes temporais das interações que serão coletadas. Como os *logs* registram todas as interações, não é necessário se pensar em estratificar uma amostra para fins de representatividade, ou seja, o pesquisador poderá ter acesso a todos os registros de todos os usuários. Por outro lado, uma característica do uso de *logs* é o grande volume de dados aos quais se tem acesso. Esse fator pode dificultar a pesquisa, já que pode sobrecarregar o pesquisador com a quantidade de dados a ser manipulada. Assim, é recomendável que se considere a aplicação de algum filtro de seleção dos dados de acordo com os objetivos da pesquisa – escolhendo quais interações se deseja observar.

O volume de dados disponível normalmente é muito grande, e os dados exigem limpeza e pré-processamento antes da análise. Por isso, é desejável que se observe a representatividade dos dados levando em conta a otimização dos recursos disponíveis para análise destes. Uma forma possível de recortar uma amostra representativa de um sistema é, por exemplo, determinar um **recorte estatístico** a partir do universo total de acessos ao sistema. Isso pode ser feito através do acompanhamento de acessos mensais por usuários únicos. Através de consulta prévia diretamente no arquivo de *log*, é possível recuperar a quantidade de IPs diferentes que acessaram o sistema no mês. Acompanhando-se esse indicador, pode-se calcular a média mensal de usuários dos serviços e, assim, determinar a amostra suficientemente representativa de acessos que devem compor o *corpus*.

#### 4.2.3 Disponibilização e uso dos dados

Tendo conhecimento das possibilidades do estudo a partir da experimentação do ambiente e interpretação do *log*, deve-se verificar a viabilidade do uso deste. Nesse momento, o responsável pela guarda do *log* deve ser contatado para identificar se o *log* registra as interações anotadas no levantamento de forma suficiente e adequada para a análise. Muitos ambientes podem optar por não registrar todas as interações ou guardar o registro por um período limitado.

Outra questão relevante na negociação é a presença no *log* de **informações sensíveis**. Em geral, dados pessoais dos usuários – aqueles que permitem sua identificação – são considerados sensíveis, como, por exemplo, o IP de acesso ao sistema. Pela natureza dessas informações, é recomendado que projetos de pesquisa que utilizem dados sensíveis em *logs* passem pela avaliação de um Comitê de Ética, preferencialmente indicando a anonimização dos participantes da pesquisa.

Por questão de privacidade dos usuários, é recomendável que seja feito um tratamento no endereço de IP do computador originário das interações. Os três primeiros algarismos do IP identificam o país de origem do acesso. Se for relevante para a pesquisa regionalizar os usuários, essa informação pode ser mantida, mas os demais algarismos devem ser embaralhados (com o mesmo padrão, para que se identifique quais interações são de cada usuário) ou substituídos por letras. Se a regionalização não for relevante ou mesmo a identificação de usuários únicos, o IP pode ser simplesmente omitido. Outra opção é fazer esse tratamento após a etapa de análise e antes da divulgação dos resultados da pesquisa, de acordo com o protocolo do projeto aprovado previamente pelo Comitê de Ética (LAPELT; MONTEIRO-KREBS, 2021).

Em sistemas que exigem *login*, é possível obter ainda mais informações, além das presentes nos *logs*, como nome, endereço, idade, gênero, profissão etc. No entanto, cada instituição estabelece

critérios próprios para definição do que entende como informação sensível. Assim, é necessário consultar o responsável pela guarda do *log* para entender quais dados poderão ser disponibilizados obedecendo a política de privacidade de dados da instituição. Outro ponto de atenção é a adequação à Lei Geral de Proteção de Dados (LGPD) (BRASIL, 2018), que entrou em vigor no Brasil em 2020. Ao lidar com dados pessoais dos usuários, pesquisadores precisam atentar para o fato de que, com o marco legal, a coleta e o uso de dados devem ser feitos sempre com consentimento explícito dos usuários. Assim, a LGPD (Lei 13.709, de 2018) garante maior controle dos cidadãos sobre suas informações pessoais.

Caso existam informações sensíveis a serem retiradas do *log* antes de sua disponibilização, deve-se analisar os impactos dessa retirada ao estudo e possíveis alternativas para contornar essas limitações. Se os objetivos da pesquisa demandarem que seja possível agrupar todas as interações de um mesmo usuário (como estudos de análise da jornada do usuário), não se pode abrir mão do uso do IP (em casos de sistemas sem login mandatório). Nesse caso, o embaralhamento dos IPs de acesso ao sistema é recomendado, que pode ser feito pela própria instituição antes de sua disponibilização. Dessa forma, garante-se a privacidade dos usuários sem comprometer a pesquisa.

Em resumo, nessa fase o pesquisador determina quais os parâmetros ou variáveis que deseja analisar, e esse planejamento é crucial para evitar retrabalho e garantir a viabilidade da pesquisa de forma legal e ética (LAIPÉLT, MONTEIRO-KREBS, 2021). Os parâmetros precisam dar conta do objetivo do estudo em questão, tendo em vista que cada sistema de informação possui critérios próprios para organização dos *logs*, para determinação do(s) período(s) de guarda dos arquivos e políticas de compartilhamento (ou não) desses dados.

### 4.3 COLETA E PREPARAÇÃO DOS DADOS

Após a verificação das viabilidades do uso dos dados e a identificação dos dados a serem colhidos, a próxima etapa é a recolha dos dados de interação de um determinado período em um *log* de transações, isto é, a coleta dos dados.

Na coleta, as informações são preparadas pela equipe que gerencia o sistema aplicativo para serem entregues ao pesquisador, de acordo com o que foi previamente especificado, com a remoção de informações sensíveis e de informações não pertinentes à pesquisa, através de comando no servidor. Então o pesquisador passa a preparar esses dados.

Depois que os dados são coletados, passa-se para o estágio de preparação de dados do processo TLA. Para a preparação de dados, o foco é importar os dados do registro de transações para um banco de dados relacional (ou outro *software* de análise), atribuindo a cada registro uma chave primária, limpando os dados (ou seja, verificando dados incorretos em cada campo) e calculando métricas de interação padrão que servirá de base para uma análise mais aprofundada (JANSEN, 2006, p. 414, tradução nossa).

Na etapa de preparação dos dados, o pesquisador realiza a limpeza e organização (ou tratamento) dos dados obtidos na fase de coleta. Procede-se, então com a normalização, isto é, a adaptação do formato de registro de *log* (normalmente em formato texto) para uma estrutura adequada para a análise (que pode ser texto, no caso de análise exclusiva de expressões de busca, e/ou tabela no caso de análises que envolvam outras variáveis).

Geralmente, os *logs* são recebidos pelo pesquisador no formato texto (.txt) e, para adaptá-los a um formato adequado à pesquisa, é necessário que estes sejam exportados para uma planilha (.csv ou

.xls). Assim, recomenda-se usar um *software* que realize as funções de acordo com o que o pesquisador deseja investigar. São exemplos: eliminar as URLs repetidas do *log*; ordenar cronologicamente as URLs; contar o intervalo de tempo entre o primeiro e o último registro de uma sessão (em horas, minutos ou segundos); exportar os dados para uma tabela (formato .csv ou .xls) organizando-os em colunas por tipo de informação; entre outras ações. É possível encomendar o desenvolvimento de um *software* a um profissional da área de tecnologia da informação, conforme a parametrização da base de dados que registra o *log*, como feito por Monteiro-Krebs (2013) e Laipelt (2015)<sup>52</sup>. Outra opção é utilizar *softwares* de mercado disponíveis para análise de *logs*, como o Graylog, Logstash, Loggly e Splunk.

A limpeza consiste em eliminar registros corrompidos, que são passíveis de identificação através da ordenação sequencial dos campos, assim, registros fora do formato padrão dos dados de cada campo ficarão agrupados (JANSEN, 2006). Desse modo, fica mais rápida a eliminação dos erros em grandes volumes de dados, em que a identificação visual é inviável. “As funções padrão de banco de dados para somar e agrupar campos-chave, como horário e endereço IP, geralmente identificam quaisquer erros adicionais” (JANSEN, 2006, p. 415, tradução nossa). Além disso, na limpeza, podem ser eliminados registros de agentes não humanos (robôs) se o objetivo da pesquisa for analisar exclusivamente interações entre humanos e o sistema. Infelizmente, não há como determinar com absoluta certeza quais interações são realizadas por humanos ou robôs (SILVERSTEIN *et al.*, 1999; JANSEN, 2006). Assim, é necessário pensar em uma técnica para estabelecer um ponto de corte no *corpus*. Jansen (2006), buscando especificamente as interações de humanos com o sistema, determinou arbitrariamente como ponto de corte a quantidade de 101 ou mais consultas. Pode-se, alternativamente, construir um histograma,

<sup>52</sup> Agradecemos a Vicente Grassi-Filho pelo desenvolvimento do *software* extrator de *logs* para o formato texto, já utilizado em diferentes projetos do grupo de pesquisa Orcalab (UFRGS).

com base no tempo de sessão, e eliminar os “*outliers*” (tanto sessões muito grandes quanto aquelas muito pequenas ou sem consulta), que podem distorcer as métricas de análise.

Laipelt e Monteiro-Krebs (2021, p. 133) explicam que “Com a decodificação das interações a partir dos parâmetros é possível separar os registros de acordo com o tipo de interação do usuário.” Assim, os arquivos extraídos podem ser importados para tabelas de um banco de dados como o Microsoft Access ou MySQL. A partir de então, podem ser realizadas consultas ao banco de dados, através da linguagem SQL. As consultas resultarão em tabelas de planilhas eletrônicas (Por exemplo: Microsoft Excel) em que as linhas normalmente representam requisições e as colunas representam os parâmetros/variáveis. Essas tabelas podem, então, ser analisadas diretamente e/ou serem transformadas em gráficos, servindo de base para a análise de dados.

#### 4.4 ANÁLISE DOS DADOS

Na etapa de análise, as tabelas e gráficos gerados são comparados e interpretados à luz dos conhecimentos teóricos concernentes à pesquisa. Jansen (2006) traz diversos exemplos de *queries* SQL para análise de *logs*. A partir dessas consultas, pode-se obter muitos resultados estatísticos que servirão de indicadores para o gestor do sistema, como média de consultas realizadas, média de páginas de resultados visualizadas, frequência e coocorrência de termos, grau de associação de pares de termos, e muitos outros.

Essas descobertas servem tanto para melhorar a arquitetura de informação do sistema quanto para planejar capacitações aos usuários, com o intuito de contribuir para que os sistemas possam ser planejados e avaliados a partir das necessidades reais destes. Apresentamos a seguir, com base nos estudos recuperados para a

redação desse capítulo, algumas alternativas de análise das interações dos usuários nos sistemas de recuperação da informação a partir dos dados registrados em *log*.

Segundo Jansen (2006, p. 410, tradução nossa), a análise de *logs* vem sendo utilizada como método “[...] para avaliar sistemas de bibliotecas, sistemas de recuperação de informação (IR) e, mais recentemente, sistemas da Web.”. Para tanto, o autor aponta três possibilidades de análise: análise no nível do termo, análise no nível da consulta e análise no nível da sessão (JANSEN, 2006, p. 418). Gouveia (2013) afirma que, assim como o *page tagging*, o método é utilizado nos estudos de métricas de acesso na Web, também designados por webmetria (*Webmetrics* ou *Web Metrics*, em inglês). A análise de *logs* pode ser utilizada para diversos fins no âmbito da Ciência da Informação. O método mostra-se útil para validar sistemas, compreender como os usuários se comportam e o que de fato esperam dos recursos oferecidos pela instituição etc.

Através de um estudo de *logs*, Foust *et al.* (2007) descobriram que 90% das pesquisas realizadas em um sistema de busca por livros eletrônicos (*e-books*) configuraram-se como palavras-chave ou frases. Apenas 10% utilizaram recursos avançados de pesquisa, com construções complexas, operadores booleanos, frases de busca estruturadas e tentativas de pesquisa de títulos de livros ou autores. Os resultados desse estudo sugerem que a maioria dos usuários não esperava recursos avançados de pesquisa e compreendia com facilidade o tipo de pesquisa no estilo “Google” no sistema de pesquisa de livros eletrônicos analisado. Ou seja, os usuários aparentemente não queriam ter que articular recursos complicados de filtragem, concatenação de operadores lógicos etc., mas esperam, cada vez mais, que o sistema faça isso por eles.

A seguir, elencamos algumas possibilidades de descobertas com a análise de *logs* no campo da CI, a partir de parâmetros que comumente estão presentes nos registros de sistemas de informação:

- a. **Análises dos acessos ao sistema:** através do IP, pode-se identificar a região de onde o sistema é acessado, possibilitando a geração de relatórios regionais (conforme mencionado na seção 4.1.4); também é possível verificar os dias de maior ou menor acesso (no mês ou na semana) através da análise das variáveis *data* e *horário* no *log*; assim como verificar quantos pesquisadores acessaram o mecanismo de pesquisa durante determinado período. Para responder a essa questão, pode-se fazer uma consulta SQL que provê “uma lista de usuários únicos e o número de consultas que eles enviaram durante o período” (JANSEN, 2006, p. 420, tradução nossa)<sup>53</sup>.
- b. **Duração e comprimento das sessões:** a variável *horário* também possibilita identificar a duração da sessão e quanto tempo em média os usuários levam para concluir uma ação (seja visualizar um documento, fazer uma reserva ou renovação, realizar o *download* de um documento etc.). Isso pode ser medido ao se subtrair o *horário final* (última interação) do *horário inicial* (primeira interação). Assim, é possível identificar se alguma ação específica influencia o tempo das sessões (se a sessão será, em média, mais longa ou curta do que sem a realização dessa ação). Monteiro-Krebs (2013) analisou a duração de sessões com e sem o uso de recomendação de obras no catálogo de uma biblioteca universitária, concluindo que o uso de recomendação implica acesso otimizado (visualização de mais registros em menos tempo); adicionalmente, pode-se

53 Jansen (2006, p. 426) exemplifica essa consulta através da SQL Query 03 (*qry\_03\_list\_of\_unique\_ips*):  
“SELECT tbl\_searching\_episodes.uid, Count(tbl\_searching\_episodes.search\_url)  
AS CountOfsearch\_url  
FROM tbl\_searching\_episodes  
GROUP BY tbl\_searching\_episodes.uid  
ORDER BY Count(tbl\_searching\_episodes.search\_url) DESC”.

identificar o comprimento das sessões, realizando uma consulta pela quantidade de pesquisas realizadas por cada usuário<sup>54</sup>.

- c. **Visualização, leitura ou *download* de documentos:** esse parâmetro pode medir sozinho ou cruzado com outras variáveis a quantidade de acessos aos registros de documentos e, em caso de obras disponíveis para acesso ao texto integral, quantas leituras do texto ou mesmo *downloads* foram feitos pelo usuário. Um exemplo desse tipo de estudo é o de Nicholas, Rowlands e Jamali (2010), que empregaram três métricas a fim de obter estimativas mais precisas de “uso” de *e-books*: número de páginas visualizadas, número de sessões conduzidas e quantidade de tempo gasto on-line. Esse tipo de estudo é caracterizado como DLA, pois, além dos *logs*, utiliza questionário, dados de circulação da biblioteca e grupo focal.
- d. **Dados demográficos e de pesquisa:** em um sistema que pede *login*, no qual se tem acesso aos dados demográficos do usuário, essas informações podem ser cruzadas com todas as demais variáveis identificadas no *log*. Conforme explicado na introdução desse capítulo, estudos que cruzam os dados de *login* do usuário com dados obtidos no *log* (variáveis acima listadas), chamam-se *deep log analysis* (DLA) (NICHOLAS *et al.*, 2006a; WARWICK *et al.*, 2008; NICHOLAS *et al.*, 2009; NICHOLAS *et al.*, 2006b). Uma possibilidade de uso da DLA é cruzar estratégias de busca (no *log*) com dados demográficos para identificar padrões por perfil de usuário (por área de conhecimento, nível de especialização, tipo de vínculo com a instituição etc.).

54 Jansen (2006, p. 426) exemplifica essa ação com a SQL Query 05 (qry\_05\_session\_length):  
“SELECT qry\_03\_list\_of\_unique\_ips.CountOfsearch\_url, Count(qry\_03\_list\_of\_unique\_ips.CountOfsearch\_url) AS CountOfCountOfsearch\_url  
FROM qry\_03\_list\_of\_unique\_ips  
GROUP BY qry\_03\_list\_of\_unique\_ips.CountOfsearch\_url  
ORDER BY Count(qry\_03\_list\_of\_unique\_ips.CountOfsearch\_url) DESC”.

- e. **Comportamento do usuário:** através do IP, pode-se agrupar todas as interações de um usuário com o sistema. Pelas variáveis data e horário, é possível ordenar cronologicamente todas as ações de um único usuário. Quando não se tem acesso ao IP, a sessão pode exercer esse papel, porém, tem limitações. A primeira limitação do uso de sessão no lugar do IP é que não é possível afirmar que uma sessão seja de um único usuário ou um acesso de vários usuários em um mesmo computador (por exemplo, quando o acesso é realizado a partir de locais públicos como bibliotecas e *lan houses*). A segunda é que, apenas com o dado de sessão, não se pode determinar quais sessões foram realizadas por um mesmo usuário (podem ser várias) e, por isso, não é possível agrupar todas as sessões de um mesmo usuário e visualizar seu comportamento informacional na íntegra. Pode-se também realizar uma análise de exibição de página, também conhecida como análise de cliques, que, de acordo com Jansen (2006, p. 420), serve para medir o comportamento de visualizações de página dos usuários da Web. Para tanto, mensura-se a duração da visualização do documento desde o momento em que o usuário clica em uma URL em uma página de resultados até o momento em que ele retorna ao mecanismo de pesquisa. Se o pesquisador tiver interesse em conhecer a linguagem do usuário, pode fazê-lo através da análise das expressões de busca. Estudos dessa natureza visam melhorar a indexação (remissivas, variantes denominativas, variantes morfológicas – numeral, gênero –, alterações gramaticais etc.).
- f. **Estratégias de busca:** visa perceber padrões de navegação (usuários mais ou menos avançados); jornada do usuário (caminhos percorridos no site, ou as diferentes tentativas de recuperação da informação através da análise das expressões de busca); ver se o usuário utiliza todos os recursos disponíveis no sistema de recuperação da informação (SRI), como filtros,

operadores booleanos, comprimento da expressão de busca; identificar termos simples e compostos. Jansen e Spink (2006) fazem uma análise comparativa de nove estudos do tipo TLA para descobrir características e mudanças nos padrões de busca na Web (através dos *logs* de buscadores). A análise realizada por Laipelt (2015, p. 160-161), por sua vez, permitiu inferir características do perfil do usuário através de sua jornada: agrupando todas as interações com o sistema, pode-se verificar o nível de especialização do usuário a partir do seu domínio sobre a terminologia da área do direito.

- g. **Uso do acervo:** pode-se analisar as expressões de busca para encontrar os títulos mais buscados. Um exemplo de estudo para identificar obras buscadas com maior ou menor frequência é o de Warwick *et al.* (2008), no qual os pesquisadores cruzaram essa informação com um questionário de percepção sobre uso de recursos digitais (usando método DLA). Além disso, é possível cruzar uma obra mais buscada (no *log*) com os períodos em que esta estava emprestada (no catálogo), verificando assim obras não retiradas ou reservadas por indisponibilidade (se a busca pela obra poderia ter se efetivado em retirada, caso o item estivesse disponível e sem fila de reserva).
- h. **Tendências temáticas:** a partir da identificação dos assuntos mais procurados, é possível identificar se o acervo está atendendo ou não as necessidades dos usuários, podendo suprir eventual demanda reprimida. A demanda reprimida também poderia ser identificada através da lista de reservas no catálogo, porém, esse indicador somente informa a quantidade insuficiente de itens já existentes no acervo, enquanto no *log* se amplia a visão, possibilitando identificar novos itens para aquisição que contemplem essas necessidades.

- i. **Penetração no site:** a partir da quantidade de itens ou páginas visualizadas, é possível identificar o comportamento de busca de informações dos usuários. Identificando quantos cliques o usuário precisa dar para acessar a informação desejada e os caminhos percorridos por ele (páginas visitadas), é possível avaliar criticamente a usabilidade do site, e implementar melhorias para poupar o tempo do leitor. A penetração do site pode ser combinada com o número de visitas, tempo de permanência no site, tipo de itens e conteúdo visualizados para indicar características do perfil do usuário e suas demandas.

## 5 CONSIDERAÇÕES FINAIS

Com a disponibilidade cada vez maior de informação através de dispositivos diversos, os sistemas de informação têm evoluído de forma gradual à medida que os usuários exigem maior precisão e personalização em seus resultados de busca. Com o objetivo de munir pesquisadores e profissionais da informação com recursos de investigação atuais e muitas vezes disponíveis dentro da própria unidade de informação, nesse capítulo discorreu-se sobre a análise de *logs* como uma alternativa para identificar comportamentos de usuários e antecipar suas demandas.

As etapas metodológicas foram apresentadas e exemplificadas, também foram expostas diversas possibilidades de uso dos parâmetros do *log* tanto separadamente quanto de forma conjunta (cruzando as variáveis). Buscou-se, ainda, apresentar a análise de *logs* para explorar as potencialidades das expressões de busca para identificação de tendências temáticas e diferenças de perfis entre os usuários.

Apesar de a obtenção de um *log* para análise não ser uma tarefa das mais simples, por envolver uma série de questões técnicas e até

políticas comentadas nesse trabalho, acredita-se que o esforço é recompensado com a riqueza de *insights* que se pode obter com esse recurso. Isso foi possível demonstrar nesse capítulo, considerando as práticas consolidadas na área da Ciência da Informação e a expertise dos autores na realização de pesquisas envolvendo o uso e a análise de *logs*.

## 6 REFERÊNCIAS

AIRES, R. V. N. X.; ALUÍSIO, S. M. Como incrementar a qualidade dos resultados das máquinas de busca: da análise de *logs* à interação em português. **Ciência da Informação**, v. 32, n. 1, p. 5-16, 2003. Disponível em: <http://www.brapci.ufpr.br/brapci/v/a/689>. Acesso em: 27 Jul 2017.

BRASIL. **Lei nº 13.709**, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Presidência da República; 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm). Acesso em: 22 nov. 2021.

CETIC. **Pesquisa sobre o uso das Tecnologias de Informação e Comunicação nos domicílios brasileiros - TIC Domicílios 2020**. S. l.: Cetic.br, 25 nov. 2021. Disponível em: [https://cetic.br/media/docs/publicacoes/2/2011124201505/resumo\\_executivo\\_tic\\_domicilios\\_2020\\_.pdf](https://cetic.br/media/docs/publicacoes/2/2011124201505/resumo_executivo_tic_domicilios_2020_.pdf). Acesso em: 10 dez. 2021.

CHEN, LS. Applying swarm intelligence to a library system. **Library Collections, Acquisitions, and Technical Services**, v. 34, n. 1, p. 1-10, 2010. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/14649055.2010.10766254>. Acesso em: 20 ago. 2018.

DAVIS, P. M. Information-seeking behavior of chemists: A transaction log analysis of referral URLs. **JASIST**, v. 55(4), 326–332, 2004. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.10384>. Acesso em: 21 dez. 2018.

FOUST, J. E. *et al.* Improving e-book access via a library-developed full-text search tool. **Journal of the Medical Library Association**, Bethesda, v. 95, n. 1, p. 40-45, 2007. Disponível em: <https://www.ncbi-nlm-nih-gov.ez45.periodicos.capes.gov.br/pmc/articles/PMC1773047/>. Acesso em: 02 jul. 2018.

GOUVEIA, F. C.. Altmetria: métricas de produção científica para além das citações. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 214-227, maio

2013 – Disponível em: <http://revista.ibict.br/liinc/article/view/3434/3004>. Acesso em: 27 jul. 2017.

HUYNH, T.; MILLER, J. Empirical observations on the session timeout threshold. **Information Processing & Management**, v. 45, n. 5, p. 513-528. 2009.

JANSEN, B. J. Search log analysis: what it is, what's been done, how to do it. **Library & Information Science Research**, Pennsylvania, v. 28, p. 407-432, 2006. Disponível em: <http://lincs.hum.iit.edu/sites/default/files/JansenSearchLog.pdf>. Acesso em: 03 maio 2013.

JANSEN, B. J., & POOCH, U. A review of Web searching studies and a framework for future research. **Journal of the American Society for Information Science and Technology (JASIST)**, v. 52, p. 235–246, 2001.

JANSEN, B. J.; TAKSA, I.; SPINK, A. Research and Methodological Foundations of Transaction Log Analysis. *In*: JANSEN, Bernard J.; SPINK, Amanda; TAKSA, Isak. **Handbook of web log analysis**. Hershey, PA: Information Science Reference, 2009. p. 1-16.

JANSEN, B.J.; SPINK, A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. **Information Processing & Management**, v. 42, n. 1, p. 248-263, jan. 2006.

JANSEN, BJ.; SPINK, A.; Blakely, C.; Koshman, S. Defining a session on web search engines. **Journal of the American Society for Information Science and Technology (JASIST)**, v. 58, n. 6, p. 862-871, Apr. 2007.

LAIPELT, R. C. F. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, Porto Alegre, v. 21, n. 3, p. 150-170, set/ dez. 2015. Disponível em: <http://www.redalyc.org/html/4656/465645968009/>. Acesso 16 jul. 2017.

LAIPELT, R. C. F.; MONTEIRO-KREBS, L. **Termos sob a superfície**: elementos teóricos, metodológicos e terminológicos para a Representação do Conhecimento. Rio de Janeiro: Editora Interciência, 2021.

MENG-XING, H.; CHUN-XIAO, X.; YONG, Z.. Supply chain management model for digital libraries. **The Electronic Library**, v. 28, n. 1, p. 29-37, 2010. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/02640471011023351>. Acesso em: 20 ago. 2018.

MONTEIRO-KREBS, L. **Sistema de recomendação para bibliotecas universitárias**. 2013. 95f. Trabalho de Conclusão de Curso (Graduação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <http://www.lume.ufrgs.br/handle/10183/78367>. Acesso em: 26 jul. 2017.

MONTEIRO-KREBS, L.; ROCHA, R. P. da; RIBEIRO, C. Quem leu este também leu...: sistema de recomendação na biblioteca universitária. **Perspectivas em Ciência da Informação**, v. 22, n. 1, p. 151-169, mar. 2017. DOI: 10.1590/1981-5344/2496. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2496>. Acesso em: 17 abr. 2019.

NICHOLAS, D. *et al.* Characterising and evaluating information seeking behaviour in a digital environment: Spotlight on the 'bouncer'. *Information Processing & Management*, v. 43, n. 4, p. 1085-1102, 2007. Disponível em: <https://doi.org/10.1002/asi.20564>. Acesso em: 22 dez. 2021.

NICHOLAS, D. *et al.* The information seeking behaviour of the users of digital scholarly journals. **Information Processing & Management**, v. 42, n. 5, p. 1345-1365, September 2006a. Disponível em: <https://doi.org/10.1016/j.ipm.2006.02.001>. Acesso em: 22 dez. 2021.

Nicholas, D. *et al.* Online use and information seeking behaviour: institutional and subject comparisons of UK researchers. **Journal of Information Science**, v. 35, n. 6, p. 660-676, 2009.

NICHOLAS, D.; HUNTINGTON, P.; WATKINSON, A. Scholarly journal usage: the results of deep log analysis. **Journal of Documentation**, London, v. 61, n. 2, p. 248-280, 2005. Disponível em: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/00220410510585214>. Acesso em: 26 jul. 2017.

NICHOLAS, D.; HUNTINGTON, P.; WATKINSON, A. Revisiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication. **Information Processing & Management**, v. 41, n. 6, p. 1441-1461, dez. 2005. Disponível em: <https://doi.org/10.1016/j.ipm.2005.03.014>. Acesso em: 20 ago. 2018.

NICHOLAS, D.; ROWLANDS, I.; JAMALI, H.R.. E-textbook use, information seeking behaviour and its impact: Case study business and management. **Journal Of Information Science**, v. 36, n. 2, p. 263-280, abr. 2010.

NICHOLAS, D.; ROWLANDS, I.; JAMALI, H.R.. E-textbook use, information seeking behaviour and its impact: Case study business and management. **Journal of Information Science**, v. 36, n. 2, p. 263-280, abr. 2010.

NICHOLAS, David *et al.* What deep log analysis tells us about the impact of big deals: case study OhioLINK. **Journal of Documentation**, v. 62, n. 4, p. 482-508, 2006b. Disponível em: <https://doi.org/10.1108/00220410610673864>. Acesso em: 22 dez. 2021.

PETERS, T.A.. The history and development of transaction log analysis. **Library Hi Tech**, v. 11, n. 2, p. 41-58, 1993.

SILVERSTEIN, C. *et al.*. Analysis of a very large Web search engine query log. **SIGIR Forum**, v. 33, n. 1, p. 6-12, 1999.

SPINK, A., JANSEN, B. J.. **Web search**: public searching of the web. New York: Kluwer, 2005. 198 p.

VAKKARI, P.; TALJA, S.. Searching for electronic journal articles to support academic tasks. A case study of the use of the Finnish National Electronic Library (FinELib). **Information Research**, v. 12, n. 1, 2006. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/02640471011023351>. Acesso em 20 ago. 2018.

WARWICK, C. *et al.*. Library and information resources and users of digital resources in the humanities. **Program-electronic Library And Information Systems**, v. 42, n. 1, p. 5-27, 2008.

# 7

Renato Tarciso Barbosa de Sousa

**A classificação automática  
de documentos de arquivo  
é uma solução para os problemas  
que os usuários encontram  
com a classificação funcional?**

**Algumas reflexões e caminhos a percorrer**

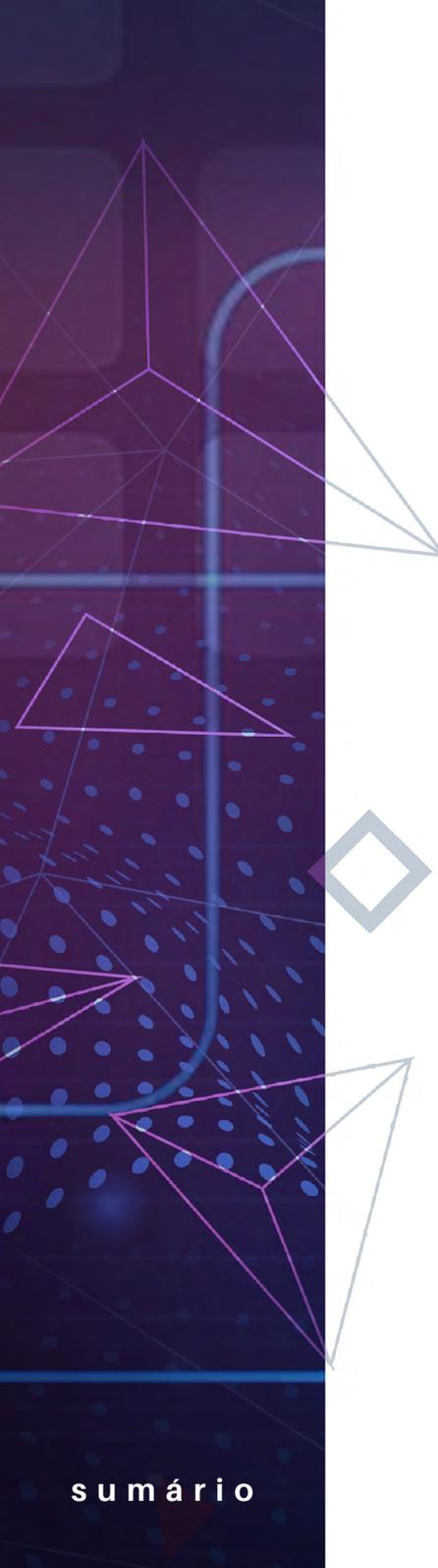
*Is automatic classification of records  
a solution to the problems users  
encounter with functional classification?*

*Some thoughts and ways forward*

### Resumo:

A classificação funcional de documentos de arquivo está pacificada na literatura da área, apesar de sua aplicação prática ter se tornado comum a partir da década de 1990. Estudos de caso, que começaram a surgir a partir dos anos 2000, têm revelado sistemas híbridos que misturam função, estrutura, assunto, tipo e espécie documental. O objetivo deste capítulo é identificar as possibilidades de inserir, nos sistemas informatizados de gestão de documentos, a classificação automática de documentos de arquivo. Quanto à natureza da pesquisa, ela é do tipo aplicada e qualitativa, em relação à abordagem do problema, e exploratória e descritiva, no tocante aos seus objetivos, pois o tema classificação automática ainda está em fase preliminar de desenvolvimento e porque torna claro um problema. A técnica de coleta de dados é a pesquisa bibliográfica relacionada à classificação de documentos de arquivo, à abordagem funcional, às questões de usabilidade do instrumento de classificação e a classificação automática de documentos. As primeiras experiências de uso da classificação automática apontam para o uso da *deep learning* com base na inteligência artificial. Essas possibilidades precisam da junção dos conhecimentos em tecnologia da informação e em arquivologia.

**Palavras-Chave:** Classificação de documentos de arquivo; classificação funcional; classificação automática.



**Abstract:**

*The functional classification of records is pacified in the literature of the area, although its practical application has become common since the 1990s. Case studies, which began to emerge from the 2000s, have revealed hybrid systems that mix function, structure, subject, type and document species. The objective is to identify the possibilities of inserting automatic classification of archive documents into computerized document management systems. As for the nature of the research it is of the applied and qualitative type in relation to the approach to the problem and exploratory and descriptive in relation to its objectives, because the automatic classification theme is still in a preliminary phase of development and because it makes a problem clear. The data collection technique is literature research related to the classification of archival documents, the functional approach, the usability issues of the classification instrument, and the automatic classification of documents. The first experiences of using automatic classification point to the use of deep leaning based on artificial intelligence. These possibilities need the pooling of knowledge in information technology and archival science.*

**Keywords:** *Classification of archival documents; functional classification; automatic classification.*

## 1 INTRODUÇÃO

Milton Santos, um dos principais geógrafos brasileiros, em sua obra mais conhecida chamada *Por uma geografia nova* (1978), faz a seguinte afirmação:

O apego às velhas ideias parece uma enfermidade incurável. (...) Caímos naquele defeito de considerar velhas formas de pensar como inevitáveis. Ao invés de perseguir um saber novo, preferimos deliciar-nos com a reprodução do saber velho. (...) O novo é, de certa forma, o desconhecido e só pode ser conceitualizado com imaginação, e não com certezas (SANTOS, 1978, p. 110).

Em 2015, Reed deu um título interessante ao artigo publicado na revista *Acervo: Fazer o mesmo não vai funcionar*, em que ela discute a passagem, na Austrália, dos sistemas de *recordkeeping* manuais para os digitais. O mundo digital tem demonstrado que fazer mais do mesmo não servirá para avançarmos. Tanto a citação quanto o título revelam a necessidade de reinvenção, de inovação, de monitoramento contínuo e de um repensar constante da nossa teoria e da nossa prática.

Os planos de classificação baseados na função ou planos de classificação funcionais começam a aparecer no cenário arquivístico desde meados da década de 1990. Lembramos, inclusive, que sua formulação teórica se deu há cerca de oitenta anos, com Campbell, em 1941, e Schellenberg, em 1956. A força propulsora do uso, em maior escala, desse tipo de classificação, que antes era dominada por instrumentos baseados em assuntos, foi a divulgação da ISO 15.489-1, em 2001. A norma defendeu, como melhor prática para a classificação de documentos, um esquema baseado na análise dos processos de negócio, funções e atividades realizadas pela organização (IFOULD; JOSEPH, 2016).

Entretanto, a partir de 2004, estudos apontando as dificuldades de uso desse tipo de abordagem para a classificação de documentos de arquivo começam a aparecer (CANABRIA, 2004). Depois

disso, vários autores, em inúmeras partes do mundo (Canadá, Inglaterra, Finlândia, Malásia, entre outros), publicaram trabalhos sobre a usabilidade do plano de classificação funcional.

Em um primeiro momento, a falta de compreensão do instrumento acontecia no ambiente de documentos em papel. Situação que não se alterou nos sistemas de gestão de documentos híbridos (documento em papel e documento digital) ou mesmo em situações de predominância completa dos documentos digitais. A passagem do mundo do papel para o digital não repercutiu em uma melhora do nosso instrumental, mas, nesse novo ambiente, abriu-se a possibilidade concreta de desafiar essa questão. A diferença é que podemos utilizar tecnologias para mitigar, em certa medida, os problemas de uso.

A utilização de sistemas informatizados tem provocado uma descentralização importante das atividades de gestão de documentos. Aquilo que era feito nos setores de protocolo (registro, tramitação, classificação) e nos setores de arquivo (classificação, descrição, avaliação, transferência, recolhimento e controle) passaram a acontecer no momento da criação ou da captura do documento. Nesse sentido, o usuário passa a jogar um papel fundamental no funcionamento do sistema. Infelizmente, não temos tido programas de formação contínua que os capacitem a operar com essas novas tarefas.

É importante, então, que as estratégias e instrumentos utilizados sejam inteligíveis ou automatizados. Da mesma forma que, para operar um automóvel, não é necessário ser um engenheiro mecânico, não se pode exigir do usuário que ele seja arquivista. Qualquer pessoa pode dirigir um automóvel sem saber quais são as leis e teorias que fazem o carro se mover. Acreditamos que isso não é uma discussão, apenas, de questões tecnológicas, mas da necessidade de inteligência arquivística embarcada nesses sistemas.

Por que estamos preocupados com a questão da classificação? Por dois motivos. A classificação de documentos de arquivo não

é uma função autônoma, ela está ligada umbilicalmente à avaliação e à descrição. Essas funções arquivísticas sem a classificação ficam soltas no ar, isto é, experimentam tal nível de subjetividade que as esvazia de sentido. Outro motivo está no fato da classificação (agrupamento) preservar o vínculo arquivístico, sem o qual o documento não é nada além da informação que está contida nele. Os documentos são agrupados a partir de um amalgama chamado função-atividade-transação, o difundido modelo FAT, de Schellenberg (2005). É esse agrupamento e esse vínculo obtidos da atividade geradora dos documentos que é possibilitado pela classificação.

A pergunta que motiva e dá título a esse trabalho é a seguinte: a classificação automática de documentos de arquivo é uma solução para os problemas que os usuários encontram com a classificação funcional? O objetivo é identificar as possibilidades de inserir, nos sistemas informatizados de gestão de documentos, a classificação automática de documentos de arquivo. Precisamos verificar o que a literatura arquivística tem falado sobre esse tipo de classificação, que desonera o usuário do trabalho de classificar diretamente os documentos que acumula.

Quanto à natureza da pesquisa, ela é do tipo aplicada por pretender produzir conhecimento que pode alterar determinada realidade. Ela é qualitativa, em relação à abordagem do problema, e exploratória e descritiva, no tocante aos seus objetivos, pois o tema classificação automática ainda está em fase preliminar de desenvolvimento e porque torna claro um problema. A técnica de coleta de dados é a pesquisa bibliográfica relacionada à classificação de documentos de arquivo, à abordagem funcional, às questões de usabilidade do instrumento de classificação e a classificação automática de documentos.

A fundamentação teórica do trabalho foi elaborada a partir da revisão da literatura sobre os temas acima indicados. Esta deu-se nos idiomas português, inglês, francês e espanhol, sem nenhuma delimitação de tempo, origem ou abordagem.

Esse trabalho está estruturado em cinco seções, incluindo a introdução. A segunda delinea a matriz teórica com o intuito de embasar e fundamentar o estudo. A terceira seção aborda os problemas, apontados na literatura, sobre o uso da classificação funcional e, a quarta, identifica as possibilidades de utilização da classificação automática. A quinta seção traz as nossas considerações finais.

## 2 FUNDAMENTAÇÃO TEÓRICA

A classificação constitui uma categoria de estudo de diversas áreas: filosofia, biologia, história, biblioteconomia, arquivologia, entre outras. Não pretendemos identificar a multiplicidade de conceitos já produzidos a partir de cada uma das contribuições fornecidas por aquelas disciplinas do conhecimento humano, mesmo porque seria uma tarefa difícil por conta dos limites de tempo e espaço. Privilegiaremos apenas as categorias teóricas com as quais pretendemos dialogar e as que poderão auxiliar-nos na compreensão sobre o objeto de estudo.

Parte-se do pressuposto, nesse trabalho, que a classificação de documentos de arquivo não pode limitar sua finalidade somente ao acesso aos documentos. A classificação de documentos de arquivo, diferentemente do que ocorre, por exemplo, na biblioteca, precisa fundamentar outras operações do quefazer arquivístico. Encontramos três objetivos a serem atingidos com a classificação: manter o contexto de produção dos documentos, fundamentar a avaliação e a descrição e, também, permitir o processo de busca e a recuperação da informação.

Para alguns autores, principalmente Duranti (1997), a prática de classificar documentos se origina da necessidade de explicitar o “vínculo arquivístico” (*archival bond*), que existe entre todos os documentos que participam da mesma atividade desde o momento de sua criação.

E, portanto, a recuperação de documentos de arquivo é somente um benefício colateral da classificação, sendo o objetivo principal colocar os documentos individuais nos agrupamentos aos quais pertencem, baseando-se na missão e nas funções do criador.

Heredia Herrera considera haver uma relação indissolúvel entre a classificação e a descrição. Para ela, “a descrição é a representação da classificação. De tal maneira, que é difícil, ao menos de forma correta, descrever sem classificar” (HEREDIA HERRERA, 2011, p. 66, tradução nossa). É possível observar, na norma ISAD(G), a importância da classificação para a descrição. Todo trabalho descritivo é baseado nos níveis de classificação de um arquivo. As regras da descrição multinível dependem diretamente do instrumento de classificação de documentos de arquivo (CONSELHO INTERNACIONAL DE ARQUIVOS, 2000). No que tange à avaliação, Sabourin (2001), em relato sobre a aplicação da macroavaliação no governo canadense, demonstra a necessidade da classificação funcional para a realização da macroavaliação ou avaliação funcional. Precisou-se de um novo modelo de classificação baseado na função para organizar documentos por função, subfunção, atividade, transação e processo de negócios e não por assunto, por estruturas hierárquicas de objetos ou por produtos finais.

Gagnon-Arguin e Bannouri (1998) entendem que a classificação de documentos de arquivo tem objetivos operacionais e objetivos estratégicos. Em relação aos operacionais, eles identificam a organização dos documentos, a recuperação e a interpretação da informação. Quanto aos estratégicos, teríamos, então, a classificação como um instrumento de gestão da informação, de respeito à legislação sobre o acesso aos documentos e base da constituição do fundo.

Heón (2003, p. 220-221) defende que o objetivo da classificação é garantir que todos os documentos estejam relacionados ao seu fundo original. Assim, dentro dele, ordenados da forma que lhes foi dada por seu criador.

A norma australiana de classificação de documentos (COUNCIL OF STANDARDS AUSTRALIA, 2011, p. 4) afirma que a classificação não é apenas para descrever o conteúdo do documento ou permitir a pesquisa e a recuperação. Mais do que isso, ela é usada para oferecer suporte a todo um conjunto de processos de *recordkeeping*, incluindo a identificação dos documentos que precisam ser capturados e criados para apoiar os negócios e fornecer uma estrutura para determinar e implementar decisões de retenção, segurança e acesso. A classificação é utilizada, também, para garantir que os vínculos entre as funções, que estão sendo executadas, e os documentos resultantes sejam consistentes e mantidos. Portanto, o objetivo não é descrever o assunto ou o conteúdo do documento, mas colocar o documento no contexto da função que o criou.

Pearce-Moses (2005) define classificação como “A organização de materiais dentro de categorias de acordo com um esquema que identifica, distingue e relaciona essas categorias”. O *Multilingual archival terminology*, do Conselho Internacional de Arquivos, apresenta seis definições para classificação, destaco duas delas: 1 – A organização sistemática de documentos em grupos ou categorias de acordo com métodos, procedimentos ou convenções representadas em um plano ou esquema (Glossário de termos arquivísticos e de *recordkeeping*, do Departamento de Trabalhos Públicos, do Estado de Queensland, Austrália); 2 – Identificação e arranjo sistemáticos das atividades de negócios em categorias logicamente estruturadas, de acordo com convenções, métodos e procedimentos, representados no sistema de classificação para auxiliar os negócios, acesso contínuo e apropriadas avaliação e eliminação (Terminologia utilizada na Universidade de British Columbia, no Canadá).

O *National Archives of Australia* (2003) estabeleceu o termo classificação da seguinte forma: a identificação e organização sistemática de atividades de negócios e/ou de documentos em categorias, de acordo com convenções logicamente estruturadas, métodos

e regras processuais representados em um sistema de classificação. Para Guercio (2002, p. 437, tradução nossa), a “classificação significa o reconhecimento, a identificação e a organização funcional de documentos, de acordo com critérios lógicos e consistentes em unidades de classificação funcionais distintas”.

No mundo francófono, o Dicionário de terminologia arquivística (2002), da Direction des Archives de France, compreende a classificação como uma operação de ordenação intelectual e física dos documentos de arquivo no interior de dossiês e esses dentro dos fundos, resultado da aplicação do princípio de respeito aos fundos. Roberge (2011, p. 469) defende a classificação como ação de identificar um documento por meio de um esquema de classificação. No Glossário do Portal de Arquivística Francófona (PIAF)<sup>55</sup>, de 2015, classificação significa a “operação intelectual e material para analisar e a ordenar os documentos de arquivo conforme os princípios arquivísticos e seus resultados” (p. 23, tradução nossa).

No mundo hispânico, Heredia Herrera (2013) entende a classificação como uma função arquivística em que se estabelecem as categorias documentais baseadas nas funções e na estrutura orgânica de uma instituição e se manifesta no quadro de classificação do fundo. Ela é determinada pelo princípio da proveniência. Cruz Mundet, em uma abordagem muito mais ampla, definiu-a como a

Ação e efeito de agrupar hierarquicamente os documentos de um fundo mediante agrupamentos ou classes, desde os mais amplos aos mais específicos, de acordo com os princípios da proveniência e ordem original; para o qual se identificam os tipos documentais, evidenciam-se as relações que existem entre eles e se organizam em uma estrutura lógica, chamada quadro de classificação, que reflete hierarquicamente essas relações (CRUZ MUNDET, 2011, p. 111, tradução nossa).

55 Disponível em: [https://www.piaf-archives.org/sites/default/files/bulk\\_media/glossaire/glossaire\\_papier.pdf](https://www.piaf-archives.org/sites/default/files/bulk_media/glossaire/glossaire_papier.pdf). Acesso em: 3 jul. 2021.

A ABNT NBR 15.489-2018 percebe a classificação como a “identificação sistemática e/ou configuração de atividades de negócio e/ou documentos de arquivo em categorias de acordo com convenções, métodos e regras estruturadas logicamente” (ABNT, 2018, p. 2). A norma australiana sobre classificação de documentos (HB 5031-2011 – *Handbook Records Classification*) sustenta, no âmbito do *recordkeeping*<sup>56</sup>, que ela tem objetivos ambiciosos de gerenciar documentos e de garantir que eles possam ser encontrados, usados e compreendidos ao longo do tempo.

E, por último, as definições encontradas no Brasil. O Dicionário Brasileiro de Terminologia Arquivística, da Associação dos Arquivistas Brasileiros (1990), estabelece que se trata de um “processo que consiste em colocar ou distribuir documentos ou informações, de acordo com planos previamente adotados” (p. 28). O Dicionário Brasileiro de Terminologia Arquivística, do Arquivo Nacional (2005), apresenta, talvez, uma das mais pobres definições em relação às apresentadas aqui. Ele entende a classificação como “organização dos documentos de um arquivo ou coleção, de acordo com um plano de classificação, código de classificação ou quadro de arranjo” (p. 49). O Dicionário de Terminologia Arquivística (CAMARGO; BELLOTTO, 2010), que se aproxima das definições internacionais, entende a classificação como a “sequência de operações, que, de acordo com as diferentes estruturas, funções e atividades da entidade produtora, visam a distribuir os documentos de um arquivo” (p. 30).

Nesse trabalho, a concepção que se utiliza de classificação é a atividade intelectual de agrupar, a partir de pesquisa, os documentos de arquivo de acordo com um princípio de classificação, definindo as operações necessárias à subordinação e à coordenação dos

56 Na definição do *Glossary of archival and records terminology*, de Pearce-Moses (2005), *recordkeeping* significa a “criação, uso, manutenção e disposição sistemática de documentos para atender às necessidades e responsabilidades administrativas, programáticas, legais e financeiras” (p. 331, tradução nossa). Com muito cuidado, podemos afirmar que se trata de sinônimo de *records management*.

níveis de classificação. Não se faz distinção, nessa definição, entre os suportes documentais.

Os trabalhos de Schellenberg aprofundaram a questão da classificação. E os que defendem a classificação funcional são, de alguma forma, credores do arquivista americano, pois foi com o seu livro *Arquivos modernos: princípios e técnicas*, lançado em 1956, que esse modelo foi disseminado pelo mundo. Essa verticalização empreendida pelo arquivista americano fica clara nas observações sobre classificação de documentos públicos e nos pontos estabelecidos para elaboração de um sistema de classificação.

Encontramos a definição de classificação funcional em algumas publicações da área. O *Glossary of archival and records terminology*, de Richard Pearce-Moses, editado pela Sociedade de Arquivistas Americanos (2005), entende como um sistema para organizar materiais de acordo com a *atividade ou função*. O *Glossary of terms* (2009), do *International Records Management Trust*, definiu o termo como um sistema para organizar materiais com base na função, atividade ou tarefa executada por uma organização para cumprir seu mandato, em vez de se basear por departamento, nome ou assunto. O Dicionário Brasileiro de Terminologia Arquivística (2005, p. 132) não apresenta nenhuma definição para classificação funcional, mas fica claro quando afirma, no verbete plano de classificação, que este é “elaborado a partir do estudo das estruturas e funções de uma instituição (...)”, deixando nítida a opção por aquela abordagem.

A classificação funcional está pacificada na literatura da área, inclusive caminhamos para uma naturalização da classificação baseada na função. Vários autores fizeram a defesa desse modelo de classificação de documentos de arquivo, dentre eles: Duranti *et. al.* (2002), National Archives of Australia (2003), Shepherd e Yeo (2003), Roberge (2011), Caravaca (2017). As normas ISO 26.122 (2008) e a ISO 15.489 (2016) também seguiram a linha desses autores. Esta última norma,

inclusive, recomenda, claramente, que o desenvolvimento do plano de classificação de documentos de arquivo seja baseado em uma análise de funções, atividades e processos de trabalho.

A formulação teórica da classificação funcional é antiga, mas a implementação prática é relativamente recente (ALBERTS; SCHELLINCK; EBY; MARLEAU, 2010, PACKALÉN; HENTTONEN, 2016). Apesar da ideia não ser nova, a classificação baseada nas funções somente começou a ser aplicada em meados da década de 1980. Orr (2005) cita, por exemplo, nesse renascimento do interesse pela classificação funcional, o trabalho de Haas, Samuels e Simmons, que, em 1985, estruturou a classificação das atividades e processos que geravam documentos na ciência moderna e tecnologia.

### 3 OS PROBLEMAS NO USO DA CLASSIFICAÇÃO FUNCIONAL

Apesar da pacificação existente sobre a abordagem funcional na classificação de documentos, temos poucos exemplos que tenham sido implementados com sucesso (ORR, 2005). A questão da usabilidade de planos de classificação funcionais é particularmente significativa para usuários leigos. Afinal, “a teoria não ajuda se as pessoas não a puderem aplicar” (HURLEY; SYMONDSON, 2002, p. 25 *apud* ORR, 2005, p. 70).

Orr (2005) fez uma pesquisa a respeito do que havia sido escrito na literatura da área sobre o tema, as visões dos especialistas e a experiência dos gestores de documentos da Austrália, Canadá e Reino Unido. Os resultados demonstraram que as desvantagens de aplicar o método funcional estavam assentados no sentimento dos usuários de que o instrumento não era intuitivo, que eles achavam confuso e que sentiam uma redução na produtividade com o tempo gasto para a compreensão desse tipo de classificação.

A aplicação da análise funcional para a construção de planos de classificação pode ter como resultado um sistema abstrato e autorreferencial, uma classificação que representaria as funções da organização, mas que seria totalmente incapaz de capturar os modos de trabalho em uma unidade organizacional real. Como consequência disso, os usuários teriam dificuldades de aplicar a classificação (FOSCARINI, 2010). Os sistemas funcionais tendem a criar silos de informação em uma organização, tornando mais difícil a produção de visões horizontais da informação. Quando os processos de trabalho são horizontais e cruzam múltiplas funções, a informação criada não pode ser agrupada (ALBERTS; SCHELLINCK; EBY; MARLEAU, 2010).

Ifould e Joseph (2016, p. 18) apresentaram, no quadro abaixo, uma síntese da literatura arquivística da percepção dos usuários sobre as dificuldades encontradas para registrar, pesquisar e encontrar informação orgânica com planos de classificação funcionais.

Quadro 1 – Percepções dos usuários sobre a classificação funcional

Percepções	Fontes
Os usuários consideram que os esquemas de classificação funcional não são intuitivos e nem fáceis de utilizar.	GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b> , v. 22, n. 2, p. 116–129, 2012. JOSEPH, Pauline. <b>EDRMS search behaviour: implications for records management principles and practices</b> , 2010. Tese – University of Western Australia, Perth, 2010.
Os usuários consideram os esquemas de classificação funcional complicados ou demasiadamente complicados.	GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b> , v. 22, n. 2, p. 116–129, 2012. JOSEPH, Pauline. <b>EDRMS search behaviour: implications for records management principles and practices</b> , 2010. Tese – University of Western Australia, Perth, 2010.
Os usuários nem sempre sabem onde arquivar o documento em um esquema de classificação funcional.	GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b> , v. 22, n. 2, p. 116–129, 2012. JOSEPH, Pauline. <b>EDRMS search behaviour: implications for records management principles and practices</b> , 2010. Tese – University of Western Australia, Perth, 2010.

<p>Os usuários acham os esquemas de classificação funcional confusos por ser possível classificar os documentos em muitas categorias.</p>	<p>GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b>, v. 22, n. 2, p. 116–129, 2012.          ORR, Stuart Anthony. <b>Functions-based classification of records: is it functional?</b> 2005. Dissertação (Master of Science in Records Management) – Divisão de Informação e Estudos de Comunicação, Northumbria University, Newcastle, 2005.          SMYTH, Z. A. Implementing EDRM: has it provided the benefits expected? <b>Records Management Journal</b>, v. 15, n.3, p. 141-149, 2005.</p>
<p>Os usuários acham que o pior método é o funcional.</p>	<p>GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b>, v. 22, n. 2, p. 116–129, 2012.</p>
<p>Os usuários consideram desnecessário o plano de classificação funcional e o uso é demorado.</p>	<p>GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. <b>Records Management Journal</b>, v. 22, n. 2, p. 116–129, 2012.</p>
<p>Os planos de classificação funcional não são intuitivos ou nem amigável para o usuário. Ele fica desconfortável com a abordagem funcional.</p>	<p>JONES, Pauline. The role of virtual folders in developing an electronic document and records management system: Meeting user and records management needs, <b>Records Management Journal</b>, v.18, n.1, p. 53-60, 2008.</p>
<p>Os usuários não compreendem como o plano de classificação funciona e como usar. Não gostam de usar a classificação funcional, porque é difícil de aplicar e é pouco clara.</p>	<p>JOSEPH, Pauline. <b>EDRMS search behaviour: implications for records management principles and practices</b>, 2010. Tese – University of Western Australia, Perth, 2010.          ORR, Stuart Anthony. <b>Functions-based classification of records: is it functional?</b> 2005. Dissertação (Master of Science in Records Management) – Divisão de Informação e Estudos de Comunicação, Northumbria University, Newcastle, 2005.          PACKALÉN, Saara, HENTTONEN, Pekka, Recordkeeping professional understanding of and justification for functional classification: Finnish public sector organizational context, <b>Archival Science</b>, v. 16, p. 403-419, 2016.</p>
<p>Os usuários consideram a função um conceito estranho e difícil de compreender e não intuitivo.</p>	<p>CALABRIA, Tina. Evaluating Caloundra City Council's EDMS classification. 2004. Disponível em: <a href="https://www.steptwo.com.au/files/kmc_caloundracouncil.pdf">https://www.steptwo.com.au/files/kmc_caloundracouncil.pdf</a>. Acesso em: 7 mai. 2020.          FOSCARINI, Fiorella. Understanding functions: An organizational culture perspective. <b>Records Management Journal</b>, v. 22, p. 20–36, 2012.          ORR, Stuart Anthony. <b>Functions-based classification of records: is it functional?</b> 2005. Dissertação (Master of Science in Records Management) – Divisão de Informação e Estudos de Comunicação, Northumbria University, Newcastle, 2005.</p>
<p>Os usuários consideram que os planos de classificação funcional são inflexíveis em relação com as antigas unidades compartilhadas com estrutura pessoal, tornando os planos impopulares.</p>	<p>FOSCARINI, Fiorella. Understanding functions: An organizational culture perspective. <b>Records Management Journal</b>, v. 22, p. 20–36, 2012.          MORELLI, Jeff. Hybrid filing schemes: the use of metadata signposts in functional file plans, <b>Records Management Journal</b>, v. 17, n. 1, p. 17–31, 2007.</p>

Para os usuários, a classificação não é intuitiva, é desnecessariamente complexa e mais demorada do que por outros meios.	BAK, Greg. Continuous classification: capturing dynamics relationships among information resources. <b>Archival Science</b> , v. 12, p. 287-318, 2012.
Pelo fato da classificação funcional permitir que um documento seja organizado num único local (fixidez), os usuários também têm capacidade de pesquisa limitada.	ALBERTS, Inge, SCHELLINCK, Jen, EBY, Craig, MARLEAU, Yves. Bridging Functions and Processes for Records Management. <b>Canadian Journal of Information and Library Science</b> , Toronto, v. 34, n. 4, p. 365-390, dez. 2010.

Fonte: adaptado de IFOULD; JOSEPH, 2016.

Analisando as percepções apresentadas pelos vários autores, fica evidente a dificuldade que os usuários têm de utilizar o plano de classificação funcional. Ele parece não representar efetivamente o que os funcionários fazem em seu trabalho cotidiano. O nível de abstração é muito alto, impedindo o vínculo direto entre o documento e a atividade que o gerou. Se os sistemas não geram confiança, os documentos permanecerão com indivíduos ou grupos de trabalho (OLIVER; FOSCARINI, 2013), criando ilhas informacionais sem nenhuma conexão e sem ter todo o seu potencial explorado de maneira consistente.

Esta dificuldade é entendida por Bak (2012) pelo fato dos teóricos da gestão de documentos promoverem a classificação funcional sem definições autorizadas de funções. Para Duranti e Franks (2015), um dos fatores que contribuem para tornar a análise das funções difícil tem a ver com a relatividade das noções de missão (geralmente associado à função) e processo ou meio (comumente relacionado à atividade). As autoras sugerem que os estudos em sociologia, administração e psicologia social podem ajudar no entendimento do conceito de função.

A abordagem funcional atende a dois dos três objetivos da classificação de documentos de arquivo: a manutenção do vínculo arquivístico e a fundamentação de outras funções arquivísticas (avaliação e descrição, principalmente). A recuperação dos documentos, como um dos objetivos, não é completamente preenchida, mesmo no ambiente tradicional de documentos em papel. O agrupamento favorecia, principalmente, a busca contextual dos documentos, mas não auxiliava em

buscas pelo conteúdo. Caravaca (2017) entende que o plano de classificação é o primeiro instrumento de acesso aos documentos do arquivo. Nos sistemas informatizados de gestão de documentos, essa tarefa não está diretamente vinculada à classificação, mas à articulação e relacionamento dos metadados. Esse novo ambiente e a implementação dos metadados permitiram, inclusive, a quebra da barreira da fixidez dos agrupamentos dados pela classificação aos documentos em papel.

Entretanto, as tentativas anteriores de fundamentar os planos de classificação em assuntos, estrutura organizacional, tipos documentais, espécies documentais ou a mistura de todos esses elementos não obtiveram sucesso. O contexto de criação e uso dos documentos não pode ser mantido a partir desse tipo de classificação. O vínculo arquivístico não era estabelecido, tornando os documentos objetos isolados em seu próprio conteúdo.

Os gestores de documentos, de acordo com Warland e Mokhtar (2013), parecem estar presos entre uma teoria da classificação e o lugar difícil da prática de classificação do usuário final. Eles afirmam que não mais de 5% de todos os documentos digitais criados ou recebidos pelas organizações estão em conjuntos classificados em sistema de gestão de documentos. Os restantes são armazenados, não classificados institucionalmente, em unidades de rede ou de computadores “pessoais”, nos sistemas de negócio e nos servidores de *e-mails* (SOUSA, 2020).

De qualquer forma, mesmo com as vantagens do plano de classificação funcional em relação à manutenção do vínculo arquivístico, à fundamentação de outras funções arquivísticas e à estabilidade e continuidade, não resolvemos ainda a importante questão da usabilidade desses instrumentos, como demonstram os estudos de caso sobre a aplicação deles (CALABRIA, 2004; ORR, 2005; FOSCARINI, 2009; ALBERTS; SCHELLINCK; EBY; MARLEAU, 2010; IFOULD; JOSEPH, 2016). Os arquivistas precisam descobrir o que seus usuários pensam e dizem sobre os serviços prestados pelos sistemas

de gestão de documentos. É fundamental iniciar um diálogo com os usuários em vez de se concentrar somente nos sistemas e nas tecnologias envolvidas (OLIVER; FOSCARINI, 2013).

Desenhar sistemas e processos de trabalho que facilitem aos produtores a criação e armazenamento de bons documentos é vital, mas desafiante na prática. Isso requer habilidades de engenharia de sistemas de alto nível junto a uma sólida compreensão dos requisitos de gestão de documentos, processos de trabalho e comportamento humano (CUNNINGHAM, 2021).

Acreditamos que o vínculo arquivístico e a fundamentação de outras funções arquivísticas encontram, na abordagem funcional, a estabilidade necessária para serem implementados. Por outro lado, o plano de classificação de documentos de arquivo não pode deixar de satisfazer as necessidades e expectativas dos usuários. É sempre importante lembrar que o instrumento é operado diretamente pelo usuário no momento em que ele cria ou recebe o documento fruto de suas atividades na organização. Observamos que, nos sistemas informatizados, a descentralização de responsabilidades pela gestão de documentos se aprofundou, exigindo do usuário uma participação maior e vital para o funcionamento do sistema como um todo.

Os caminhos possíveis para fugir dessa encruzilhada (vinculação do documento à atividade que o gerou e a usabilidade do instrumento) são, primeiro (sem ordem de valor), de não adotar a análise funcional como o ponto de chegada para a definição de um plano de classificação funcional, mas como ponto de partida. A análise funcional é importante para definir os níveis mais altos da estrutura do plano, mas, nos níveis mais baixos, outros elementos seriam mais interessantes, tais como: série documental, tópico, critério geográfico, necessidades dos usuários, entre outros. O segundo, é ter um melhor alinhamento entre a teoria da administração e a de gestão de documentos, como aconselham Alberts *et al.* (2010). E, por último, a

classificação automática, que já começou a ser tratada pela literatura da área. É sobre isso que vamos tratar na próxima seção.

## 4 A CLASSIFICAÇÃO AUTOMÁTICA DE DOCUMENTOS DE ARQUIVO

Classificar é uma atividade humana. Os seres humanos (quase) instintivamente organizam as coisas e os fenômenos em categorias. Portanto, nós a classificamos. Sobreponemos ou vinculamos com algum tipo de forma e estrutura e, ao fazermos isso, os dotamos (coisas e fenômenos) de um significado e contexto, que nos ajudam a dar-lhes sentido, a tornar a sua gestão mais eficiente (WARLAND; MOKHTAR, 2013).

O conceito de classificação é apropriado sem que se tenha consciência sobre ele. A classificação trafega pelos vários momentos da vida dos seres humanos. A maioria das pessoas não é consciente do fato que classificamos em uma grande extensão em nossas vidas diárias. Sem classificação o progresso humano seria impossível (SOUSA, 2005).

Para Langridge (1977, p. 11), o fato da inconsciência é um indício da natureza fundamental do processo de classificação. E mais, “sem classificação não poderia haver nenhum pensamento humano, ação e organização que conhecemos. A classificação transforma impressões sensoriais isoladas e incoerentes em objetos reconhecíveis e padrões recorríveis”. Buscando a Psicologia, Langridge cita o inglês Patrick Meredith para quem “grande parte da arte de aprender consiste em regularizar hábitos pessoais de classificação”. Astério Campos (*apud* PIEDADE, 1983, p. 16) considera que “a classificação, entendida como processo mental de agrupamento de elementos portadores de características comuns e capazes de ser reconhecidos como entidade ou conceito, constitui uma das fases fundamentais do pensar humano”.

Desde os sistemas de gestão de documentos em papel, a classificação é um elemento fundamental e esse papel permaneceu ou mesmo ficou mais evidenciado no ambiente digital. Entretanto, substituir os agrupamentos criados pelo ato classificatório por metadados relacionados ao conteúdo dos documentos é uma falsa solução, pois esvazia todo o esforço para a manutenção do vínculo arquivístico e a fundamentação das outras funções arquivísticas (avaliação e descrição).

A falta de compreensão de que eram necessárias novas atitudes e soluções com a expansão do ambiente digital gerou, em um primeiro momento, a ideia de que seria importante apenas reproduzir as pastas, as gavetas e as caixas existentes no mundo físico em diretórios e subdiretórios na raiz dos computadores. Não se atentou para as grandes possibilidades que a Tecnologia da Informação poderia oferecer ao mundo arquivístico. Apresentamos essa discussão somente para contextualizar, pois as dificuldades dos usuários no uso e na aplicação do plano de classificação funcional continuavam, tinham apenas migrado para outro ambiente.

Agrupamos os documentos para lhes dar sentido e contexto. Como visto na seção anterior, os métodos, que não passam, em alguns momentos, de recomendações vagas, parecem falhar nessa tarefa. Cunningham (2021) reconheceu isso na Austrália, para ele nada funcionou como se esperava, inclusive com a inevitável relação entre análise funcional e classificação.

A falibilidade desses métodos está levando as organizações a abandonarem a classificação pela utilização de metadados, que resolvem em parte o problema. Entretanto, esse tipo de alternativa fornece não mais do que uma forma de agrupamento temporário, não preserva, na maioria das vezes, o contexto de produção e uso dos documentos.

A partir da década de 2010, algumas experiências começam a ser feitas na busca de alternativas a essa encruzilhada existente entre a

classificação funcional e a sua usabilidade. Como vimos anteriormente, a classificação funcional é uma ideia que leva em consideração a relação entre três entidades: o agente acumulador (pessoa física ou jurídica), a função/atividade e o documento. Isso está bastante claro para os arquivistas, como aponta a literatura, mas está embaçado para os usuários, que acham isso pouco intuitivo e difícil de compreender.

Encontramos, na literatura, alguns poucos exemplos sobre o uso de classificação automática de documentos de arquivo. Na língua inglesa, encontramos três termos: *auto-categorization*, *auto-classification* ou *automatic classification*. Conceito que, ainda, não ganhou uma definição nos glossários e dicionários da área.

Guercio (2017) chama a atenção para o fato de que a atividade de reflexão teórica, tão expressiva entre 1995 e 2005, sobre os documentos digitais, foi deixada de lado nos últimos anos. Isso permitiu, ao mercado (fornecedores de *software*), propor soluções operacionais que, de fato, têm sido limitadas a replicar passivamente o existente e não embarcando nessas ferramentas inteligência arquivística, na forma de requisitos para o funcionamento desses instrumentos. Não iniciando, assim, processos de inovação.

Encontramos um estudo realizado na Universidade de Richmond, que examinou se a análise assistida por tecnologia no *e-discovery* poderia ser mais eficaz e mais eficiente do que a análise manual exaustiva. Uma análise manual alcançou 65% de precisão e 65% de revocação, que é o nível que os seres humanos concordam uns com os outros. Em comparação, uma análise baseada no algoritmo informático, de documentos do arquivo de correio eletrônico de uma empresa, e tendo sido analisado 1,9% dos documentos por seres humanos, para ajudar o sistema a aprender, alcançou taxas médias de revocação e precisão de 76,7% e 84,7%, respectivamente. Assim, se os gestores de documentos pudessem ajudar um sistema a aprender a classificar documentos digitais, as taxas de sucesso para a classificação desses documentos

poderiam ser tão elevadas como indicado nos resultados da pesquisa da Universidade de Richmond (WARLAND; MOKHTAR, 2013).

O *blog* dos Arquivos do Estado de Nova Gales do Sul<sup>57</sup>, na Austrália, veiculou uma experiência interessante sobre o uso da classificação automática (*auto-classification*) para organizar documentos que não tinham recebido tratamento. Na verdade, eram 400 mil documentos digitais contidos em 31 mil pastas, algumas com até 14 níveis de profundidade, de uma agência governamental. Devido a questões administrativas, os documentos precisavam ser migrados e classificados em questão de meses. O que tornava a classificação manual impossível.

Optou-se, então, pela classificação automática a partir de um módulo do software TRIM. A solução envolvia três componentes. O primeiro foi o reconhecimento óptico de caracteres, um OCR, que transformou arquivos de imagem em texto legível. O arquivo foi, em seguida, indexado por um servidor de indexação de conteúdos. E, por último, encaminhado para o módulo de classificação automática.

A precisão do sistema de classificação automática dependia da definição de um conjunto de termos para vinculá-los a cada categoria (classes do instrumento de classificação). Segundo os relatos, esta foi a parte mais difícil, que consumiu mais recursos do projeto como um todo e foi um fator crítico para o sucesso.

Ribeiro e Assis (2018) discutem, de forma interessante, o *deep learning* e suas potencialidades para a classificação de documentos de arquivo. Eles entendem que, com o grande volume de documentos digitais acumulados diretamente em várias plataformas, muitas vezes de forma não estruturada e organizada, a tarefa de classificá-los poderia ser facilitada com o auxílio de algoritmos de *deep learning*, que fazem parte da chamada “inteligência artificial leve”. Acreditam, inclusive, que

57 Using auto-classification to classify unmanaged records. Disponível em: <https://futureproof.records.nsw.gov.au>. Acesso em: 20 nov. 2021.

é possível criar sistemas que analisem uma quantidade massiva de dados e, a partir dessas informações, classificar documentos.

Como sugestão, Ribeiro e Assis propõem que:

[...] para alcançar uma eficácia e eficiência na utilização de tais técnicas, é necessário que haja estudos prospectivos da área de Arquivologia em torno da Inteligência Artificial, possibilitando assim o entendimento de como tais algoritmos funcionam e os aplicando em seu cotidiano, juntamente com profissionais de Tecnologia da Informação (TI). (RIBEIRO; ASSIS, 2018).

Oladejo e Hadzidedic (2020), na mesma linha dos autores brasileiros, entenderam que as técnicas de classificação automática de documentos utilizam algoritmos, que aprendem com as classificações humanas. Dessa forma, podem realizar a tarefa humana de classificação a partir de uma formação adequada. Citaram, inclusive, um projeto intitulado SEEK! que criou uma solução de aprendizagem automática com um algoritmo de recomendação baseado em regras para decidir automaticamente a pasta de destino e os metadados de um documento, com base no seu conteúdo.

Para Payne e Baron (2017), as novas ferramentas de categorização automatizada (*auto-categorization*) são necessárias para ultrapassar as limitações das tradicionais categorizações e manter a força do vínculo arquivístico. Eles entendem que essa abordagem traz vantagens significativas, que incluem a capacidade de:

- indexar automaticamente um único documento ou um grupo de documentos;
- permitir a atualização automatizada da massa de relações entre documentos;
- desenvolver linguagem natural consistente que pode melhorar o acesso dos usuários aos dados e repositórios;

- lidar com múltiplas linguagens de maneira eficaz;
- acessar uma plataforma independentemente por diversos usuários e para inúmeros fins.

Os autores acima citam como métodos de categorização automática a classificação baseada em regras, a correlação difusa, o vetor espacial, a similaridade, o baseado em árvore e a rede neural. Os métodos de categorização automática são concebidos para reduzir significativamente os custos que as organizações têm em relação aos seus orçamentos para gestão de documentos. Isso diminui a despesa global de armazenamento e arquivamento (PAYNE; BARON, 2017). Algumas fronteiras foram abertas, mas está claro que precisamos de mais pesquisa e aplicação para tornar os métodos de classificação automática eficientes.

## 5 CONSIDERAÇÕES FINAIS

Vivemos, atualmente, uma situação híbrida, da mesma forma que temos grandes massas documentais acumuladas em suporte papel estamos formando outras no ambiente digital. O fato é que assistimos meio inertes a criação de volumes cada vez maiores de documentos digitais com pouco tratamento arquivístico. Os sistemas de gestão de documentos não conseguiram capturar todos aqueles registros. Muitos desses documentos passam longe do controle arquivístico, dando origem a repositórios de documentos de arquivo sem nenhum requisito de gestão de documentos.

Isso pode ser observado no caso dos sistemas de negócio, nos servidores de *e-mails*, nos computadores “pessoais” e nos diretórios de rede. Nesses espaços, a organização, a preservação e a recuperação estão sob a jurisdição das decisões arbitrárias, da linguagem

natural, da falta de conhecimento sobre o que é um documento de arquivo, entre outros elementos.

O tratamento desses conjuntos documentais não tem condições de ser feito manualmente. Precisamos lançar mão da tecnologia da informação para processar esses enormes volumes de documentos digitais e que crescem em progressão geométrica, como é o caso dos *e-mails*. Soluções têm sido encontradas, mas que tem uma fundamentação mais prática do que teórica. O exemplo maior disso é o projeto do Arquivo Nacional Americano intitulado *Capstone*.

O objetivo deste trabalho foi o de identificar as possibilidades de inserir, nos sistemas informatizados de gestão de documentos, a classificação automática de documentos de arquivo. Para isso verificamos o que a literatura arquivística tem escrito sobre esse tipo de classificação, que desonera o usuário do trabalho de classificar diretamente os documentos que acumula.

Apesar de não estar no escopo desse trabalho, não só os usuários têm dificuldade na compreensão e aplicação da classificação funcional mas também os arquivistas, de acordo com Packalén e Henttonen (2016). Esse problema ganha especial importância quando percebemos que a classificação, além de matricial, está intimamente vinculada à avaliação e à descrição. Uma classificação errada pode resultar em uma aplicação da tabela de temporalidade equivocada e a uma descrição pouco consistente.

As primeiras experiências de uso da classificação automática apontam para o uso da *deep learning* com base na inteligência artificial. Entretanto, essas possibilidades precisam da junção dos conhecimentos em tecnologia da informação e em arquivologia. O canto de sereia de tratar o documento como uma entidade isolada, que tem valor apenas pelo conteúdo que veicula, é muito sedutor.

Outra possibilidade é adiantar o trabalho do arquivista para o momento da criação do documento. Partindo da ideia de que o documento é resultado das atividades, que são cotidianamente repetidas, portanto, podemos criar padrões de documentos, isto é, tipos documentais já classificados.

Muito trabalho precisa ser feito para que os métodos de classificação automática dos documentos de arquivo alcancem um grande desempenho, tornando desnecessária a intervenção manual, que no caso dos grandes volumes já acumulados seria impossível ou demandaria grandes recursos financeiros e materiais. Estão aí alguns dos caminhos que podemos seguir, mas somente a pesquisa pode construir soluções com requisitos de gestão de documentos.

## 6 REFERÊNCIAS

ALBERTS, Inge, SCHELLINCK, Jen, EBY, Craig, MARLEAU, Yves. Bridging Functions and Processes for Records Management. **Canadian Journal of Information and Library Science**, Toronto, v. 34, n. 4, p. 365-390, dez. 2010.

ARQUIVO NACIONAL (Brasil). **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO 15489-1: 2018**. Informação e documentação – Gestão de documentos de arquivo. Parte 1: Conceitos e princípios. Rio de Janeiro: ABNT, 2018

ASSOCIAÇÃO DOS ARQUIVISTAS BRASILEIROS. **Dicionário brasileiro de terminologia arquivística**. São Paulo: CENADEM, 1990.

ASSOCIACIÓN ESPAÑOLA DE NORMALIZACIÓN Y CERTIFICACIÓN (AENOR). **UNE-ISO/TR 26.122**. Información y documentación. Análisis de los procesos de trabajo para la gestión de documentos. Madrid: AENOR, 2008.

BAK, Greg. Continuous classification: capturing dynamics relationships among information resources. **Archival Science**, v. 12, p. 287-318, 2012.

BAK, Greg. La clasificación de documentos electrónicos: documentando relaciones entre documentos. **Tábula**, n. 13, p. 59-77, 2010.

CALABRIA, Tina. **Evaluating Caloundra City Council's EDMS classification**. 2004. Disponível em: [https://www.steptwo.com.au/files/kmc\\_caloundracouncil.pdf](https://www.steptwo.com.au/files/kmc_caloundracouncil.pdf). Acesso em: 7 maio 2020.

CAMARGO, Ana Maria, BELLOTTO, Heloisa Liberali. **Dicionário de terminologia arquivística**. São Paulo: Centro de Memória da Educação, 2010.

CAMPBELL, Eduard G. Functional classification of archival material. **The Library Quarterly**: information, community, policy, v. 11, n. 4, p. 431-441, 1941.

CARAVACA, Maria Mata. **Policies and requirements for archival sedimentation in a hybrid records management environment**: a critical analysis of international writings, 2017. Tese (Doutorado em Library and Documentation Sciences) – Sapienza Università di Roma, Roma, 2017.

CONSELHO INTERNACIONAL DE ARQUIVOS (CIA). **ISAD(G) Norma geral internacional de descrição arquivística**. Rio de Janeiro: Arquivo Nacional, 2000.

CONSELHO INTERNACIONAL DE ARQUIVOS (CIA). **Multilingual archival terminology**. Disponível em: <http://www.ciscra.org/mat/>. Acesso: em: 24 abr. 2020.

COUNCIL OF STANDARDS AUSTRALIA. **Handbook Records Classification**. HB 5031-2011, Sydney, 2011.

CRUZ MUNDET, José Ramón. **Diccionario de archivística**: con equivalências em inglês, francês, alemão, português, catalão, euskera y gallego. Madrid: Alianza Editorial, 2011.

CUNNINGHAM, Adrian. Cómo de lleno está el vaso? Cambios y desafíos para los profesionales de los documentos frente a la transformación digital en la era de los datos. **Tábula**, n. 24, 2021, p. 21-43.

DIRECTION DES ARCHIVES DE FRANCE. **Dictionnaire de terminologie archivistique**. Paris, 2002.

DURANTI, Luciana *et al.* **Preservation of the integrity of electronic records**. Dordrecht: Kluwer, 2002.

DURANTI, Luciana, FRANKS, Patrica C. (ed.). **Encyclopedia of Archival Science**. Lanham: Rowman & Littlefield, 2015.

DURANTI, Luciana. The archival bond. **Archives and Museum Informatics**, v. 11, p. 213-218, 1997.

FOSCARINI, Fiorella. **Function-based records classification systems**. An exploratory study of records management practices in central banks, 2009. Tese (Doutorado em Library, Archival and Information Studies) – University of British Columbia. Vancouver: British Columbia, 2009.

FOSCARINI, Fiorella. La clasificación de documentos basada em funciones: comparación de la teoría y la práctica. **Tábula**, n. 13, p. 41-57, 2010.

FOSCARINI, Fiorella. Understanding functions: An organizational culture perspective. **Records Management Journal**, v. 22, p. 20-36, 2012.

GAGNON-ARGUIN, Louise, BANNOURI, Rabii. **La classification des documents institutionnels dans les ministères du Gouvernement du Québec et le défi informatique**. Montreal: Conseil du Trésor, 1998.

GUERCIO, Maria. La classificazione nell'organizzazione dei sistemi documentari digitali: criticità e nuove prospettive. **JLIS.it**, v. 8, n. 2, mai., 2017.

GUERCIO, Maria. Records classification and content management: old functions and new requirements in the legislations and standards for electronic record-keeping systems. *In: Proceedings of the DLM-Forum 2002*. Access and preservation of electronic information: best practices and solutions. Barcelona, 2002.

GUNNLAUGSDOTTIR, Johanna. Functional classification scheme for records: a way to chart documented knowledge in organizations. **Records Management Journal**, v. 22, n. 2, p. 116-129, 2012.

HEÓN, Gilles. La classification. *In: COUTURE, Carol et al. Les fonctions archivistiques*. Montreal: Presses de L'Université du Québec, 2003. p. 220-254.

HEREDIA HERRERA, Antonia. **Lenguaje y vocabulario archivísticos**: algo más que un diccionario. Sevilla: Junta de Andalucía, 2011.

HEREDIA HERRERA, Antonia. **Manual de archivística básica**: gestión y sistemas. Cidade do México: Univesidad Autónoma de Puebla, 2013.

IFOULD, Peta, JOSEPH, Pauline. User difficulties working with a business classification scheme: a case study", **Records Management Journal**, v. 26, n. 1, 2016.

INTERNATIONAL RECORDS MANAGEMENT TRUST (IRMT). **Glossary of terms**. Londres: IRMT, 2009.

JOSEPH, Pauline. **EDRMS search behaviour**: implications for records management principles and practices, 2010. Tese – University of Western Australia, Perth, 2010.

LANGRIDGE, Derek. **Classificação**. Abordagem para estudantes de Biblioteconomia. Rio de Janeiro: Interciência, 2006.

MORELLI, Jeff. Hybrid filing schemes: the use of metadata signposts in functional file plans. **Records Management Journal**, v. 17, n. 1, p. 17-31, 2007.

NATIONAL ARCHIVES OF AUSTRALIA. **Overview of Classification Tools for Records Management**. Canberra: National Archives of Australia, 2003. Disponível em: <https://www.naa.gov.au/sites/default/files/2019-10/classification-tools.pdf>. Acesso em: 1 maio 2020.

OLADEJO, Babatunde, HADZIDEDIC, Suncica. Electronic records management: a state of the art review. **Records Management Journal**, v. 31, n. 1, p. 74-88, 2021.

OLIVER, Gillian, FOSCARINI, Fiorella. Effective records management: working collaboratively within organizations. **Comma**, Liverpool, n. 1, 2013. Disponível em: <https://doi.org/10.3828/comma.2013.1.14>. Acesso em: 15 maio 2020.

ORR, Stuart Anthony. **Functions-based classification of records: is it functional?**, 2005. Dissertação (Master of Science in Records Management) – Divisão de Informação e Estudos de Comunicação, Northumbria University. Newcastle: Northumbria University, 2005.

PACKALÉN, Saara, HENTTONEN, Pekka. Recordkeeping professional understanding of and justification for functional classification: Finnish public sector organizational context, **Archival Science**, v. 16, p. 403-419, 2016.

PAYNE, Nathaniel, BARON, Jason R. Auto-categorization methods for digital archives. *In*: IEEE International Conference on Big Data, 2017, Boston. **Anais [...]**. Boston: IEEE International Conference on Big Data, 2017.

PEARCE-MOSES, Richard (org.). **A glossary of archival and Records terminology**. Chicago: The Society of American Archivists, 2005.

PIECADE, Maria Antonietta Requião. **Introdução à Teoria da Classificação**. Rio de Janeiro: Interciência, 1983.

REED, Barbara. Just doing the same won't work lets make the digital record-keeping compelling! **Acervo**, Rio de Janeiro, v. 28, n. 2, jul./dez., 2015, p. 93-109.

RIBEIRO, Patrick Dourado, ASSIS, João Marcus Figueiredo. Deep learning e suas potencialidades para a classificação arquivística. *In*: Congresso Internacional em Humanidades Digitais, 1, 2018, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2018.

ROBERGE, Michel. **Le schéma de classification hiérarchique des documents administratifs**. Conception, développement, déploiement et maintenance. Québec: Éditions Michel Roberge, 2011.

SABOURIN, Paul. Constructing a functions-based records classification system: Business Activity Structure Classification System. **Archivaria**, Ottawa, v.51, p. 137-154, 2001.

SCHELLENBERG, Theodore Roosevelt. **Arquivos modernos**. Princípios e técnicas. Rio de Janeiro: FGV, 2005.

SHEPHERD, Elizabeth, YEO, Geoffrey. **Managing records**. A handbook of principles and practice. Londres: Facet, 2003.

SMYTH, Z. A. Implementing EDRM: has it provided the benefits expected? **Records Management Journal**, v. 15, n.3, p. 141-149, 2005.

SOUSA, Renato Tarciso Barbosa de. A classificação no ambiente digital ainda é a base da gestão de documentos e de arquivos? Alguns apontamentos para o debate. *In*: SIMPÓSIO INTERNACIONAL DE ARQUIVOS, 1, São Paulo, 2020. **Anais** [...] São Paulo: Even 3, 2020.

SOUSA, Renato Tarciso Barbosa de. **Classificação em arquivística**: trajetória e apropriação de um conceito. 2005. Tese (Doutorado em História Social) – Universidade de São Paulo, São Paulo, 2005.

WARLAND, Andrew, MOKHTAR, Umi Asma. Can technology classify records better than a human? *IRMS Bulletin*, n. 171, p. 16-19, 2013.

The background is a dark purple gradient with a complex geometric pattern. It features a network of thin white lines forming various polygons and triangles. Scattered throughout are larger, semi-transparent purple shapes, including diamonds and triangles. A dense field of small, light purple dots is visible in the lower-left and central areas, creating a textured effect.

Parte **2**

Abordagem aplicada  
na organização  
do conhecimento

# 8

Isabel Cristina Pereira dos Santos

**Reduções sintagmáticas e indexação:**  
uma análise a partir de *logs* de pesquisa  
dos usuários de um sistema  
de recuperação da informação

*Syntagmatic reductions and indexing:*  
*an analysis from user search logs*  
*of an information retrieval system*

DOI: 10.31560/pimentacultural/2022.95613.8

### Resumo:

Este capítulo se propõe a analisar o tratamento das reduções sintagmáticas no processo de indexação e a orientar metodologicamente como extrair e identificar essas reduções, tendo como fonte de terminologia os *logs* de pesquisa de usuários de um sistema de recuperação de informação (SRI) da área de Economia, estabelecendo-se, assim, uma forma de se trabalhar com *logs* no contexto da Organização do Conhecimento. Apresenta-se um instrumento para a identificação de lexias candidatas a variantes denominativas extraídas dos *logs* de pesquisa, baseado na teoria de Judit Freixa. Aplicam-se, como metodologias, a linguística de *corpus*, a garantia literária, mais especificamente, a garantia terminológica, a garantia do usuário e a garantia acadêmica. Analisam-se, detalhadamente, duas lexias da subárea de Microeconomia — *cobb douglas* e *poupança* —, por se entender que, entre os casos analisados na pesquisa, são os que mais podem enriquecer a discussão sobre reduções sintagmáticas. Conclui-se que a análise de *logs* e a metodologia de garantias são grandes aliadas para a indexação de reduções sintagmáticas, uma vez que podem trazer à luz relações semânticas entre termos e esclarecer conceitos que, quando reduzidos, podem perder sua especificidade e adquirir outro sentido.

**Palavras-Chave:** Reduções sintagmáticas; *logs*; indexação; variação denominativa; garantia terminológica; garantia do usuário.

**Abstract:**

*This chapter aims to analyze the treatment of syntagmatic reductions in the indexing process and to methodologically guide how to extract and identify these reductions, having as a source of terminology the search logs of users of an information retrieval system (IRS) in the area of Economy, thus establishing, a way of working with logs in the context of the Knowledge Organization. An instrument for the identification of lexias candidates for denominative variants extracted from the search logs of users of an IRS, based on the theory of Judit Freixa, is presented. The applied methodologies are the corpus linguistics, literary warrant, more specifically, terminological warrant, user warrant and academic warrant. Two lexias from the Microeconomics subarea — cobb douglas and poupança — are analyzed in detail, as it is understood that, among the cases analyzed in the research, they are the cases that can most improve the discussion on syntagmatic reductions. It is concluded that log analysis and the methodology of warrants are great allies for the indexing of syntagmatic reductions, as they can bring to light semantic relationships between terms and clarify concepts, which can lose their specificity and acquire another sense when reduced.*

**Keywords:** Syntagmatic reductions; logs; indexing; denominational variation; terminological warrant; user warrant.

## 1 INTRODUÇÃO

Este capítulo se propõe a analisar o tratamento das reduções sintagmáticas no processo de indexação. Ele é um recorte da dissertação de mestrado desta autora intitulada “Reconhecimento de termos e de variantes denominativas a partir dos *logs* de pesquisa dos usuários da *Revista Análise Econômica*: uma análise da variação no âmbito da Microeconomia” (2021). Além de abordar o uso de reduções de sintagmas, esse capítulo também se propõe a orientar, por meio de sua metodologia, como extrair e identificar essas reduções, tendo como fonte de terminologia os *logs* de pesquisa de usuários de um sistema de recuperação de informação, estabelecendo-se, assim, uma forma de se trabalhar com *logs* no contexto da Organização do Conhecimento.

Os *logs* são os arquivos que contêm o histórico de pesquisa dos usuários. Na dissertação supracitada, entre outros estudos, como os de Laipelt (2015a; 2015b), eles têm se mostrado uma importante fonte para a extração de terminologia e de variação denominativa, servindo inclusive como garantia do usuário no processo de legitimação dos termos. Os *logs* proporcionam, ao analista, uma terminologia *in vivo*, uma terminologia “pulsante”, uma vez que carregam consigo as expressões de busca dos usuários, também chamadas de *lexias*, que nada mais são do que a representação da linguagem natural, caracterizada por supressões de letras, acentos, preposições, siglas e partes do sintagma, além de apresentarem possibilidades de variações denominativas dos termos pesquisados.

No processo de indexação, as reduções sintagmáticas são um grande desafio ao indexador, uma vez que se admite não apenas um trabalho prescritivo, destinado a representar unidades unívocas, mas também um trabalho descritivo das unidades propostas pelo usuário, considerando o emprego de termos variantes como formas remissivas para a recuperação da informação. Nesse caso, as variantes denominativas

e/ou suas reduções precisam ser analisadas com bastante atenção, pois podem vir a ser realmente uma representação dos termos normalizados (uma redução do sintagma) ou um outro termo (um outro conceito). E, quando da extração de terminologia dos *logs* de pesquisa, essa relação entre sintagma e termo reduzido é bastante frequente na análise.

Para o reconhecimento das reduções sintagmáticas extraídas dos *logs*, esta pesquisa se baseou na tese de doutorado de Freixa (2002), que propõe uma análise exaustiva da variação terminológica. Fundamentada em bases teóricas da Teoria Comunicativa da Terminologia (TCT), a autora propõe uma classificação formal detalhada para a variação denominativa existente nos discursos especializados, mais precisamente, da área ambiental, dividindo essa classificação em cinco blocos, de acordo com o tipo de alteração linguística produzida. Nesse capítulo, serão destacados o segundo e o terceiro bloco, que tratam das variações morfossintáticas e das reduções sintagmáticas.

## 2 AS REDUÇÕES SINTAGMÁTICAS

Conforme os princípios metodológicos fundamentais da TCT, a categoria gramatical básica dos termos é a nominal, podendo apresentar unidades adjetivas e verbais de caráter terminológico relacionadas a esses termos nominais. Existem também locuções de valor terminológico compostas de preposições e sintagmas nominais. Em uma perspectiva denominativa, a maioria dos termos é substantivo (CABRÉ, 2001).

Em linguagens documentárias, como dicionários, glossários, tesouros, lista de palavras-chave, o termo preferido geralmente é um termo monoléxico, ou seja, um substantivo (ex.: *poupança*), ou um termo poliléxico, ou seja, um sintagma nominal (ex.: *caderneta de poupança*). Nas áreas de especialidade, os sintagmas nominais são mais numerosos, e há predominância de substantivos também, porém, da mesma forma, adjetivos podem adquirir *status* de termo (KRIEGER; FINATTO, 2017).

O uso da variação por redução nos *logs* de pesquisa se diferencia do uso da variação por redução no discurso. A variação no discurso, segundo Freixa (2013), dá-se principalmente por questões de estilo de escrita do autor, que tem por objetivo evitar repetição dos termos, usando outras denominações; economizar termos, fazendo uso de siglas, abreviaturas, acrônimos (variantes gráficas) e reduções por anáfora (variantes lexicais), conforme o exemplo:

A **teoria da escolha social** (*social choice*) estuda como as preferências individuais se agregam para formar uma preferência coletiva. **Essa teoria** remonta à publicação do paradoxo da votação, por Condorcet (1785) (RAE1704, 2020).<sup>58</sup>

Nesse exemplo, “essa teoria” retoma a “teoria da escolha social” anteriormente mencionada no discurso. Trata-se de um mecanismo de coesão, no qual “teoria” mantém o sentido original do sintagma, constituindo-se assim uma redução anafórica. Essa afirmação somente é possível porque o sintagma e sua redução coexistem em um mesmo contexto textual. Para mapear essas reduções presentes nos textos, é necessária uma leitura atenta de todo o texto, a fim de localizar as anáforas do termo preferido, ou seja, se ele foi repetido de forma reduzida no mesmo contexto ou em outro contexto no mesmo texto.

No entanto, para as variantes que ocorrem nos *logs*, não há um contexto textual, mas sim um contexto informacional, composto por sessões, históricos e expressões de busca ou lexias (Quadro 2 e 4), as quais são a representação gráfica de uma construção mental do usuário, sem revisões ou correções. Em suas estratégias de busca, o usuário cria associações semânticas e morfológicas nas suas tentativas de recuperar a informação e, dessas associações, podem surgir variantes denominativas que são reconhecidas pelo indexador apenas dentro de uma sessão de *logs*, sem um contexto textual especializado. A validação no contexto

58 RAE1704: VARGAS, J.; HERSCOVICI, A. A tragédia dos *commons* revisitada: uma análise crítica. **Revista Análise Econômica**, Porto Alegre, v. 35, n. 67, p. 105-128, mar. 2017. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/54987/40801> Acesso em: 05 out. 2019.

de uso é realizada pelo indexador durante o processo de garantia literária, que abrange a garantia terminológica (contexto textual), a garantia do usuário (contexto informacional) e a garantia acadêmica (consulta a especialistas), etapas subseqüentes ao reconhecimento terminológico.

Em termos de recuperação da informação, ao utilizar termos reduzidos, o usuário gera um “ruído” na recuperação, trazendo à tona um número elevado de documentos aleatórios (CHAUMIER, 1988, p. 63), uma vez que torna o termo de busca muito genérico. Em termos de indexação, o indexador preza sempre pelo termo mais específico, no caso, o sintagma, e, quando esse sintagma se reduz, a redução faz com que ele perca informações semânticas, perdendo assim o seu sentido original.

Por esse motivo, no processo de identificação de variantes denominativas presentes nos *logs*, é preciso direcionar uma maior atenção aos casos de reduções, principalmente aos casos de redução de extensão, como, por exemplo, em *finanças pessoais* / *finanças*, em que um termo poliléxico (*finanças pessoais*) se converte a um termo monoléxico (*finanças*), perdendo o seu carácter específico e tornando-se mais genérico. Nesse caso, claramente ocorre perda de sentido do termo poliléxico, uma vez que, sem um contexto textual no qual se possa identificar o carácter anafórico do termo reduzido (*finanças*), não é possível afirmar que esse termo seja uma variante do termo poliléxico (*finanças pessoais*).

Para identificar os tipos de variação que poderiam existir entre os pares de lexias extraídos dos *logs*, utilizou-se, como referencial teórico, a Classificação Formal de Variantes Denominativas de Freixa (2002). O Quadro 1 foi elaborado a partir do quadro de Freixa (2014) e abrange apenas os blocos II e III da classificação. Por meio do quadro de Freixa (2014), foi possível elaborar um instrumento para identificar lexias candidatas a variantes extraídas dos *logs* de pesquisa e fundamentar as possíveis variações existentes, tais como variações morfosintáticas entre termos monoléxicos/poliléxicos, reduções de extensão e de base e outros tipos de reduções.

Faz-se importante destacar que as lexias apresentadas no Quadro 1 são “candidatas a variantes” e foram extraídas da pesquisa de Santos (2021), exceto para o exemplo *producto ecológico / ecoproduto*, que pertence ao *corpus* de Freixa (2002)<sup>59</sup>. Ao contrário de Freixa (2002), que extraiu e validou suas variantes em um contexto textual antes de disponibilizá-las em um quadro, Santos (2021) apenas se baseou no quadro de Freixa (2014) para organizar as lexias extraídas dos *logs*, sem validá-las como variantes, pois, como mencionado, no caso dos *logs*, não há um contexto textual, mas sim um contexto informacional, devendo as lexias serem validadas na etapa de garantia, que ocorre posteriormente às suas inserções no quadro.

**Quadro 1 – Instrumento para a identificação de lexias candidatas a variantes denominativas extraídas dos *logs* de pesquisa dos usuários de um sistema de recuperação da informação – Variações morfofossintáticas e reduções**

II VARIAÇÕES MORFOSSINTÁTICAS	
<b>2. Estrutura diferente</b>	
f) Monoléxico / poliléxico	<i>producto ecológico / ecoproduto</i>
III REDUÇÕES	
<b>1. Reduções da extensão</b>	<i>finanças pessoais / finanças bolsa de valores / bolsa demanda marshalliana / demanda oaxaca blinder / oaxaca</i>
<b>2. Reduções da base</b>	<i>função de produção cobb-douglas / cobb-douglas função de produção de cobb-douglas / cobb-douglas função produção de cobb-douglas / cobb-douglas função de produção do tipo cobb-douglas / cobb-douglas caderneta de poupança / poupança conta de poupança / poupança</i>
<b>3. Outras reduções</b>	<i>teoria econômica evolucionista / teoria evolucionista</i>

Fonte: Adaptado de Freixa (2014, p. 316-317).

<sup>59</sup> No Quadro 1, nas categorias para as quais não foram encontrados exemplos de variação em Economia, tais como “Estrutura diferente: monoléxico / poliléxico”, mantiveram-se os exemplos propostos por Freixa (2014), em sua classificação original, por se entender que candidatas a variantes para essas categorias podem ser encontradas em outro *corpus* de estudo (SANTOS, 2021).

Para classificar os casos de redução, Freixa (2002) adotou três critérios: primeiro, conforme a parte da unidade em que ocorre a redução, se na extensão ou na base do sintagma; segundo, conforme o número de elementos reduzidos; e terceiro, conforme a estrutura do sintagma. As reduções de extensão são as mais comuns. De acordo com a autora, as reduções são principalmente anafóricas, ou seja, aquelas que, quando aplicadas em um contexto textual, retomam um sintagma usado anteriormente no mesmo discurso, como, por exemplo, *bolsa de valores / bolsa* e *função de produção cobb-douglas / cobb-douglas*, em que *bolsa* e *cobb-douglas* são os elementos anafóricos que retomam e remetem às formas completas do termo (sintagmas) que já foram mencionadas antes no contexto.

No Quadro 1, o bloco II trata das variações morfossintáticas, ou seja, das relações formais entre os constituintes do termo, que podem acarretar mudanças na estrutura do termo. No tipo “Estrutura diferente”, ocorre a alternância entre a estrutura substantivo e adjetivo (*producto ecológico*) e uma unidade monoléxica (*ecoproducto*), em que o adjetivo (*ecológico*) da unidade sintagmática é representado por um prefixo (*eco*) no monoléxico (*ecoproducto*) (FREIXA, 2002).

Laipelt (2015b), que classificou variantes denominativas da área do Direito do Trabalho e do Direito Previdenciário, chama a atenção para a necessidade de se fazer uma análise atenta, pois aspectos muito específicos da variação podem deixar dúvidas quanto às possibilidades de classificação de variantes em algumas categorias. Sobre as variações morfossintáticas com estruturas diferentes, as quais apresentam reduções, a autora explica o seguinte:

[...] ao compararmos duas variantes entre si, é preciso observar com atenção as alterações produzidas, visto que a presença ou ausência de um único aspecto é fator determinante para correta classificação. Por exemplo, qual a diferença entre uma alteração de estrutura morfossintática que ocasiona a mudança de uma unidade monoléxica para uma unidade poliléxica, e as

reduções de extensão, visto que na categoria reduções também encontramos unidades poliléxicas que, devido à redução, se tornam monoléxicas? A diferença está no resultado da alteração obtida. No caso das alterações morfossintáticas, um dos termos da unidade poliléxica, geralmente, o adjetivo, de acordo com Freixa (2002), se transforma em prefixo da forma monoléxica, sem que o termo perca sua especificidade, como se pode verificar no exemplo a seguir: *produto ecológico / ecoproduto*. No caso de várias alterações complexas, é o radical do substantivo que se torna prefixo da unidade monoléxica resultante, como pode-se verificar nesse exemplo: *Agricultura ecológica / Agroecologia* (LAPELT, 2015b, p. 50, grifo do autor).

O bloco III trata das variações denominativas oriundas das reduções dos sintagmas. Freixa (2002) explica que a redução pode ser entendida como a supressão de um ou mais constituintes de um sintagma, e que essa supressão contribui para a economia linguística, assim como ocorre com os acrônimos, mecanismo utilizado para encurtar um termo.

A aplicação do Quadro 1 pretende auxiliar no processo de reconhecimento dos termos, pois possibilita, ao indexador, organizar as lexias por categorias, realizar uma pré-análise e conjecturar possibilidades de variações que podem existir entre elas até uma confirmação final, que se dará por meio da aplicação das garantias. Tendo em vista a etapa de garantias, ao se analisar os pares de lexias dispostos no instrumento de identificação, deve-se fazer a seguinte pergunta: Cada lexia do par de lexias apresenta uma definição própria?

Se a resposta for “sim”, pode-se comparar as duas definições e verificar se o conceito é o mesmo; caso o conceito não seja o mesmo, considera-se que se trata de dois termos independentes, não sendo, portanto, variantes. Se a resposta for “não”, quer dizer que as definições são iguais, sendo os termos, portanto, variantes. Aqui, pode-se considerar as reduções como variantes também.

### 3 METODOLOGIA DA PESQUISA

Esta é uma pesquisa baseada em *corpus* e de natureza aplicada, pois visa ao emprego concreto em trabalhos no âmbito da Organização do Conhecimento. Para análise da terminologia e justificação das reduções, aplicaram-se, como metodologias, a linguística de *corpus*, a garantia literária, mais especificamente, a garantia terminológica, a garantia do usuário e a garantia acadêmica (BARITÉ, 2020, 2009, 2007; BARITÉ ROQUETA, 2011; BARITÉ, M. *et al.* 2010; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005; BEGHTOL, 1986). Nesse contexto, para que esta pesquisa possa ser replicada por pesquisadores que pretendam empreender trabalho semelhante, descrevem-se a seguir as etapas metodológicas<sup>60</sup>.

#### **Etapla 1 – Organização do corpus de estudo**

Realizou-se a extração das lexias dos *logs* de pesquisa dos usuários da *Revista Análise Econômica*, no período de 31 de maio de 2019 a 26 de novembro de 2020, totalizando 1.162 *logs*, para, a partir delas, selecionar candidatas a variantes denominativas. A extração foi realizada por meio de um extrator de *logs*<sup>61</sup>, o qual disponibiliza os *logs* em uma planilha em Excel, com extensão .csv, separados em colunas, que designam, respectivamente: o IP do usuário que fez a pesquisa no *site* da revista; o ano, o mês, o dia, a hora, o minuto, e o segundo em que o usuário fez a pesquisa; o país do acesso; e o termo pesquisado (sempre entre aspas).

#### **Etapla 2 – Organização do corpus de referência**

Realizou-se o *download* dos artigos publicados no *site* da *Revista Análise Econômica* no período 2015-2019. Cada artigo recebeu um

60 Para mais detalhes dessa metodologia, ver Santos (2021).

61 Elaborado pelo profissional da área de Informática Vicente Grassi Filho.

código correspondente à edição a qual pertence, a fim de identificar a referência do artigo durante a análise da colocação do termo em determinado contexto. Para a compilação do *corpus* de referência e extração dos termos, foi utilizado o *software Sketch Engine* (2021).

### Etapa 3 – Reconhecimento terminológico:

- 3.1) Análise das lexias de Economia extraídas dos *logs* de pesquisa a partir do instrumento para a identificação de lexias candidatas a variantes denominativas (Quadro 1), separando-se as lexias candidatas a variantes nos seguintes grupos: Grupo A – Lexias candidatas a variantes encontradas na mesma sessão de busca; e Grupo B – Lexias candidatas a variantes encontradas em diferentes sessões de busca;
- 3.2) Análise das lexias de Economia extraídas dos *logs* de pesquisa (*corpus* de estudo) e suas ocorrências no *corpus* textual especializado (*corpus* de referência), e aplicação do Quadro 1, separando-se as lexias candidatas a variantes no seguinte grupo: Grupo C – Lexias candidatas a variantes encontradas comparando-se o *corpus* de estudo com o *corpus* de referência;
- 3.3) Classificação das candidatas a variantes dos Grupos A, B e C conforme o Sistema de Classificação JEL (*Journal of Economic Literature*), método padrão de classificação da literatura acadêmica de Economia (AMERICAN ECONOMIC ASSOCIATION, 2020). A JEL foi usada como instrumento de apoio para o reconhecimento terminológico das candidatas a variantes, auxiliando na organização da classificação dos termos da Economia e no recorte de um subdomínio para fins de análise, uma vez que se obteve um número elevado lexias;
- 3.4) Reconhecimento do subdomínio com maior ocorrência de candidatas a variantes e seleção das candidatas pertencentes a ele, fazendo-se, assim, o recorte para análise.

### Etapa 4 – Garantia literária

Validação das candidatas a variantes nos dicionários e nos livros-texto de Microeconomia, subdomínio recortado.

#### Dicionários:

- a. SANDRONI, P. (Org.). *Dicionário de Economia do século XXI*. 8. ed. Rio de Janeiro; São Paulo: Record, 2016.
- b. HOUAISS, A. (Org.). *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

#### Livros-texto sobre a área recortada:

- a. GOOLSBEE, A.; LEVITT, S.; SYVERSON, C. *Microeconomia*. 2. ed. Trad. Teresa Cristina Padilha de Souza. São Paulo: Atlas, 2018.
- b. PINDYCK, R. S.; RUBINFELD, D. L. *Microeconomia*. 8. ed. Trad. Daniel Vieira. São Paulo: Pearson Education do Brasil, 2013.
- c. VARIAN, H. R. *Microeconomia: uma abordagem moderna*. 9. ed. Trad. Regina Célia Simille de Macedo. Rio de Janeiro: Elsevier, 2015.

#### Etapa 5 – Preenchimento das fichas terminológicas

A ficha foi adaptada conforme as necessidades dessa pesquisa. Para a sua elaboração, levaram-se em consideração as informações que um indexador precisa saber sobre um determinado termo.

#### Etapa 6 – Consulta a especialistas para validar as candidatas a variantes selecionadas

Seguindo as recomendações da Terminologia e da Organização do Conhecimento, essa pesquisa contou com a colaboração de dois docentes da área de Economia – o professor Ivan Colangelo Salomão, da Universidade Federal do Paraná (UFPR), e o professor Sabino da Silva Pôrto Junior, da Universidade Federal do Rio Grande do Sul (UFRGS) –, os quais gentilmente aceitaram participar do reconhecimento terminológico como consultores especialistas. Aplicada a metodologia proposta, foram escolhidas duas lexias para serem analisadas detalhada-

mente – *cobb douglas*, pertencente ao Grupo B de lexias, e *poupança*, pertencente ao Grupo C de lexias –, por se entender que são os casos que mais podem enriquecer a discussão sobre reduções sintagmáticas.

## 4 ANÁLISE DAS REDUÇÕES ENCONTRADAS EM LEXIAS CANDIDATAS A VARIANTES DOS GRUPOS B E C

Nesta seção, analisam-se, mais detalhadamente, duas lexias candidatas a variantes pertencentes a Categoria D – Microeconomia – da Classificação JEL: *cobb douglas* e *poupança*. A intenção dessa análise é verificar se ambas as lexias são reduções de sintagmas ou se são conceitos independentes.

### 4.1 LEXIA COBB DOUGLAS<sup>62</sup>

Esta análise pretende investigar se a lexia *cobb douglas* é uma redução sintagmática ou um conceito independente. Para tanto, o primeiro passo é olhar para o *corpus* de estudo e verificar os históricos de busca dos usuários. Conforme o Quadro 2, essa lexia aparece em outubro de 2020, no histórico de busca de um usuário, que será chamado de “Usuário 2”.

62 O termo eponímico *cobb-douglas* é formado pela junção dos sobrenomes do economista Paul Douglas e do matemático Charles Cobb. Douglas (1892-1976) foi professor da Universidade de Chicago e senador dos Estados Unidos (EUA); Cobb (1875-1949) foi professor da Faculdade de Amherst, em Massachusetts, EUA (VARIAN, 2015).

Quadro 2 – Contexto informacional da lexia *cobb douglas* (Grupo B)

Usuário 2								
Lexia	IP*	Ano	Mês	Dia	Hora	Min.	Seg.	Local
“economia da produção”	177.129.xx.yyy	2020	10	21	16	18	30	Viçosa – MG
“cobb douglas”	<b>177.129.xx.yyy</b>	<b>2020</b>	<b>10</b>	<b>21</b>	<b>16</b>	<b>19</b>	<b>29</b>	<b>Viçosa – MG</b>
“elasticidade de produção”	177.129.xx.yyy	2020	10	21	16	19	47	Viçosa – MG
“CES”	177.129.xx.yyy	2020	10	21	16	24	7	Viçosa – MG

Fonte: Elaborado pela autora, 2021.

Nota: \* Por questão de privacidade e anonimato do usuário, optou-se por substituir os últimos números do endereço IP por letras.

Ao se observar o Quadro 2, o segundo passo é buscar por *cobb douglas* no *Dicionário de Economia do século XXI* (2016), a fim de conhecer a sua definição:

**COBB-DOUGLAS (Função de Produção)**. Uma função com a fórmula  $Q = A.L^a.K^b$ , em que Q é a produção, A, a e b são constantes e L e K são, respectivamente, o trabalho e o capital. [...]. Veja também **Função Homogênea**. (SANDRONI, 2016, p. 276, grifo do autor).

**FUNÇÃO DE PRODUÇÃO COBB-DOUGLAS**. Veja **Cobb-Douglas (Função de Produção)**. (SANDRONI, 2016, p. 662, grifo do autor).

Pelas definições do dicionário, é possível verificar que as lexias *cobb douglas* / *função de produção* se relacionam semanticamente. Ao se olhar para os *logs* novamente, verifica-se que outro usuário, que será chamado de Usuário 1, já havia pesquisado por *função de produção* em junho de 2019, conforme o Quadro 3.

Quadro 3 – Contexto informacional da lexia *função de produção* (Grupo B)

Usuário 1								
Lexia	IP*	Ano	Mês	Dia	Hora	Min.	Seg.	Local
“micro”	200.235.xxx.yy	2019	6	26	13	51	26	Viçosa – MG
“função de produção”	<b>200.235.xxx.yy</b>	<b>2019</b>	<b>6</b>	<b>26</b>	<b>13</b>	<b>53</b>	<b>59</b>	<b>Viçosa – MG</b>
“microeconomia”	200.235.xxx.yy	2019	6	26	14	10	12	Viçosa – MG

Fonte: Elaborado pela autora, 2021.

Nota: \* Por questão de privacidade e anonimato do usuário, optou-se por substituir os últimos números do endereço IP por letras.

Ou seja, as lexias *cobb douglas* / *função de produção* foram pesquisadas por dois usuários diferentes, sendo, por isso, classificadas no Grupo B. Nesse exemplo, uma vez que se trata de termos especializados bastante restritos ao domínio da Economia, e não há um parentesco formal entre as lexias que torne visível uma relação entre elas, nem a analista possui um conhecimento prévio e extratextual capaz de chegar a uma constatação, a busca por definições se tornou uma prática imprescindível para que se pudesse criar possíveis relações entre as lexias. Porém, somente será possível legitimar essa relação como uma relação entre variantes nas etapas “Garantia Literária” e “Consulta a especialistas”.

Pelas entradas do dicionário, observa-se que a lexia *cobb douglas* tem como variantes *cobb-douglas* (*função de produção*) e *função de produção cobb-douglas*. Em ambas, o termo eponímico *cobb-douglas* apresenta hífen, diferente da expressão reduzida *cobb douglas* proposta pelo usuário. No primeiro caso, que aparenta ser o termo preferido, pois apresenta a definição, a estrutura do termo *cobb-douglas* (*função de produção*) causa estranhamento ao consulente, uma vez que a expressão *função de produção* aparece entre parênteses, dando a impressão de que ela não seria parte integral do termo, mas uma variante. Na verdade, a expressão *função de produção* não deveria aparecer entre parênteses na estrutura do termo, uma vez que, como será visto a seguir, ela constitui outro termo, com sentido próprio.

Já, no segundo caso, desfaz-se essa ambiguidade: em *função de produção cobb-douglas*, o termo descritivo antecede ao eponímico, formando um sintagma com um sentido completo, que leva o consultante a constatar que *cobb douglas* seria um “tipo de” função de produção, não uma variação por redução de base, como leva falsamente a crer a primeira entrada. Essa entrada, contrariando a escolha do dicionarista, poderia ocupar satisfatoriamente o lugar de termo preferido. Como se verá mais adiante nas colocações, os usuários e especialistas não costumam usar parênteses na formação de termos, sendo *cobb-douglas (função de produção)* uma construção bastante precária, pois este não seria o termo mais usual nos contextos de uso.

No entanto, é preciso verificar também se, para a lexia *função de produção*, existe uma definição própria. Caso exista, *cobb douglas / função de produção* não serão variantes. Veja-se:

**FUNÇÃO DE PRODUÇÃO.** É a relação entre a produção de um bem e os insumos ou fatores de produção necessários para produzi-lo. Uma função de produção pode ser apresentada na forma genérica  $Q = f(L, K, t)$ , em que  $Q$  é o produto,  $L$  é a força de trabalho,  $K$  é o capital e  $t$  é o progresso técnico. Outros fatores de produção, como as matérias-primas, podem fazer parte também da função de produção. [...]. Veja também **Cobb-Douglas (Função de Produção)**. (SANDRONI, 2016, p. 661, grifo do autor).

Como se pode perceber, o dicionário apresenta também entrada para *função de produção*. Apesar de as definições não deixarem clara para um leigo a diferença entre os termos, conclui-se que, por terem entradas diferentes, *cobb douglas/função de produção* representam conceitos diferentes.

No entanto, para sanar dúvidas quanto à definição, foi preciso ir mais além. Nesse caso, os livros-texto de Microeconomia se mostraram grandes aliados. Segundo a literatura microeconômica, a princípio, a função Cobb-Douglas foi aplicada para se estudar o comportamento da produção, sendo **um dos tipos mais comuns de**

**função de produção** utilizados pelos economistas (GOOLSBEE; LEVITT; SYVERSON, 2018; VARIAN, 2015, grifo nosso). Todavia, os livros revelaram uma nova possibilidade para a lexia *cobb douglas* proposta pelo usuário: o termo reduzido *cobb-douglas* não se refere apenas à *função de produção*, mas se refere também a outros termos, que o incorporam em sua estrutura, conforme o Quadro 4.

**Quadro 4 – Colocações do termo reduzido *cobb-douglas* nos livros-texto de Microeconomia**

Referência	Terminologia
<b>Goolsbee, Levitt e Syverson (2018)</b>	<i>função Cobb-Douglas</i> <i>função utilidade convencional de Cobb-Douglas</i> <i>função utilidade de Cobb-Douglas</i> <i>função produção de Cobb-Douglas</i> <i>fórmula funcional de Cobb-Douglas</i> <i>equação padrão de Cobb-Douglas</i> <i>funções de Cobb-Douglas</i>
<b>Varian (2015)</b>	<i>preferências Cobb-Douglas</i> <i>função de utilidade Cobb-Douglas</i> <i>função de utilidade de Cobb-Douglas</i> <i>curvas de indiferença Cobb-Douglas</i> <i>função Cobb-Douglas</i> <i>forma Cobb-Douglas</i> <i>forma funcional Cobb-Douglas</i> <i>consumidor Cobb-Douglas</i>
<b>Varian (2015)</b>	<i>funções de demanda Cobb-Douglas</i> <i>demanda Cobb-Douglas</i> <i>função de produção de Cobb-Douglas</i> <i>isoquantas de Cobb-Douglas</i> <i>função oferta da empresa Cobb-Douglas</i> <i>tecnologia de Cobb-Douglas</i>
<b>Pindyck e Rubinfeld (2013)</b>	<i>função utilidade de Cobb-Douglas</i> <i>função de utilidade Cobb-Douglas</i> <i>função de produção de Cobb-Douglas</i> <i>funções de produção de Cobb-Douglas</i>

Fonte: Elaborado pela autora, 2021.

Com os resultados apresentados no Quadro 4, não se pode afirmar que todo *cobb-douglas* está relacionado à *função de produção*,

pois ele pode se referir também à *função utilidade de Cobb-Douglas*, à *forma funcional Cobb-Douglas*, a *preferências Cobb-Douglas*, a *isoquantas de Cobb-Douglas*, à *tecnologia de Cobb-Douglas* etc. Daí depreende-se que a *lexia cobb douglas*, proposta pelo usuário, seja uma **redução** de vários outros termos. Esses resultados fizeram com que se repensasse a proposta do usuário registrada nos *logs* de pesquisa, ou seja, o que realmente o usuário gostaria de recuperar quando utilizou a **redução cobb douglas** em sua estratégia de busca?

Para responder essa questão, é preciso compreender a diferença entre os termos *função produção de Cobb-Douglas* e *função utilidade de Cobb-Douglas*. Pela literatura microeconômica, entende-se, grosso modo, que *função produção de Cobb-Douglas* e *função utilidade de Cobb-Douglas* são representações que se diferenciam por seus objetivos e pelas variações da forma funcional Cobb-Douglas, a qual sofre adaptações de suas variáveis conforme os objetivos de análise econômica.

Entre as infinitudes de possibilidades relacionadas no Quadro 4, cabe ao indexador retornar ao histórico de busca do usuário, a fim de verificar se ele utilizou outros termos em sua estratégia que podem delimitar o conceito pretendido quando da pesquisa do termo reduzido. Nesse contexto, no Quadro 2, verificou-se que as *lexias* relacionadas a *cobb douglas* pelo usuário foram *economia da produção*, *elasticidade de produção* e *CES* (sigla em inglês para *elasticidade de substituição constante*), ou seja, todas pertencentes à teoria da produção, o que conecta imediatamente a *lexia cobb douglas* ao termo *função de produção cobb-douglas* e descarta a sua conexão com *função de utilidade Cobb-Douglas*, termo pertencente à teoria do consumidor.

Em síntese, pensando-se em termos de indexação, se, por um lado, a *lexia cobb douglas*, proposta pelo usuário, não pode ser considerada redução de base dos sintagmas dispostos no Quadro 4, porque o usuário, ao reduzir o sintagma pretendido, deu a ele um

sentido indefinido, devido às inúmeras possibilidades discursivas (Quadro 4), por outro lado, o contexto informacional da lexia delimitou a sua análise neste estudo à sua relação com a função de produção, podendo ele, nesse contexto, ser considerado uma redução de *função de produção cobb-douglas*.

## 4.2 LEXIA POUPANÇA

As lexias que constituem o Grupo C são as que, na comparação do *corpus* de estudo com o *corpus* de referência, surgiram como candidatas a variantes. No Quadro 5, apresenta-se o histórico de busca do usuário.

Quadro 5 – Contexto informacional da lexia *poupança* (Grupo C)

Lexia	IP*	Ano	Mês	Dia	Hora	Min.	Seg.	Local
“investimento”	200.137.xx.yy	2019	7	16	17	47	21	Vitória – ES
“poupança”	200.137.xx.yy	2019	7	16	17	50	14	Vitória – ES

Fonte: Elaborado pela autora, 2021.

Nota: \* Por questão de privacidade e anonimato do usuário, optou-se por substituir os últimos números do endereço IP por letras.

De acordo com os *logs*, o usuário utilizou a expressão *poupança* em sua estratégia de busca. Por não haver candidatas a variantes na mesma sessão de busca, nem em sessões de busca diferentes, aplicou-se a regra do Grupo C de lexias, em que se buscaram colocações de *poupança* no *corpus* formado pelos artigos da revista, de onde foi possível estabelecer os pares de lexias *poupança / caderneta de poupança* e *poupança / conta de poupança*. A intenção aqui é saber se a lexia *poupança*, proposta pelo usuário, é uma redução de base de *caderneta de poupança* e *conta de poupança* ou se representa outro conceito.

O primeiro passo é verificar a garantia literária; se há definições para esses termos no dicionário especializado, no dicionário do léxico geral e nos livros-texto de Microeconomia. Veja-se:

**POUPANÇA.** Em economia, parte da renda nacional ou individual que não é utilizada em despesas, sendo guardada e aplicada depois de deduzidos os impostos. [...]. Veja também **Consumo**; **Investimento**. (SANDRONI, 2016, p. 1271, grifo do autor).

**<sup>1</sup>poupança** s.f. 1 ato ou efeito de poupar 2 despesa moderada; economia. 3 *infrm.* característica do que não gosta de gastar e prefere amealhar; sovínice 4 ECON fração da renda nacional ou individual que não é aplicada em serviços e bens de consumo. ETIM *poupar* + *ança*. PAR. *popança* (s.f) **<sup>2</sup>poupança** s.f. ECON red. de caderneta de poupança. SIN/VAR ver sinonímia de *pecúlio* (HOUAISS, 2009, p. 1534, grifo do autor).

**CADERNETA DE POUPANÇA.** Contas sobre cujos depósitos são creditados mensalmente (lei de agosto de 1983) juros e correção monetária, uma vez observada a condição de que saques e depósitos sejam feitos em épocas predeterminadas. [...]. Veja também **BNH**; **Desilusão Monetária**; **Plano Collor**; **Plano Real**; **Plano Verão** (SANDRONI, 2016, p. 191, grifo do autor).

**caderneta de poupança** ECON tipo de depósito bancário em que a quantia não movimentada por determinado período de tempo rende juros e recebe correção monetária. ETIM *caderno* + *eta* \ê\ (HOUAISS, 2009, p. 357, grifo do autor).

Sumariamente, observa-se que *poupança*, na área de Economia, apresenta duas definições:

- a. “[...] parte da renda nacional ou individual que não é utilizada em despesas, sendo guardada e aplicada depois de deduzidos os impostos” (SANDRONI, 2016, p. 1271); “fração da renda nacional ou individual que não é aplicada em serviços e bens de consumo” (HOUAISS, 2009, p. 1534); e
- b. redução do termo *caderneta de poupança* (HOUAISS, 2009, p. 1534).

A definição “b” para *poupança*, oriunda do *Dicionário Houaiss*, dicionário do léxico geral, apresenta dois problemas: primeiro, não há um texto que a defina e, para “decifrá-la”, o consulente precisa verificar a definição de *caderneta de poupança*, a fim de confirmar se realmente *poupança* pode ser considerada uma redução desse termo; segundo, tendo *poupança* a sua própria definição no *Houaiss*, para a Terminologia, ela seria outro termo, outro conceito, não podendo assim ser considerada uma redução de *caderneta de poupança*, ainda mais estando etiquetada na mesma área de conhecimento, que é a Economia.

A redução proposta pelo *Dicionário Houaiss* é plenamente aceitável em contextos informais, em que o intercâmbio entre *poupança* e *caderneta de poupança* assume o mesmo sentido; no entanto, em contextos especializados, esses termos assumem sentidos específicos, objetivando cada um a sua univocidade, respeitando o contexto comunicativo da Economia, área de conhecimento em que ele está inserido.

Pindyck e Rubinfeld (2013) definem *caderneta de poupança* como um ativo isento de risco, que paga juros (geralmente mensais) que são reinvestidos na própria conta. Dos três livros-texto de Microeconomia consultados, esses autores foram os únicos que usaram o termo em seu texto:

Às vezes, é necessário decidir o grau de risco que estamos dispostos a assumir. Por exemplo, o que você deveria fazer com suas economias? Será que deveria investir seu dinheiro em algo seguro, como uma **caderneta de poupança**, ou em algo mais arriscado, porém com maior potencial de rendimentos, como o mercado de ações? (PINDYCK; RUBINFELD, 2013, p. 151).

Também nos artigos da *Revista Análise Econômica* foram encontradas três ocorrências para o termo *caderneta de poupança*:

[...] o financiamento imobiliário depende essencialmente de recursos direcionados da **caderneta de poupança** e do FGTS, cujos montantes, em grande parte, são provenientes de bancos

públicos (a Caixa Econômica Federal detém aproximadamente 70% desse mercado); [...] (RAE1616, 1983, grifo nosso)<sup>63</sup>.

Nas obras consultadas, não se encontrou definição para o termo *conta de poupança*, porém se encontraram ocorrências para ele. Dos três livros-texto consultados, Goolsbee, Levitt e Syverson (2018) foram os únicos que usaram esse termo em seu texto:

Suponha que, quando você tinha um ano de idade, sua avó tenha lhe dado uma pequena moeda de prata de um dólar. Seus pais colocaram essa moeda de prata de um dólar em uma **conta de poupança** com taxa de juros garantida de 9%, e logo depois esqueceram isso (GOOLSBEE; LEVITT; SYVERSON, 2018, p. 1393, grifo nosso).

Também no universo de artigos da *Revista Análise Econômica* foram encontradas duas ocorrências para *conta de poupança*:

Uma evolução semelhante ocorreu com o número de clientes com **conta de poupança**, que registrou alta de 41,1% na mesma base de comparação, alcançando 82,1 milhões no final de 2007 (58,2 milhões em 2002) (RAE1506, 1983, grifo nosso)<sup>64</sup>.

Após essa reflexão, para se entender os conceitos dos termos *caderneta de poupança* e *conta de poupança*, pelos contextos apresentados, entende-se que elas são variantes entre si. Sofrem, apenas, alteração de base *caderneta* / *conta*, não alterando o sentido original do sintagma caso sejam intercambiadas no mesmo contexto de uso.

Retornando a análise do conceito *poupança*, o segundo passo agora é verificar a garantia do usuário. Nesse caso, os *logs* de pesquisa podem auxiliar o indexador a conhecer qual o sentido de *poupança*

63 RAE1616: ARAUJO, A.; FERRARI FILHO, F.; BUENO, E. Existe uma bolha imobiliária no Brasil? Uma análise teórica e empírica. *Revista Análise Econômica*, Porto Alegre, v. 34, n. 66, p. 149-172, set. 2016. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/54872/38539>. Acesso em: 20 out. 2019.

64 RAE1506: CAGNIN, R. F.; FREITAS, M. C. P. Tributação das transações financeiras: a experiência brasileira com o IOF e a CPMF. *Revista Análise Econômica*, Porto Alegre, v. 33, n. 63, p. 139-169, mar. 2015. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/35893/33366>. Acesso em: 20 out. 2019

que o usuário está buscando. De acordo com os *logs*, o usuário utilizou a expressão *poupança* em sua estratégia de busca, imediatamente após a expressão *investimento* (Quadro 5). Segundo Goolsbee, Levitt e Syverson (2018), pela definição econômica,

[investimento] implica a aquisição de algum tipo de capital, na esperança de um retorno futuro. Todavia, [...], poupança e investimento [...] são, de fato, interligados. Essa ligação resulta do funcionamento dos mercados de capitais. Essencialmente, os poupadores proporcionam fundos que os investidores utilizam para comprar capital. Os investidores compensam os poupadores pagando de volta uma parcela dos retornos do investimento. (GOOLSBEE; LEVITT; SYVERSON, 2018, p. 1393).

Relacionando-se essas duas lexias, tudo leva a crer que o usuário estaria pensando *poupança* como um “tipo de” investimento. Nesse caso, acredita-se que o usuário, assim como o *Dicionário Houaiss*, não faz distinção entre a lexia *poupança* e os sintagmas *caderneta de poupança* / *conta de poupança*. O que é compreensível, pois a tendência do usuário de um sistema de informação é reduzir o termo de busca. No entanto, observou-se que, nos três livros-texto de Microeconomia consultados, nenhum dos autores utiliza o termo *poupança* com o sentido de *caderneta de poupança*, mas sim com o sentido de parte de uma renda que não é utilizada. Isso também ocorre com os artigos da revista:

Mais especificamente, a contribuição de Keynes pode ser entendida a partir do debate sobre a relação **poupança**/investimento e a problemática do financiamento (RAE1508, 1983, grifo nosso).<sup>65</sup>

Portanto, o timing dos impostos tem impactos sobre as decisões de consumo e **poupança** das famílias. (RAE1520, grifo nosso)<sup>66</sup>.

65 RAE1508: MISSIO, F. J.; JAYME JR., F. G.; OLIVEIRA, A. M. H. C. Desenvolvimento financeiro e crescimento econômico: teoria e evidência empírica para as unidades federativas do Brasil (1995-2004). *Revista Análise Econômica*, Porto Alegre, v. 33, n. 63, p. 191-227, mar. 2015. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/39374/33368>. Acesso em: 20 out. 2019.

66 RAE1520: MARQUES JÚNIOR, L. S. Equivalência ricardiana e os efeitos da política fiscal na economia brasileira. *Revista Análise Econômica*, Porto Alegre, v. 33, n. 64, p. 215-241, set. 2015. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/41890/34912>. Acesso em: 20 out. 2019.

À medida que o investimento é realizado, a atividade industrial é estimulada e geram-se rendas que serão divididas entre consumo e **poupança**. (RAE1701, grifo nosso)<sup>67</sup>.

O terceiro passo da análise corresponde à garantia acadêmica ou consulta a especialistas. Nas palavras do professor Ivan Colan-gelo Salomão, especialista consultado, “*Caderneta* ou *conta de poupança* não são exatamente a mesma coisa que *poupança*. As duas primeiras podem ser lidas como um tipo de investimento, ao passo que o termo *poupança*, desacompanhado, significa todo e qualquer tipo de não gasto, ou seja, todos os recursos que não são empenhados em despesas ou consumos.”

A opinião do especialista vem a corroborar as informações levantadas nos *corpora* de estudo e de referência. Em suma, nos contextos de uso da área de Economia, *poupança* não pode ser considerada uma redução dos termos *caderneta de poupança* e *conta de poupança*, uma vez que representa um conceito diferente desses dois termos, tendo sua própria definição.

## 5 CONSIDERAÇÕES FINAIS

A indexação de reduções sintagmáticas em listas de remissivas deve ser um processo extremamente cuidadoso, uma vez que nem toda redução pode ser considerada uma variante denominativa do termo preferido. Este, na maioria das vezes, quando reduzido, adquire outro conceito, como no caso da *lexia poupança*.

Partindo-se dos *logs* de pesquisa, conforme o usuário cria estratégias de busca, ele tende a reduzir os termos. A primeira impressão é

67 RAE1701: AIDAR, G. L.; TERRA, F. H. B. A Teoria da Firma pós-keynesiana: uma revisão dos elementos relevantes. **Revista Análise Econômica**, Porto Alegre, v. 35, n. 67, p. 21-45, mar. 2017. Disponível em: <https://seer.ufrgs.br/AnaliseEconomica/article/view/55545/40798>. Acesso em: 03 out. 2019.

a de que entre duas lexias como *poupança / caderneta de poupança* ocorre variação por redução de base, pois se trata de um termo monoléxico e de um sintagma, em que há grande possibilidade de ocorrer uma variação desse tipo. Entretanto, a análise terminológica mostrou que o termo *poupança* adquire conceitos diferentes nos contextos especializado e não especializado, não se confirmando como uma redução sintagmática de *caderneta de poupança / conta de poupança* no domínio da Economia. Nesse caso, a consulta ao especialista foi decisiva para a confirmação dos conceitos.

A lexia *cobb douglas* apresentou uma análise bastante complexa, uma vez que no decorrer do estudo, surgiu uma gama de possibilidades de sintagmas para a redução proposta pelo usuário. Frente a tantas possibilidades, somente o contexto informacional dos *logs* foi capaz de suprir as ambiguidades trazidas por essa redução, uma vez que ele está implícito na consulta. É o contexto informacional, ou seja, as relações semânticas entre as lexias em uma mesma sessão de busca, que pode vir a determinar o assunto pretendido na recuperação.

No caso, o conjunto de lexias *cobb douglas*, *economia da produção*, *elasticidade de produção* e *CES* (sigla em inglês para *elasticidade de substituição constante*) constitui um grupo de termos empregados em teoria da produção, os quais determinaram a relação entre *cobb douglas* e *função de produção*, em detrimento à relação *cobb douglas* e *função de utilidade* e *cobb douglas*. Na aplicação da metodologia, ficou evidente que os *logs* podem servir como garantia do usuário em justificativas para o emprego de determinado termo em detrimento de outro, mas isso requer uma análise diferenciada para cada termo devido às suas especificidades.

Em síntese, se existe um *corpus* que dê sentido ao termo pelo seu contexto de uso, consegue-se verificar se um termo é uma redução por ele ser um elemento anafórico, ou seja, um elemento de retomada do termo completo ou sintagma. No entanto, se a finalidade for

a inclusão do termo em um catálogo de bibliotecas, recomenda-se tomar muito cuidado com as reduções, justamente porque nos *logs*, como já mencionado, não existe um contexto de uso; as reduções podem ser variantes ou não, e o uso de termos simples ou reduzidos pode gerar ruídos para o usuário durante a recuperação da informação, uma vez que amplia a sua estratégia de busca, recuperando documentos fora do assunto realmente desejado.

## 6 REFERÊNCIAS

AMERICAN ECONOMIC ASSOCIATION. **JEL Classification System/EconLit Subject Descriptors**. Fev. 01, 2020. Disponível em: <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>. Acesso em: 07 jan. 2021.

BARITÉ, M. Garantía literaria y normas para construcción de vocabularios controlados: aspectos epistemológicos y metodológicos. **Scire**, v. 15, n. 2, p. 13-24, jul.-dic. 2009.

BARITÉ, M. La garantía literaria: vigencia y proyección teórico-metodológica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais...** Salvador: Ancib, 2007.

BARITÉ, M. *et al.* Garantia literária: elementos para uma revisão crítica após um século. **TransInformação**, Campinas, v. 22, n. 2, p. 123-138, maio/ago. 2010.

BARITÉ ROQUETA, M. Garantía académica. *In*: BARITÉ ROQUETA, M. **La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento**: modelo y aplicación. 2011. 382 f. Tesis (Doctorado en Información Científica) – Departamento de Biblioteconomía e Documentación, Facultad de Documentación y Comunicación, Universidad de Granada, Granada, 2011.

BEGHTOL, C. Semantic validity: concepts of warrant in bibliographic classification systems. **Library Resources & Technical Services**, v. 30, n. 2, p. 109-125, Apr. – Jun. 1986. Disponível em: <https://utexas.instructure.com/courses/1160641/files/38750130/download?verifier=F7wJRuw5ku8BSNKMe-fyIbtPERFBVykHRIvWkDVv&wrap=1>. Acesso em: 01 out. 2020.

CABRÉ, M. T. Sumario de principios que configuran la nueva propuesta teórica. *In*: CABRÉ, M. T.; FELIU, J. (Ed.). **La terminología científico-técnica**:

reconocimiento, análisis y extracción de información formal y semántica. Barcelona: IULA, UPF: 2001. p. 19-25.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

FREIXA, J. La variación denominativa en terminología: tipos y causas. In: ISQUERDO, A. N.; CORNO, G. O. M. (Org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. Campo Grande: UFMS, 2014. p. 311-329.

FREIXA, J. **La variació terminològica**: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient. Tese (Doutorado) – Universidade de Barcelona, Barcelona, 2002.

FREIXA, J. Otra vez sobre las causas de la variación denominativa. **Debate Terminológico**, n. 9, p. 38-46, feb. 2013.

GOOLSBEE, A.; LEVITT, S.; SYVERSON, C. **Microeconomia**. 2. ed. Trad. Teresa Cristina Padilha de Souza. São Paulo: Atlas, 2018.

HOUAISS, A. (Org.). **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009.

KINGHOST. **Localizar IP (GeoIP)**. Disponível em: <https://king.host/wiki/geoip>. Acesso em: 12 jan. 2021.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à terminologia**: teoria & prática. 2. ed. São Paulo: Contexto, 2017.

LAIPELT, R. C. F. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, Porto Alegre, v. 21, n. 3, p. 150-170, set./dez. 2015a. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/59806/36047>. Acesso em: 27 maio 2020.

LAIPELT, R. C. F. **Metodologia para seleção de termos equivalentes e descritores de tesouros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015b. 213 f. Tese (Doutorado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2015b.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the construction, format and management of monolingual controlled vocabularies**: an American National Standards developed by the National Information Standards Organization. Bethesda: NISO Press, 2005.

PINDYCK, R. S.; RUBINFELD, D. L. **Microeconomia**. 8. ed. Trad. Daniel Vieira. São Paulo: Pearson Education do Brasil, 2013.

REVISTA ANÁLISE ECONÔMICA. Porto Alegre: FCE/UFRGS, 1983. Quadrimestral. Disponível em: <https://www.ufrgs.br/rae>. Acesso em: 19 jul. 2020.

SANDRONI, P. (Org.). **Dicionário de Economia do século XXI**. 8. ed. Rio de Janeiro; São Paulo: Record, 2016.

SANTOS, I. C. P. **Reconhecimento de termos e de variantes denominativas a partir dos logs de pesquisa dos usuários da Revista Análise Econômica**: uma análise da variação no âmbito da Microeconomia. 2021. 260 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.

SKETCH ENGINE: Language *corpus* management and query system. Disponível em: <https://www.sketchengine.eu>. Acesso em: 23 jan. 2021.

VARIAN, H. R. **Microeconomia**: uma abordagem moderna. 9. ed. Trad. Regina Célia Simille de Macedo. Rio de Janeiro: Elsevier, 2015.

# 9

Carine Melo Cogo Bastos

Thiago Henrique Bragato Barros

## As taxonomias no contexto da arquivologia

*Taxonomies  
in the context of archives*

DOI: 10.31560/pimentacultural/2022.95613.9

### Resumo:

A organização do conhecimento encontra-se no bojo da atividade de representação e acesso aos documentos arquivísticos. As discussões, neste capítulo, apresentam a relação da Arquivologia com a Organização do Conhecimento. Este trabalho é sobre a organização do conhecimento, relacionando conceitos de autores da área e de áreas próximas à Arquivologia. Discorre igualmente, sobre sistemas de organização do conhecimento na disciplina Organização do Conhecimento. As abordagens aqui apresentadas buscam contextualizar as taxonomias na Arquivologia. Por fim, apresenta conceitos sobre a representação do conhecimento e os sistemas de organização, aprofundando a taxonomia como um instrumento de gestão na organização do conhecimento no contexto da Arquivologia, com os aportes teóricos da Arquivologia e Organização do Conhecimento dos seguintes autores: Araújo (2014), Barité (2001), Borko (1968), Barros (2016), Barros e Sousa (2019), Brascher e Café (2008), Campos e Gomes (2007), Dahlberg (1978, 2006), Guimarães (2005, 2008, 2012), Hjørland (2002, 2008), Navarro (1995), Smit (2012), Sousa e Araújo (2007, 2017) e Vitoriano (2017).

**Palavras-Chave:** Arquivologia; organização do conhecimento; taxonomias.



**Abstract:**

*The organization of knowledge is at the heart of the activity of representation and access to archival documents. The discussions in this chapter present the relationship of Archivology to the Knowledge Organization. This work is about the organization of knowledge, relating concepts from authors in the area and from areas close to archivology. It also discusses knowledge organization systems in the Knowledge Organization discipline. The approaches presented here seek to contextualize taxonomies in archivology. Finally, it presents concepts about knowledge representation and organization systems, deepening taxonomy as a management tool in the organization of knowledge in the context of archivology. With theoretical contributions from Archival and Knowledge Organization by the following authors: Araújo (2014), Barité (2001), Borko (1968), Barros (2016), Barros e Sousa (2019), Brascher e Café (2008), Campos e Gomes (2007), Dahlberg (1978, 2006), Guimarães (2005, 2008, 2012), Hjørland (2002, 2008), Navarro (1995), Smit (2012), Sousa e Araújo (2007, 2017) e Vitoriano (2017).*

**Keywords:** *Archival science; knowledge organization; taxonomies.*

## 1 INTRODUÇÃO

Um arquivo institucionalizado passível de organização será preservado para um determinado grupo ou instituição a fim de que se garanta a proteção de direitos individuais e da memória coletiva. Será por meio da organização do conhecimento e do domínio do contexto de produção documental que o arquivista, dentro do quadro funcional de um arquivo, conseguirá estruturar conhecimento, modelar sistemas de organização e disponibilizar o acesso ao acervo arquivístico. Os documentos de arquivo permitem que os apagamentos de um determinado grupo venham a ficar em evidência, fazendo com que as informações e a memória social possam emergir através da organização do conhecimento.

Como decorrência tanto do impacto das reflexões da Arquivologia integrada quanto das tecnologias da informação e dos tipos especiais de arquivo, algumas mudanças ocorreram: o arquivo passou a ser cada vez mais caracterizado como um sistema de informação, e o objeto de estudo da Arquivologia passou a ser a informação arquivística (ARAÚJO, 2014). Outra forma de entender informação é pensá-la “[...] como fenômeno social coletivo, estruturas de conhecimento e instituições de memória das comunidades” (ALMEIDA *et al.*, 2007, p. 23). Desta maneira,

[...] estudos recentes vêm mostrando a necessidade de estudar os arquivos como construções sociais, e para tanto, “a arquivologia deveria deixar de ser vista como uma ciência que ‘documenta a verdade dos fatos’” para ser uma ciência voltada para a discussão dos documentos como resultado e reflexo de uma multiplicidade de tensões sociais que se produzem em tempo e lugar determinados, isto é, que vê os documentos como construtos produzidos segundo determinadas regras do discurso, determinadas “condições de documentar”. Nessa mesma perspectiva, cada vez mais vêm sendo desenvolvidos estudos vinculando as questões arquivísticas às questões de construção de identidade por meio da memória no plano conceitual ou em estudos relativos a identidades étnicas de determinados grupos sociais

e mesmo vinculando memória e arquivos a partir do uso de documentos arquivísticos em obras de arte (ARAÚJO, 2014, p. 87).

Organização do Conhecimento, para Bräscher (2008), é como o processo de modelagem do conhecimento que visa à construção de representações do conhecimento. Dessa maneira, “[...] a organização do conhecimento enquanto campo disciplinar, representa uma especialização para a Ciência da Informação” (BARROS; SOUSA, 2019, p. 79). Ainda, de acordo com os mesmos autores, a Organização do Conhecimento não só pode como tem relação com a Arquivologia e com os arquivos, principalmente quando se pensa nas possibilidades de abordagens referentes aos sistemas de organização, já que os sistemas de gestão, classificação, acesso e controle arquivísticos são justamente isto: sistemas conceituais baseados em características das instituições produtoras de documentos (BARROS; SOUSA, 2019).

Cabe à Arquivologia produzir instrumentos de recuperação da informação relacionados aos acervos arquivísticos custodiados pelas instituições de arquivo, tanto públicas quanto privadas. Com o advento das tecnologias de informação, a ênfase nos documentos de arquivo mudou, recaindo aos sistemas informatizados o gerenciamento de documentos arquivísticos e se aproximando cada vez mais a Organização do Conhecimento da Arquivologia.

Além disso, o usuário tem um importante papel nesse processo de estruturação, organização e recuperação do conhecimento, visto que é entendido como um dos personagens principais na busca da informação, manifestando seus interesses nas pesquisas e incitando os arquivistas a estarem cada vez mais conectados com as necessidades e maneiras de representar o conhecimento, a fim de dar acesso aos seus usuários, auxiliados pelas abordagens e teorias da Organização do Conhecimento. Na Arquivologia, as necessidades de acesso à informação estão vinculadas diretamente às dificuldades dos profissionais, no decorrer dos anos, de estruturar esquemas de classificação de documentos e fazer uma recuperação eficaz da informação.

Ao diferenciar informação de conhecimento, dizemos que informação é algo pontual, tem um prazo de vida curto e serve para o hoje, mas não nos garante nenhuma utilidade daqui a um tempo. Já o conhecimento tem um propósito e significado diferente de informação; sua utilização é atemporal, agrega valor e se transforma em benefício real, possuindo o poder de até mesmo mudar a forma de pensar das pessoas. Dessa maneira, conhecimento “[...] é um saber acumulado da humanidade, mas também é para satisfazer as necessidades sociais permanentes que requerem e geram novos conhecimentos” (BARITÉ, 2001, p. 42, tradução nossa).

A Organização do Conhecimento, na Ciência da Informação, procura adequar as diversas práticas e atividades sociais vinculadas ao acesso ao conhecimento. Hjørland (2008) define organização do conhecimento como algo

[...] sobre descrever, representar, arquivar e organizar documentos e representações de documentos, bem como assuntos e conceitos tanto por humanos quanto por programas de computador. Para esses fins, são desenvolvidos regras e padrões, incluindo sistemas de classificação, listas de títulos de assuntos, tesouros e outras formas de metadados (HJØRLAND, 2008, p. 86).

Já Dahlberg (2006) conceitua organização do conhecimento como

[...] a ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento inerentes (características) e a aplicação de conceitos e classes de conceitos ordenados dessa forma para atribuição de conteúdos de referentes (objetos/assuntos) de todos os tipos (DAHLBERG, 2006, p. 12, tradução nossa).

Esteban Navarro e Garcia Marco (1995), por sua vez, dizem que a Organização do Conhecimento se apresenta como uma plataforma de integração das ciências documentais, aproximando-se da Arquivologia, quando esta busca estudar esquemas de classificação para representação do conhecimento. Os instrumentos elaborados nos arquivos, os

catálogos, guias e inventários, possuem o mesmo objetivo que a própria organização do conhecimento, no caso, facilitar o acesso, a recuperação da informação e a gestão do conhecimento para os usuários.

## 2 TAXONOMIAS NA ARQUIVOLOGIA

O conhecimento é registrado sempre em documentos, independentemente do seu formato ou suporte; é um conjunto organizado de informações disponíveis, admitindo os mais diversos usos, de maneira indiscriminada (BARITÉ, 2001). Os documentos de arquivo são preservados, em um primeiro momento, por razões administrativas devido ao seu valor legal ou probatório, bem como por seu valor histórico. Dessa forma, os arquivistas, ao elaborar instrumentos de busca e estruturar sistemas de representação das informações dos documentos e suas temáticas, atuam como intermediários e organizadores do conhecimento. O armazenamento e a organização da informação estão diretamente relacionados ao objetivo de provar fatos, contar algo ou difundir o conhecimento para gerar novos conhecimentos. Assim, “[...] este conhecimento, se registrado, adquire um novo estatuto de informação que poderá ser socializado e potencialmente incorporado por algum indivíduo para gerar novo conhecimento” (SMIT, 2012, p. 95).

Guimarães (2008) diz que, no âmbito da Ciência da Informação, a organização e a representação do conhecimento apresentam natureza mediadora, configurando-se em um conjunto de processos que estabelecem a intermediação entre um conhecimento que, uma vez produzido, foi materializado e socializado, de tal forma que possa servir de base para a geração de um novo conhecimento. Por sua vez, esse novo conhecimento, uma vez materializado e socializado, pode igualmente ser objeto de nova organização e representação, caracterizando aquilo que se pode denominar como fluxo helicoidal da informação.

As classificações em arquivo não são uniformes, dado que estão automática e diretamente relacionadas à organização produtora desses documentos. Dessa maneira, os arquivistas utilizam legislações, estatutos e organogramas do órgão que os auxiliam nessa atividade.

De acordo com Barros e Souza (2019), a base para a construção de sistemas de organização do conhecimento consiste nas estruturas organizacionais, visto que estas oferecerem subsídios para a modelagem dos sistemas de organização do conhecimento. Igualmente, contribuem para a construção de melhores esquemas de classificação no desenvolvimento do fazer arquivístico, produzindo melhores resultados na representação, disseminação e acesso as informações por seus usuários finais. Os arquivistas, nesse contexto, atuam diretamente com a atividade de representação em função da interpretação que esse profissional faz do acervo arquivístico de uma instituição.

Faz-se necessário também entender conceito, o qual é fundamental para a construção, organização e representação do conhecimento. O conceito é uma “[...] unidade de conhecimento que surge pela síntese dos predicados necessários relacionados com determinado objeto e que, por meio de sinais linguísticos, podem ser comunicados” (DAHLBERG, 1978, p. 12, tradução nossa). Dahlberg também define o termo como “[...] a forma verbal de um conceito, o componente que convenientemente, sintetiza e representa um conceito com o propósito de designá-lo e comunicá-lo” (DAHLBERG, 1978, p. 76, tradução nossa).

Na Arquivologia, é a partir dos documentos de arquivo que se podem elaborar representações sobre o que são abordados nestes, como uma forma de garantir a recuperação da informação, pois é necessário tornar mais eficiente o armazenamento e o acesso às informações. A classificação de documentos baseada no princípio de proveniência e auxiliada pelo contexto de produção documental seria a base para elaboração de esquemas de representação do conhecimento. Assim,

[...] a representação do conhecimento é feita por meio de diferentes tipos de Sistemas de Organização do Conhecimento (SOC), que são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles (BRÄSCHER; CAFÉ, 2008, p. 8).

Nesse sentido, tais sistemas são instrumentos ou ferramentas que auxiliarão os usuários a encontrarem a informação de maneira mais rápida, eficaz e segura, pois foram elaborados levando em consideração o contexto de geração da informação registrada em documento de arquivo.

A grande maioria das instituições busca soluções que resolvam seus problemas de desorganização de arquivos. Com o enorme volume documental acumulado e a falta de acessibilidade aos documentos arquivísticos, recebem propostas de sistemas informatizados que prometem soluções milagrosas em curto espaço de tempo, o que, em grande parte, negligencia a estrutura que produziu, utilizou e acumulou o acervo. Este é um dos problemas que os Sistemas de Organização do Conhecimento (SOCs) podem auxiliar a sanar.

De acordo com Hodge (2000, apud BRÄSCHER e CAFÉ, 2008, p.8) onde este ressalta que os SOCs são “[...] o coração de toda biblioteca, museu e arquivo”, uma vez que são “mecanismos de organização da informação”. Na descrição de conteúdo, os SOCs cumprem a função de padronizar a representação da informação, no que concerne à identificação do assunto do documento. Relacionado à recuperação da informação, Vickery define os sistemas de organização do conhecimento como “[...] instrumentos complementares que ajudam o usuário a encontrar seu caminho no texto” Vickery (2008, apud BRÄSCHER e CAFÉ, 2008, p. 8).

Na Biologia, a taxonomia é um sistema de classificação desenvolvido por Lineu (naturalista sueco no século XVII) para agrupar e categorizar as espécies de seres vivos, em que se propôs uma hierarquia

de semelhanças entre eles, dividindo-os em cinco grupos: reino, classe, ordem, gênero e espécie. Aquino *et al.* (2009) acrescentam que as taxonomias surgiram no campo da Biologia e, por lá, são utilizadas há muito tempo, tendo se tornado alvo de estudos na Ciência da Informação, em que o seu aparecimento e uso estão relacionados com as formas automatizadas de criação da informação.

A taxonomia pode ser definida como “[...] um vocabulário controlado de uma determinada área do conhecimento e, acima de tudo, um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema sob uma premissa lógica” (TERRA *et al.*, 2005, p. 1). Assim, na Ciência da Informação, a taxonomia é um sistema de organização do conhecimento que classifica e facilita o acesso às informações, no qual se organizam as informações por categorias, normatizando as relações hierárquicas existentes entre elas. Nesse contexto, a taxonomia é uma solução que pode ser escolhida para atender à demanda de maneira sucinta ao classificar as informações contidas nos documentos de arquivo, auxiliando a representar e estabelecer os conceitos por meio de um sistema de organização do conhecimento.

A classificação hierárquica é muito utilizada em esquemas de classificação na Arquivologia. As taxonomias auxiliam os usuários a compreenderem como o conhecimento pode ser categorizado e, quando representam conceitos, oferecer um mapa que serve como guia nos processos de conhecimento. Sendo assim,

A taxonomia deverá ser usada como instrumento complementar à atividade de classificação de documentos, devendo ser pautada pelos seguintes passos: 1. Classificar o documento conforme o plano de classificação de documentos de arquivo; 2. Identificar no primeiro ou segundo nível de termos da taxonomia a área utilizada na classificação dos documentos; 3. Utilizar os termos dos diversos níveis da taxonomia como descritores (termos de indexação) em complemento à classificação dos documentos; e

4. Verificar a coerência da classificação combinada com a indexação nas dimensões funcionais e temáticas, respectivamente com os documentos (SOUSA; ARAÚJO JÚNIOR, 2017, p. 53).

A taxonomia é um dos sistemas de organização do conhecimento mais simples, menos estruturado e pode ser aplicado em qualquer universo e contexto organizacional, além de ser uma ferramenta de gestão que auxilia a realizar a organização em si, com um olhar voltado ao usuário e a quem realmente utiliza e precisa da informação. Também pode-se dizer que as taxonomias possuem um caráter dinâmico e apresentam necessidades de atualização, manutenção frequente e constante validação de conceitos que já foram preestabelecidos juntamente aos usuários, os quais são fundamentais, dado que auxiliam na classificação das informações e na modelagem dos sistemas de organização do conhecimento em conjunto aos arquivistas. Dessa maneira,

[...] as taxonomias são estruturas classificatórias para organizar as informações de uma determinada instituição num dado contexto. Nesse sentido, são diferentes, tendo em vista que refletem o tipo de organização e informação da instituição que representam (AQUINO *et al.*, 2009, p. 206).

Sendo assim, ao construir taxonomias, os arquivistas utilizarão princípios classificatórios para elaboração desses instrumentos, em que é necessário estabelecer a categorização das classes, “[...] a categorização é um processo que requer pensar o domínio de forma dedutiva, ou seja, determinar as classes de maior abrangência dentro da temática escolhida” (CAMPOS; GOMES, 2007, p. 5).

### 3 CONSIDERAÇÕES FINAIS

Estudos que explorem as taxonomias contribuirão para traçar melhores instrumentos de busca na Arquivologia, assim como auxiliar na representação do conhecimento e na construção de melhores planos

de classificação de documentos, com vistas à promoção da qualidade da pesquisa por informações. Não é uma atividade fácil, mas é necessário construir instrumentos que representem, cada vez melhor, a prática arquivística e as taxonomias desenvolvidas para domínios específicos. As taxonomias podem servir como guia e permitir agregação de novos conceitos, pois, estas têm como uma de suas características “[...] ser um instrumento de organização intelectual, atuando como um mapa conceitual dos tópicos explorados em um sistema de recuperação da informação” (CAMPOS; GOMES, 2007, p. 4).

Nesse sentido, “[...] as taxonomias representam os propósitos de organização intelectual de um dado contexto” (CAMPOS; GOMES, 2007, p. 2). Assim, é possível fazer a relação destas com a análise de domínio, pois é através da contextualização do domínio realizado na análise que se pode chegar à organização intelectual representada nas taxonomias.

No âmbito da Ciência da Informação, “[...] as taxonomias podem ser comparadas a estruturas classificatórias, como as Tabelas de Classificação, que têm como objetivo reunir documentos de forma lógica e classificada” (CAMPOS; GOMES, 2007, p. 2). Nesse sentido, têm ampla relação com a Arquivologia, pois também servem como instrumentos de organização e recuperação de informação. Outra questão em que as taxonomias podem auxiliar os arquivistas refere-se a como “ensinar” os usuários através das estruturas de conceito e hierarquia, facilitando a aplicação de práticas de gestão documental e entendimento por parte de pesquisadores e usuários.

Dessa maneira, é importante acrescentar que, para a Organização do Conhecimento, é necessário, cada vez mais, promover novas formas para representar e organizar conhecimento de grupos específicos que se encontram em desvantagem em estruturas classificatórias, principalmente em acervos arquivísticos. Por fim, a relação Arquivologia com a Organização do Conhecimento trará melhores respostas às

demandas institucionais e permitirá que os arquivistas possam auxiliar, ainda mais, na representação do conhecimento das mais diversas comunidades, grupos ou instituições.

## 4 REFERÊNCIAS

ALMEIDA, Daniela Pereira dos Reis *et al.* Paradigmas Contemporâneos da Ciência da Informação: a recuperação da informação como ponto focal. **Revista Eletrônica Informação e Cognição**, Marília, v. 6, n. 1, p.16-27, 2007. Disponível em: [https://brapci.inf.br/\\_repositorio/2010/03/pdf\\_fc-4f01292e\\_0008415.pdf](https://brapci.inf.br/_repositorio/2010/03/pdf_fc-4f01292e_0008415.pdf). Acesso em: 22 set. 2020.

AQUINO, Idalécio. J.; CARLAN, Eliana.; BRÄSCHER, Marisa. B. Princípios classificatórios para a construção de taxonomias. **Ponto de Acesso**, Salvador, v. 3, n. 3, p. 196-215, dez. 2009. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/3626/2744>. Acesso em: 29 mar. 2021.

ARAÚJO, Carlos. Alberto. Ávila. **Arquivologia, Biblioteconomia, Museologia e Ciência da Informação**: o diálogo possível. Brasília: Briquet de Lemos, 2014.

BARITÉ, Mario. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. *In*: CARRARA, Kester (org.). **Educación, universidad e pesquisa**. Marília: Unesp: FAPESP, 2001. p. 35-60.

BARROS, Thiago. Henrique. Bragato; SOUSA, Renato. Tarcisio. Barbosa de. Organização do conhecimento e Arquivologia: abordagens metodológicas. **Informação & informação**, Londrina, v. 24, n. 2, p. 76-92, maio/ago. 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/38290>. Acesso em: 18 set. 2020.

BORKO, Harold. Information Science: What is it? **American Documentation**, Washington, D.C., v.19, n.1, p. 3-5, Jan. 1968. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090190103>. Acesso em: 20 jan. 2020.

BRÄSCHER, Marisa. B. B.; CAFÉ, Lúcia. M. A. Organização da Informação ou Organização do Conhecimento? *In*: Encontro Nacional de Pesquisa em Ciência da Informação, 9., 2008, São Paulo. **Anais [...]**. São Paulo: ANCIB, 2008.

CAMPOS, Maria. Luiza. Almeida.; GOMES, Hagar. Espanha. Taxonomia e classificação: A categorização como princípio. *In*: Encontro Nacional de Pesquisa em Ciência da Informação, VIII, 2007, Salvador. **Anais**. Bahia. Brasil, 2007.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115/115>. Acesso em: 18 fev. 2020.

DAHLBERG, Ingetraut. Knowledge organization: a new science? **Knowledge Organization**, Baden, v. 33, n. 1, p. 11-19, 2006. Disponível em: [https://www.ergonverlag.de/isko\\_ko/downloads/ko3320061c.pdf](https://www.ergonverlag.de/isko_ko/downloads/ko3320061c.pdf). Acesso em: 20 jan. 2020.

ESTEBAN NAVARRO, Miguel Angel; GARCÍA MARCO, Francisco Javier. Las primeras jornadas sobre organización del conocimiento: organización del conocimiento e información científica. Scire: representación y organización del conocimiento, Zaragoza, v. 1, n. 1, p. 149-157, ene./jun. 1995. Disponível em: <http://www.ibersid.eu/ojs/index.php/scire/article/download/1038/1020/0>. Acesso em: 20 jan. 2020

GUIMARÃES, José. Augusto. Chaves. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 1, n. 1, 2008. Disponível em: <https://doi.org/10.26512/rici.v1.n1.2008.940>. Acesso em: 20 jan. 2020.

HJØRLAND, Birger. Core classification theory: a reply to Szostak. **Journal of Documentation**, United Kingdom, v. 64, n. 3, p. 333-342, 2008. Disponível em: <https://doi.org/10.1108/00220410810867560>. Acesso em: 30 set. 2020.

HJØRLAND, Birger. Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, United Kingdom, v. 58, n. 4, p. 422-462, ago. 2002. Disponível em: <https://doi.org/10.1108/00220410210431136>. Acesso em: 21 out. 2020.

HJØRLAND, Birger. What is Knowledge Organization (KO)? **Knowledge Organization**, Baden, v. 35, n. 2/3, p. 86-101, 2008. Disponível em: <https://repository.arizona.edu/handle/10150/106183>. Acesso em: 20 jan. 2020.

SMIT, Johanna. Wilhelmina. Smit. A informação na Ciência da Informação. **INCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 3, n. 2, p. 84-101, 2012. DOI: 10.11606/issn.2178-2075.v3i2p84-101. Disponível em: <https://www.revistas.usp.br/incid/article/view/48655>. Acesso em: 18 maio. 2021.

SOUSA, Renato. Tarcísio. Barbosa de.; ARAÚJO JÚNIOR, Rogério. Henrique de. A indexação e criação de taxonomias para documentos de arquivo: proposta para a expansão do acesso e integração das fontes de informação. **Brazilian Journal of Information Science: Research Trends**, Marília, v. 11, n. 4, p. 47-56, 2017. Disponível em: <http://revistas.marilia.unesp.br/index.php/bjis/article/view/7508/4788>. Acesso em: 03 jan. 2020.

TERRA, José Cláudio C. *et al.* Taxonomia: elemento fundamental para a gestão do conhecimento. **Slide Share**, S.l., 21 ago. 2005. Disponível em: <https://pt.slideshare.net/jcterra/taxonomia-elemento-fundamental-para-a-gestao-do-conhecimento>. Acesso em: 04 set. 2021.

VITORIANO, Marcia Cristina de Carvalho Pazin. Uma aproximação entre Arquivologia e Ciência da Informação: o uso dos conceitos de informação orgânica e informação arquivística. *Brazilian Journal of Information Science: Research Trends*, Marília, v. 11, n. 4, p. 57-66, 2017. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/issue/view/447>. Acesso em: 12 fev. 2020.



# 10

Talita Morgana Arruda Tavares

Fabio Assis Pinho

**Reflexão sobre folksonomia  
e engajamento político**

*Reflection on folksonomy  
and political engagement*

DOI: [10.31560/pimentacultural/2022.95613.10](https://doi.org/10.31560/pimentacultural/2022.95613.10)

### Resumo:

Este capítulo sobre folksonomia e engajamento político materializa uma pesquisa cujo objetivo é o de analisar o engajamento dos usuários do *Twitter* em temas de viés político com o uso de duas *hashtags* – #CPIdaCovid e #29M – sob a ótica da folksonomia. A pesquisa se justifica pelo fato de que o uso dessas marcações pontua acontecimentos históricos a partir da visão pessoal dos usuários da mídia e é um extrato social que utiliza a linguagem natural para marcar acontecimentos. Trata-se de uma pesquisa qualitativa, exploratória e de observação não participante, utilizando a Análise de Conteúdo (AC), com as categorias temáticas notícia, opinião, repercussão, meme ou imagem viral e outros, e as categorias a favor, neutra e contra. O estudo conclui que a folksonomia pode ser compreendida e considerada como instrumento para o engajamento político de usuários no meio digital, em especial no *Twitter*, pois é possível observar que a folksonomia pode ser considerada uma ferramenta eficaz de engajamento, já que é em torno das publicações, utilizando de marcações em forma de indexação livre das informações, que os usuários se encontram, debatem e formam opinião ou se aproximam em torno de lutas e manifestações com as quais convergem ou divergem.

**Palavras-Chave:** Folksonomia; política; *Twitter*; engajamento.

**Abstract:**

*This chapter on folksonomy and political engagement materializes research whose objective was to analyze the engagement of Twitter users in themes of political bias using two hashtags – #CPIdaCovid and #29M – from the perspective of folksonomy. The research is justified by the fact that the use of these markings punctuates historical events from the personal view of media users and is a social extract that uses natural language to mark events. It was a qualitative, exploratory, and non-participant observation research, using Content Analysis (CA), with the thematic categories news, opinion, repercussion, meme or viral image and others and the categories in favor, neutral, and against. The study concluded that folksonomy can be understood and considered as an instrument for the political engagement of users in the digital environment, especially on Twitter, as it was observed that folksonomy can be considered an effective engagement tool, as it revolves around publications using tags in the form of free indexing of information that users meet, debate and form opinions or comes together around struggles and manifestations with which they converge or diverge.*

**Keywords:** Folksonomy; policy; Twitter; engagement.

## 1 INTRODUÇÃO

Estar conectado às redes sociais, hoje, é algo tão natural e difundido que não há mais dúvidas em relação à importância de se pesquisar os impactos do digital nas áreas de conhecimento. Na Ciência da Informação (CI) – que investiga a informação e seus usos – é ainda mais inegável a relevância de pesquisas que abarquem esse campo, pois, sendo ela uma área compreendida pela interdisciplinaridade, que antes tratava sobre o livro e a biblioteca, expandiu sua atuação para teorias de classificação e indexação, além da recuperação de dados e estudos descritivos (LE COADIC, 1994, p. 23).

Com a inserção do digital na vida cotidiana e a difusão do uso das mídias sociais, houve grandes transformações, também, na comunicação interpessoal, que passou a ser mediada por aparatos eletrônicos. De acordo com Castells (2005, p. 24), o aparato tecnológico nas comunicações

[...] criou um novo sistema de redes de comunicação global e horizontal que, pela primeira vez na história, permite que as pessoas se comuniquem umas com as outras sem utilizar os canais criados pelas instituições da sociedade para a comunicação socializante (CASTELLS, 2005, p. 24).

Essa forma de comunicação viabiliza, então, que as pessoas debatam e se organizem em torno das mais diversas pautas, sem intermediários, aumentando o engajamento em temas antes restritos ao *off-line* ou a públicos específicos. Com as manifestações políticas, o cenário não é diferente. É nas redes sociais, locais marcados pela efervescência de debates, que há o encontro de discursos similares e opostos, em que, quem tem acesso à internet, expõe suas opiniões.

É nessa perspectiva, com a proposição de compreender a participação dos usuários das redes em pautas políticas, que analisamos o uso de duas *hashtags* no *Twitter* sob a ótica da folksonomia, a qual é compreendida segundo Assis e Moura (2013, p. 87), como “modalidade

de organização da informação em ambientes digitais”. Dessa forma, a pergunta de pesquisa que norteou nossa análise foi: como se dá o engajamento dos usuários do *Twitter* em temas de viés político?

Para responder a essa pergunta, o objetivo geral dessa pesquisa é o de analisar o engajamento dos usuários do *Twitter* em temas de viés político com o uso de duas *hashtags* – #CPIdaCovid e #29M – sob a ótica da folksonomia. A justificativa que aqui se postula é pelo fato de que o uso dessas marcações serve como importante objeto de estudo, já que pontua acontecimentos históricos, a partir da visão pessoal de cada usuário da rede, e é um extrato social que utiliza a linguagem natural para marcar acontecimentos.

A reflexão que trazemos nesse capítulo tem abordagem qualitativa, ancorada na Análise de Conteúdo (AC) de assuntos políticos materializados por *hashtags* no *Twitter*. Para isso, #CPIdaCovid e #29M foram as *tags* determinadas. A #CPIdaCovid se refere aos trabalhos do Senado Federal em compreender as responsabilidades do governo federal na condução da pandemia e a #29M faz referência aos protestos que aconteceram em 29 de maio de 2021 pelo Brasil contra o presidente da República Federativa do Brasil, pela continuidade do auxílio emergencial durante a pandemia do coronavírus, pela ampliação da vacinação entre outras pautas.

Consequentemente, foi definido o período em que as *hashtags* seriam analisadas. Para a #CPIdaCovid, foi escolhido o dia 13 de maio de 2021, quando o CEO da Pfizer, Sr. Carlos Murillo, prestou depoimento no Senado Federal e, de acordo com matéria de Gois (2021, p. 1) ,no portal O Globo, foi “o dia com mais manifestações, a exceção de ontem (19 de maio), foi 13 de maio, data do testemunho do representante da Pfizer, a farmacêutica”. De acordo com a consultoria de dados Bites<sup>68</sup>, foram realizados 46.275 usos da *hashtag*.

68 A Bites (<https://bites.com.br/>) é uma consultoria que captura dados abertos e retorna para o cliente em forma de relatório. A pesquisa dessa hashtag foi enviada de forma gratuita para a realização dessa reflexão.

Foi a partir desse cenário que delimitamos o *corpus* de análise. Foram utilizados *tweets* que obtiveram pelo menos um *retweet*, descartando aqueles que eram respostas a publicações de outros usuários, visando alcançar aqueles que obtiveram uma interação mais relevante na rede social. Essa busca foi realizada na própria plataforma do *Twitter*, no campo “busca avançada” e “mais recentes” e retornou 1.479 publicações.

Para a *hashtag* do ato nas ruas, a #29M, foi analisado o dia 29 de maio de 2021, dia da manifestação. Da mesma forma que a #CPI-daCovid, foi utilizada a busca avançada da plataforma considerando, também, publicações com pelo menos um *retweet*, gerando 1.476 resultados. É necessário pontuar que as publicações aqui utilizadas foram possíveis de serem recuperadas pois os perfis dos usuários estão de modo público no *Twitter*, permitindo ser encontrados nas buscas e *retweetados* por outros usuários da rede (com identidades preservadas).

Para a análise de conteúdo, as categorias definidas para ambas as *hashtags* foram as seguintes: notícia, opinião, repercussão, meme ou imagem viral e outros (sem enquadramento nas categorias anteriores). Além disso, outras três categorias foram utilizadas na pesquisa: publicações marcadas como “A favor”, “Neutra” e “Contra”. No caso da #CPI-daCovid, as marcadas como “A favor” se referem as mensagens em apoio à CPI e em oposição ao governo federal, “Neutra” são aquelas que não falam diretamente sobre a CPI ou apenas repercutem algo sem juízo de valor especificamente sobre a *hashtag* ou a CPI e “Contra” são os *tweets* contrários a realização da CPI e a favor do governo federal.

No caso da #29M, “A favor” são as mensagens em apoio às manifestações e contra o governo federal, “Neutra” são as que não falam diretamente sobre as manifestações ou a publicação não possui juízo de valor sobre elas e “Contra” são as publicações que utilizaram a *tag*, mas são contrárias ao movimento.

## 2 ENGAJAMENTO POLÍTICO, FOLKSONOMIA E MÍDIAS SOCIAIS

Maar (1982, p. 399) diz que a política anda lado a lado com a história, “com o dinamismo de uma realidade em constante transformação que continuamente se revela insuficiente e insatisfatória e que não é fruto do acaso, mas resulta da atividade dos próprios homens vivendo em sociedade” (MAAR, 1982, p. 36). Dessa maneira, seria necessária a “organização e mobilização em torno de interesses sociais, que desta forma passam a se desenvolver com objetivos políticos [...] para transformar em direito as reivindicações” (MAAR, 1982, p. 456).

Compreende-se, então, que o agir cidadão conversa com a necessidade do engajamento político por parte da população. Rosenfield (2017, p. 7) reflete sobre essa questão pontuando que “o conceito de ‘política’ se refere efetivamente ao que é coletivo, ao que é comum a todos”. Para o autor, “o indivíduo é antes de tudo cidadão, ou seja, membro de uma cidade”, sendo assim, “a comunidade política que é legada às próximas gerações provém da responsabilidade que cada um assume, hoje, em relação aos assuntos coletivos” (ROSENFELD, 2017, p. 6).

Em locais que possuem a Democracia como regime político – sendo, em resumo, o exercício do poder político por parte do povo, com a participação do povo na tomada de decisões –, a perspectiva de o indivíduo buscar agir em comunidade se torna mais concreta e necessária. A democracia, para Rosenfield (2017, p. 25),

[...] inaugura um novo sentido do político, ao determinar-se por um espaço público de discussão, de luta, de negociação e de diálogo. A reunião de todos aqueles que constituem a sociedade numa forma de organização política aberta ao seu aperfeiçoamento dá aos cidadãos um novo sentido da comunidade, não excluindo ninguém, por princípio, dos assuntos públicos (ROSENFELD, 2017, p. 25).

Nesse contexto, nossa reflexão traz o engajamento político sob a ótica da folksonomia como modalidade de organização do conhecimento nos ambientes digitais. Ela representa o uso da linguagem natural promovida pelos próprios usuários das redes de forma independente ou colaborativa. O termo é uma junção de povo (do inglês *folk*) e taxonomia (uma forma de classificação) e foi cunhado por Wal (2007, não paginado, tradução nossa):

Folksonomia é o resultado da marcação livre e pessoal de informações e objetos (qualquer coisa com um URL) para a própria recuperação. A marcação é feita em um ambiente social (geralmente compartilhado e aberto a outras pessoas). A folksonomia é criada a partir do ato de marcação pela pessoa que consome as informações (WAL, 2007, não paginado, tradução nossa).

Nesse processo, o próprio usuário de determinada rede “rotula, nomeia, designa suas palavras, dados e arquivos de forma livre, não existindo nenhuma metodologia e/ou etapas a serem realizadas para a organização e a recuperação das informações de seu interesse” (MOREIRA; ROMÃO, 2010, p. 4). Como observa O’Reilly (2005, p. 3, tradução nossa), a folksonomia seria uma categorização colaborativa “usando palavras-chave livremente escolhidas, geralmente chamadas de *tags*. A marcação permite o tipo de associações múltiplas e sobrepostas que o próprio cérebro usa, em vez de categorias rígidas” (O’REILLY, 2005, p. 3, tradução nossa).

Para Assis e Moura (2013, p. 87), a folksonomia é “uma inovação que explora o potencial das redes sociais na organização e no compartilhamento dos recursos informacionais”, dessa forma, esse tipo de organização da informação agregaria as “manifestações da linguagem contextualizada e, por vezes, caótica de sujeitos em colaboração”. Para sintetizar o conceito de folksonomia, o presente capítulo se utiliza da compreensão produzida por Corrêa e Santos (2018, p. 29), como sendo:

[...] o resultado do processo de etiquetagem livre (atribuição de etiquetas, palavras-chave) realizada pelos usuários mediante o

emprego de termos provenientes de linguagem natural – dispensando o uso de vocabulários controlados – em ambientes digitais colaborativos visando indexar recursos informacionais compartilhados de qualquer formato (textos, imagens, áudio, vídeo etc.) para fins de sua representação e recuperação (CORRÊA; SANTOS, 2018, p. 29).

A partir disso, entende-se, então, que cada usuário, a partir da visão de mundo que possui, organiza – ou categoriza, classifica – a informação de modo natural, próprio e subjetivo, de forma democrática. O que acontece a partir disso é a possibilidade de conexão de diversos indivíduos que começam a interagir a partir de qualquer local do globo em torno de um determinado tema ou *hashtag*.

Ao termo *hashtag*, somente é permitido seu surgimento após o aparecimento da Web 2.0, que seria considerada a segunda geração da internet. Tim O’Reilly (2005, p. 3) definiu a Web 2.0 como sendo a “web enquanto plataforma”. De acordo com Dias, Castro e Silva (2015, p. 6), esse espaço se configura como um “ambiente que diz respeito ao aspecto evolutivo e gradativo da fase preliminar na perspectiva da participação coletiva, na produção de informação, em redes de relacionamento”. As tecnologias que estão presentes nessa fase da Web são intuitivas e interativas e conferem maior independência aos usuários, interesse pela navegação e descoberta de conteúdos, favorecendo o surgimento de novas formas de produção e uso de informação (DIAS; CASTRO; SILVA, 2015). É a partir do estabelecimento da Web 2.0 que é possível o surgimento das redes sociais como são conhecidas hoje e, conseqüentemente, das *hashtags*. Para criar uma, é necessário colocar o símbolo “#” antes da palavra ou expressão que se queira destacar, formando uma *tag* (ou etiqueta).

A rede social aqui utilizada para a reflexão neste capítulo foi o Twitter, Inc., que pode acessada por meio do endereço: <http://www.twitter.com>. A mídia, que surgiu no ano de 2006, é considerada um serviço de *microblogging*, no qual os usuários escrevem pequenos

textos que podem conter até 280 caracteres. O site sugere ao usuário começar a escrever a partir da pergunta “O que está acontecendo?”. Cada pessoa pode criar um perfil na rede, seguir outros perfis de usuários e ser seguido por eles. Pode-se publicar de maneira privada ou pública e, caso queira direcionar a mensagem a uma conta em específico, deve-se acrescentar o símbolo “@” antes do nome do usuário destinatário da publicação.

Nessa ambiência, o engajamento político por parte dos cidadãos, o qual é permitido nas democracias, nas sociedades contemporâneas, tem se expressado não só nas ruas, por meio dos protestos, mas também nos meios digitais, na internet. As inovações tecnológicas abrem novas possibilidades para o diálogo entre os atores sociais para que eles se integrem na prática democrática e se tornam uma nova via para promover o ativismo social.

Ribeiro, Borba e Hansen (2016, p. 503-504) refletem que, em uma perspectiva otimista, a internet cria uma nova estrutura de comunicação e, com isso, surgem oportunidades diferentes de participação política além da convencional que ocorre no off-line, tendo o poder de dirimir barreiras da participação popular, ampliando o debate político. “Nesse sentido, a internet é vista como parte do espaço público e, desse modo, a ágora poderia ser reproduzida virtualmente, trazendo a possibilidade de participação ao cidadão comum, inserindo-o nos processos de deliberação.”

Importante ressaltar, então, os números de acesso à internet no Brasil e no mundo para se ter uma noção do alcance das tecnologias no atual panorama social. De acordo com pesquisa realizada em janeiro de 2021 pela agência *We are Social* e pela Hootsuite, divulgada pela Istoé Dinheiro (NÚMERO..., 2021), existem cerca 4,66 bilhões de pessoas conectadas em todo o mundo, o que significa mais da metade da população mundial. No Brasil, esse número, em 2020, de acordo com a pesquisa TIC Domicílios 2019, é de 134 milhões de usuários, ou seja,

três em cada quatro brasileiros acessam a internet no país, segundo divulgado pela Agência Brasil (BRASIL..., 2020).

Essa realidade digital, tão presente no nosso cotidiano, foi definida por Castells (2005, p. 24) como sendo uma “sociedade em rede”, que é:

[...] uma estrutura social baseada em redes operadas por tecnologias de comunicação e informação fundamentadas na microelectrónica e em redes digitais de computadores que geram, processam e distribuem informação a partir de conhecimento acumulado nos nós dessas redes (CASTELLS, 2005, p. 24).

O autor resgata que movimentos sociais começaram a se organizar na internet já na década de 1980 e basicamente os mesmos componentes faziam parte desses grupos: movimentos locais em busca de oportunidades de auto-organização, o movimento *hacker* e governos municipais com o intuito de fortalecer a participação do cidadão. Como os governos não conseguiam controlar os fluxos de comunicação ali existentes, a internet conseguiu fugir do controle estatal.

Dessa forma, Castells (2005, p. 24) observa que a nova configuração social começa a tornar a comunicação de massa horizontalizada e autocomandada, ou seja, possui uma interação que não necessita dos meios dos sistemas de mídia tradicionais para acontecer, porque ela parte da iniciativa de indivíduos ou grupos independentes. Para a política, as características dessa sociedade em rede são relevantes,

[...] uma vez que a política é largamente dependente do espaço público da comunicação em sociedade, o processo político é transformado em função das condições da cultura da virtualidade real. As opiniões políticas e o comportamento político são formados no espaço da comunicação (CASTELLS, 2005, p. 24).

O autor compreende que, por a internet estar se tornando importante meio de comunicação e de sistematização, com certeza os movimentos sociais e políticos já a utilizam como um “instrumento

privilegiado para atuar, informar, recrutar, organizar, dominar e contra-dominar” (CASTELLS, 2003). Ele reforça, porém, que esses espaços não são determinantes na formação da opinião pública, já que os indivíduos possuem autonomia na busca por fontes de informação – bem como para se tornarem as próprias fontes. Para Recuero (2009, p. 16), que estuda a interação mediada pelo computador, o advento dessa comunicação “está mudando profundamente as formas de organização, identidade, conversação e mobilização social”.

Essa mobilização, que perpassa pelas pautas políticas, surge, comumente, em um ambiente em que demandas da sociedade não estão sendo supridas. Para Castells (2013, p. 127),

[...] a conjunção de degradação das condições materiais de vida e crise de legitimidade dos governantes encarregados de conduzir os assuntos públicos leva as pessoas a tomar as coisas em suas próprias mãos, envolvendo-se na ação coletiva fora dos canais institucionais prescritos para defender suas demandas e, no final, mudar os governantes e até as regras que moldam suas vidas (CASTELLS, 2013, p. 127).

É possível afirmar que a realidade social se alterou quando atingiu o âmbito digital e com o surgimento e estabelecimento das redes sociais, locais de fluxo intenso de publicação e troca de informação, nas quais “as comunidades de interesse organizam-se com riqueza muito maior de informações específicas” (MAGRANI, 2014, p. 84). Magrani (2014, p. 87) diz ainda que os movimentos que acontecem no on-line permitem “catalisar agendas através de engajamento, diálogo, circulação e protesto” e que:

[...] indivíduos antes marginais no diálogo político possuem, com as novas mídias, meios para se mobilizar, debater opiniões entre si e com instituições governamentais e mídia tradicional e inclusive influenciar estes atores através de diferentes plataformas (MAGRANI, 2014, p. 87).

Os indivíduos passam, então, a encontrar, nas mídias sociais, na Internet, o local para expressar ideias e opiniões, reunir, debater e compreender discursos dos mais diversos. Para Recuero (2009, p. 24), as redes sociais se desenham como propagadoras de discursos, ao mesmo tempo, individuais e coletivos e afirma que o que chama de “Comunicação Mediada pelo Computador” redefiniu formas de organização e mobilização social. Essas redes, como um espaço de livre expressão, não possuem barreiras governamentais ou partidárias, o que dá visibilidade e engaja pessoas nos mais diversos acontecimentos.

Também se faz necessário pontuar, ainda mais se tratando do debate político nas redes, a visão pessimista do uso da internet e das redes sociais nesse âmbito. Magrani (2014, p. 106-162) cita alguns pontos que dão vazão a esse pensamento: a exclusão digital, a sobrecarga de informação, o filtro-bolha, a polarização e fragmentação, falta de cultura de engajamento político on-line e sistema político não dialógico, tecnização do debate e a tecnologia para o controle. A exclusão se refere a falta de acesso universal à internet, retirando alguns cidadãos do debate amplo. A sobrecarga, ou *information overload*, diz respeito à quantidade excessiva de informações disponíveis, que atrapalham o indivíduo na compreensão e tomada de decisões.

Já o filtro-bolha (*Filter Bubble*) faz referência aos algoritmos que se formam a partir dos dados de navegação, limitando a liberdade de navegação dos usuários e os deixando à margem de alguns discursos e opiniões. Este conversa com outro ponto negativo, a “polarização e fragmentação”, que é quando as pessoas buscam e compartilham conteúdo de valores e opiniões e ideologias semelhantes. Com relação a falta de cultura de engajamento político on-line e sistema político não dialógico, Magrani (2014, p. 138) observa que há ações individuais descentralizadas e as discussões nas redes não atingem: “alto nível racional-dialógico relevantes para a política ou para o incremento da

democracia”. Outro ponto negativo apontado pelo autor é a ênfase à parte técnica do debate faz com que o usuário da internet se interesse menos em se engajar em discussões, o que diminuiria o valor do que está sendo discutido e restringindo a discussão dos temas a um menor número de pessoas, em teoria mais qualificados.

Por fim, Magrani (2014, p. 152) cita a tecnologia para o controle, em que pontua a preocupação com a regulação da internet por parte dos governos para garantir os direitos dos cidadãos no ambiente on-line, como privacidade, liberdade de expressão e direitos consumeristas. “Há hoje algum consenso de que a regulação (em alguma medida) da internet é importante para que haja maior segurança de garantias aos direitos constitucionais dos cidadãos-usuários” (MAGRANI, 2014, p. 152).

### 3 HASHTAGS NO ENGAJAMENTO POLÍTICO

As duas *hashtags* para a realização desta reflexão – #CPIda-Covid e #29M – tratam de formas diferentes de engajamento político. A primeira é sobre um acontecimento que será discutido e tratado por meses, no caso da CPI, que tinha duração prevista inicialmente de 90 dias, sendo renovada, posteriormente, por mais 90 dias; e a outra é sobre um momento específico em que o evento é realizado, em que o chamamento para a realização dele foi feito basicamente on-line.

Vejamos o uso das marcações aqui escolhidas com o momento político vivenciado no Brasil, a fim de compreender o surgimento das *tags*, como pode ser visualizado a seguir no quadro 1.

Quadro 1 – *Hashtags*, contexto em que são utilizadas e o período de análise

Hashtag	Contexto	Período de coleta
<b>#CPIdaCovid</b>	Essa <i>hashtag</i> foi utilizada quando os usuários do <i>Twitter</i> queriam fazer referência aos acontecimentos da Comissão Parlamentar de Inquérito (CPI) do Senado Federal brasileiro que trata da pandemia da Covid-19 no país. A CPI da Pandemia foi instalada em 27 de abril e visa investigar ações e omissões do governo federal no combate à pandemia.	13 de maio de 2021.
<b>#29M</b>	A <i>#29M</i> surgiu para os usuários repercutirem os protestos que aconteceram em 29 de maio de 2021 pelo país contra o governo de Jair Bolsonaro. Em meio à pandemia da Covid-19, as reivindicações se concentraram em torno do <i>impeachment</i> do presidente, a continuidade do auxílio emergencial, a ampliação da vacinação, contra os cortes de verba na educação, entre outras pautas.	29 de maio de 2021.

Fonte: Os autores, 2021.

A partir da busca avançada da própria plataforma, então, foram recuperados 1.479 *tweets* utilizando a *#CPIdaCovid* e 1.476 publicações da *#29M*. Com base na Análise de Conteúdo e na definição das categorias mencionadas anteriormente, estas ficaram classificadas conforme demonstra o quadro 2 a seguir.

Quadro 2 – Distribuição e classificação de publicações em cada categoria de análise.

Categoria	#CPIdaCovid	#29M
<b>Notícia</b>	55	42
<b>Opinião</b>	879	874
<b>Repercussão</b>	386	436
<b>Meme</b>	59	22
<b>Outros</b>	100	102
<b>Total</b>	<b>1.479</b>	<b>1.476</b>

Fonte: Os autores, 2021.

Também, de acordo com a categorização realizada a partir da Análise de Conteúdo, a pesquisa definiu as publicações como “A favor”, “Neutra” e “Contra”. Para essa definição, foram considerados diversos fatores, entre eles, o uso de outras *hashtags* por parte dos usuários. O quadro 3, a seguir, mostra a quantidade de publicações nessas categorias.

**Quadro 3 – Quantidade de publicações a favor, “neutras” e contra os eventos**

Hashtag	A favor	“Neutra”	Contra
#CPIdaCovid	746	706	27
#29M	1265	210	1

Fonte: Os autores, 2021.

Dessa forma, a partir de agora, demonstra-se, com números e exemplos, como se comportaram essas publicações ao cruzar as informações. Com relação a #CPIdaCovid, obteve-se 55 notícias, todas marcadas como “neutras”. Um exemplo pode ser conferido na figura 1.

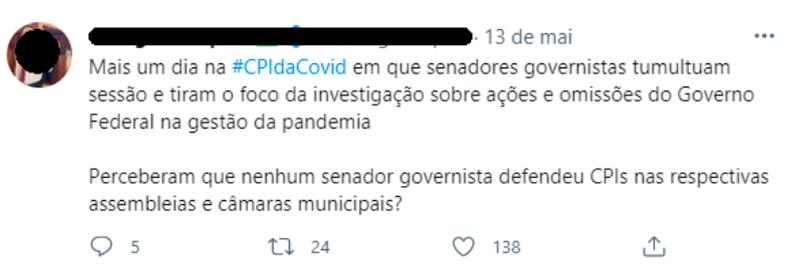
**Figura 1 – #CPIdaCovid: Exemplo das categorias notícia e “neutra”**



Fonte: Twitter, Inc. (2021).

As publicações marcadas como opinião e a favor foram a grande maioria, com 659 publicações. Elas apresentam algum juízo de valor favorável à CPI, como demonstrado na figura 2.

Figura 2 – #CPIdaCovid: Exemplo das categorias opinião e a favor



Fonte: Twitter, Inc. (2021).

Apesar de, na denominação, as categorias opinião e “neutra” possam parecer contraditórias, 196 publicações foram marcadas dessa forma, porque elas apresentam alguma opinião, mas elas não se encaixam como “A favor” ou “Contra” a CPI, poderiam ser feitas por quem apoia ou não os trabalhos da comissão ou o governo federal. Outros 24 tweets de opinião foram marcados como contrários à CPI.

Já com relação a #29M, foram identificadas 21 publicações com notícias marcadas como “a favor”. Elas são, basicamente, de portais independentes de notícias. Também foram encontradas 20 notícias marcadas como neutras, conforme exemplo na figura 6 a seguir. Apenas uma notícia foi marcada como contrária por apresentar, já no título, uma crítica a realização do evento.

As publicações marcadas como de opinião e a favor foram a maioria dos dados coletados: 852. Como “neutra” foram encontradas 22 publicações. Como categorias de repercussão e a favor, foram 301 publicações, e repercussão e neutra foram 135, como demonstrado nas figuras 3 e 4 a seguir.

Figura 3 – #29M: Exemplo das categorias repercussão e a favor



Fonte: Twitter, Inc. (2021).

Figura 4 – #29M: Exemplo das categorias repercussão e “neutra”



Fonte: Twitter, Inc. (2021).

Nas categorias meme/imagem viral e a favor”, foram encontrados 22 tweets, conforme exemplo na figura 5.

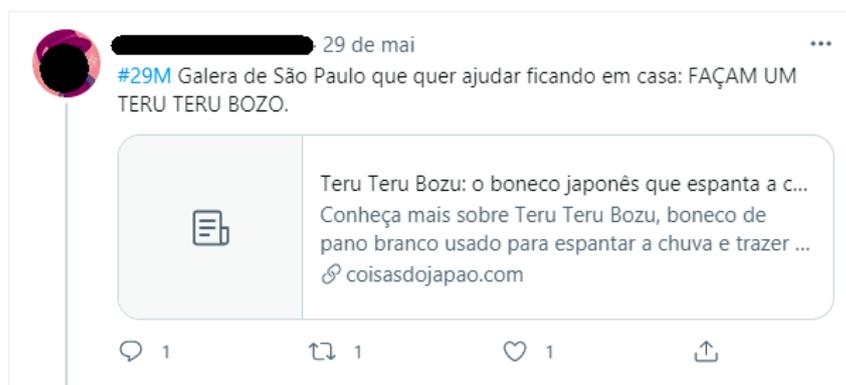
Figura 5 – #29M: Exemplo das categorias meme/imagem viral e a favor



Fonte: Twitter, Inc. (2021).

E, finalmente, como “outros”, aparecem 74 a favor e 28 neutras, como visto nas figuras 6 e 7, a seguir:

Figura 6 – #29M: Exemplo das categorias outros e a favor



Fonte: Twitter, Inc. (2021).

Figura 7 – #29M: Exemplo das categorias outros e “neutro”



Fonte: Twitter, Inc. (2021).

O debate que acontece nas redes sociais em torno de ambas as *hashtags* e dos temas que as compreendem tem repercutido não só nas redes sociais, mas na mídia tradicional. O alcance da CPI da Pandemia foi (e é) tão grande no Twitter que foi comparado com o do programa de TV Big Brother Brasil pela Folha de São Paulo. Com o título “CPI da Covid atrai público que ficou órfão do BBB e decola na era da ‘política pop’”, Moraes e Menon (2021) escrevem que: “não faltam aproximações dramáticas entre o que se tem visto no Senado no último mês e as grandes atrações do entretenimento brasileiro”.

Entrevistado pelas jornalistas, o cientista de dados Pedro Barciela afirmou que eventos como o *impeachment* de Dilma Rousseff e as eleições de 2008 tiveram um maior número de interações, porém “no universo do parlamento, ela (a CPI) pode ser, sim, um dos eventos que mais criaram engajamento” (MORAES; MENON, 2021). Também sobre a CPI, o Correio Brasiliense divulgou que “ao longo do primeiro e agitado mês de depoimentos, a CPI da Covid resultou em 11,8 milhões de publicações no Twitter” (incluindo ou não o uso de *hashtags*).

Já as manifestações sobre o 29 de maio de 2021 repercutiram no jornal O Globo. A reportagem de João Saconi informou que foram 1,8 milhões de menções do dia 1º de maio até o dia 29 no Twitter com *hashtags* a favor e contra a manifestações. “Online, o ato foi mencionado 767 mil vezes por meio da *hashtag* #29MForaBolsonaro e 29 mil vezes com a #ForaGenocida, de acordo com a contagem da Arquimedes até as 19h” (SACONI, 2021). Ele também escreve que, em contraponto, a base do governo investiu na *hashtag* #EuApoioBolsonaro2022, “que teve 165 mil menções neste sábado, em referência à busca pela reeleição do político”.

## 4 CONSIDERAÇÕES FINAIS

Os números expressivos do uso das *hashtags* e de publicações em torno dos temas políticos aqui expostos refletem a importância que as redes sociais, em especial o *Twitter*, conseguiram alcançar ao redor dessas questões. Seja da simples manifestação de pensamento, da repercussão de um fato ou da convocação para manifestações de rua, é possível compreender que o uso que se faz das mídias sociais vem impactando até na mais básica compreensão do ser cidadão, de atuar enquanto cidadão.

A partir do exposto, foi possível, então, observar que, sim, a folksonomia pode ser considerada uma ferramenta eficaz de engajamento, já que é em torno das publicações, utilizando essas marcações em forma de indexação livre das informações, que os usuários se encontram, debatem e formam opinião ou se aproximam em torno de lutas e manifestações com as quais convergem ou divergem. Por meio da folksonomia, então, pode-se observar e analisar diversos fenômenos que ocorrem no “mundo real” ou *off-line*, mas que se moldam, fortalecem-se e ganham relevância no digital, além de se estabelecer como

método atual e necessário, para estudos acerca dos debates políticos. Dessa forma, foi possível atingir o objetivo geral do trabalho de analisar o engajamento dos usuários do *Twitter* em temas de viés político, a partir do uso de *hashtags*, sob a ótica da folksonomia.

É imperioso que, dado o advento do uso das redes sociais, as discussões levantadas nessas mídias não sejam inferiorizadas ou alcancem um lugar que fique à margem dos debates políticos. Pelo que foi visto, a partir dos exemplos de eventos, com destaque no meio digital, e pelos exemplos aqui expostos, o mundo já está em transformação e essas mudanças perpassam pelos indivíduos que se organizam e expõem suas opiniões nessas redes.

Importante também destacar que essa reflexão por si só não se propõe a esgotar todas as interpretações das *hashtags* com viés político ou até mesmo das aqui demonstradas, já que, pelo grande volume de dados, houve a necessidade de se realizar um recorte das publicações para viabilizar a análise, o que significa dizer que mais pesquisas podem se derivar a partir do exposto, dada a importância de se tratar de temas sociais no âmbito da Ciência da Informação, nesse caso, a partir da folksonomia sob a ótica da Organização e Representação do Conhecimento. Esta reflexão demonstra, ainda, que instrumentos utilizados na Organização e Representação do Conhecimento podem ser atuais e relevantes se colocados em cenários e posições diversas, acompanhando as mudanças tecnológicas, políticas e sociais.

## 5 REFERÊNCIAS

ASSIS, Juliana de; MOURA, Maria Aparecida. Folksonomia: a linguagem das tags. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 18, n. 36, p. 85-106, 2013. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2013v-18n36p85/24523>. Acesso em: 25 set. 2021.

BRASIL tem 134 milhões de usuários de internet, aponta pesquisa. **Agência Brasil**, Brasília, 26 maio 2020. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2020-05/brasil-tem-134-milhoes-de-usuarios-de-internet-apon-ta-pesquisa>. Acesso em: 26 set. 2021.

CASTELLS, Manuel. **A galáxia da internet**. Zahar: Rio de Janeiro, 2003.

CASTELLS, Manuel. A sociedade em rede: do conhecimento à política. In: CASTELLS, Manuel; CARDOSO, Gustavo (Org.). **A sociedade em rede**: do conhecimento à ação política. Belém: Imprensa Nacional, 2005.

CASTELLS, Manuel. **Redes de indignação e esperança**. Rio de Janeiro: Zahar, 2013.

DIAS, Luana Gomes; CASTRO, Hallison Phelipe Lopes de; SILVA, Márcio Bezerra da. Categorização de serviços da web 2.0: uma proposta de apoio aos bibliotecários. **Revista Folha de Rosto**, Juazeiro do Norte, v. 1, n. 2, p. 5-16, 2015. Disponível em: <https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/38/51>. Acesso em: 24 set. 2021.

GOIS, Ancelmo. CPI da Covid já gerou 9 milhões de comentários no Twitter; 33% são de apoio ao governo. **O Globo**, Rio de Janeiro, 20 maio 2021. Disponível em: <https://blogs.oglobo.globo.com/ancelmo/post/cpi-da-covid-ja-gerou-9-milhoes-de-comentarios-no-twitter-33-sao-de-apoio-ao-governo.html>. Acesso em: 24 set. 2021.

LE COADIC, Yves-François. **A ciência da informação**. Brasília: Briquet de Lemos/Livros, 1994.

MAAR, Wolfgang Leo. **O que é política**. São Paulo: Brasiliense, 1982.

MAGRANI, Eduardo. **Democracia conectada**: a internet como ferramenta de engajamento político-democrático. Curitiba: Juruá Editora, 2014. Disponível em: <https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/14106/Democracia%20conectada.pdf>. Acesso em: 30 set. 2021.

MORAES, Carolina; MENON, Isabella. CPI da Covid atrai público que ficou órfão do BBB e decola na era da 'política pop'. **Folha de S. Paulo**, São Paulo, 27 maio 2021. Disponível em: <https://www1.folha.uol.com.br/ilustrada/2021/05/cpi-da-covid-deslanca-na-era-da-politica-pop-e-atrai-publico-interessado-em-bbb.shtml>. Acesso em: 24 set. 2021.

MOREIRA, Vivian Lemes; ROMÃO, Lucília Maria Sousa. Discurso e(m) folksonomia: o Brasil vestido com algumas tags. **E-Compós**, Brasília, v. 13, n. 1, 2010. Disponível em: <https://1library.org/document/q76koxvy-discurso-e-folkosonomia-brasil-vestido-com-alguas-tags.html>. Acesso em: 25 set. 2021.

NÚMERO de usuários de Internet no mundo chega aos 4,66 bilhões. **Istoé Dinheiro**, São Paulo, 03 fev. 2021. Disponível em: <https://www.istoedinheiro.com.br/numero-de-usuarios-de-internet-no-mundo-chega-aos-466-bilhoes/>. Acesso em: 26 set. 2021.

O'REILLY, Tim. **What is web 2.0**: design patterns and business models for the next generation of software. 30 set. 2005. 5 p. Disponível em: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1>. Acesso em: 24 set. 2021.

RECUERO, Raquel. **Redes sociais na internet**. Porto Alegre: Sulina, 2009.

RIBEIRO, Ednaldo Aparecido; BORBA, Julian; HANSEN, Jaqueline Resmini. Participação *on-line* e *off-line* no Brasil: relações e condicionantes. **Revista do Serviço Público**, Brasília, v. 67, n. 4, p. 497-523, 2016. Disponível em: <https://revista.enap.gov.br/index.php/RSP/article/view/1262>. Acesso em: 25 set. 2021.

ROSENFELD, Denis Lerrer. **O que é democracia**. São Paulo: Editora Brasiliense, 2017.

SACONI, João. Nas redes sociais, manifestações de maio somaram 1,8 milhão de menções. **O Globo**, Rio de Janeiro, 29 maio 2021. Disponível em: <https://blogs.oglobo.globo.com/sonar-a-escuta-das-redes/post/nas-redes-sociais-manifestacoes-de-maio-somaram-18-milhao-de-mencoes.html>. Acesso em: 24 set. 2021.

CORRÊA, Renato Fernandes; SANTOS, Raimunda Fernanda dos. Análise das definições de folksonomia: em busca de uma síntese. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 23, n. 2, p. 1-32, 2018. Disponível em: <https://www.scielo.br/j/pci/a/wq76G649MfqdWHWtQkwwgGB/?format=pdf&lang=pt>. Acesso em: 24 set. 2021.

TWITTER, INC. **About**. 2021. Disponível em: <https://about.twitter.com/en>. Acesso em: 13 maio 2021.

WAL, Thomas Vander. **Folksonomy**. 2007. Disponível em: <http://www.vanderwal.net/folksonomy.html>. Acesso em: 24 set. 2021.

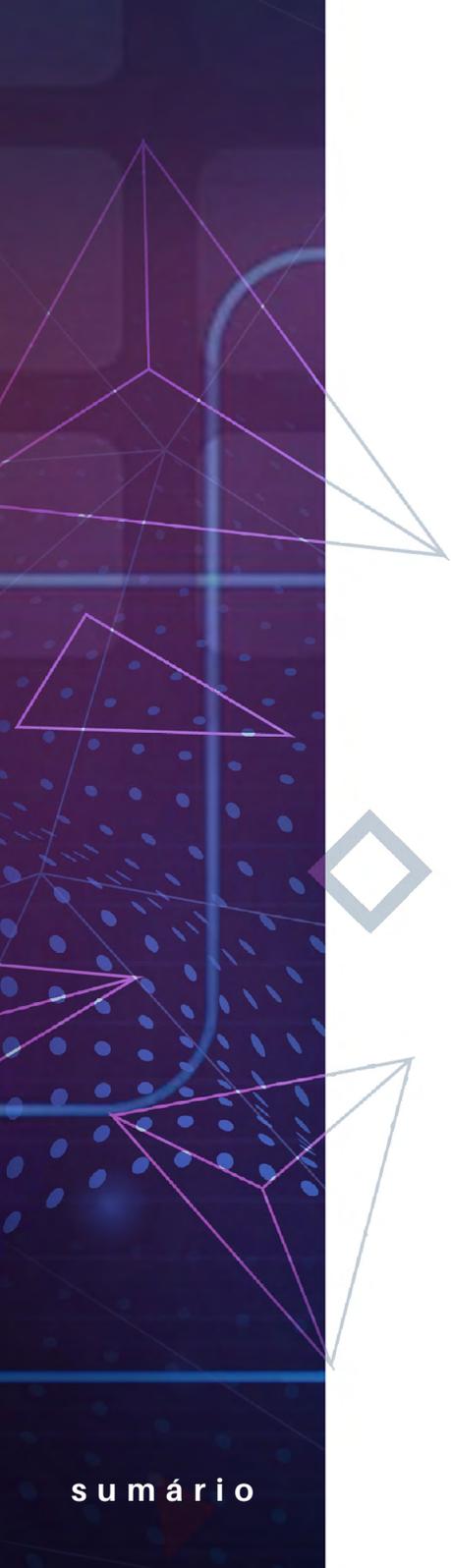
# 11

Andréa Fontoura da Silva  
Bruna Carballo Dominguez de Almeida  
Joana Peregrina Hernandes

**Organização e recuperação  
da informação:  
desafios na Arquivologia  
e na Biblioteconomia**

*Organization and information retrieval:  
challenges in Archival Science  
and Librarianship*

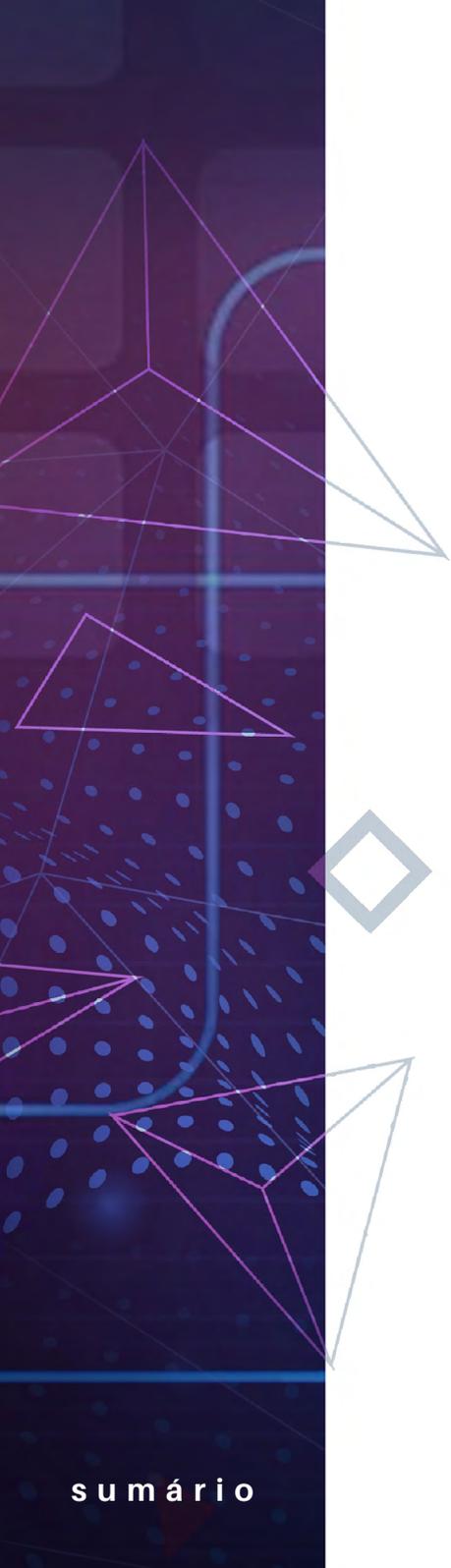
DOI: 10.31560/pimentacultural/2022.95613.11



**Resumo:**

Busca-se discutir a Organização do Conhecimento e recuperação da informação com foco no acesso de usuários em arquivos e bibliotecas. Aborda os conceitos de Organização do Conhecimento (OC) e estudo de usuários. Relaciona as atividades de arquivos e bibliotecas como primordiais para o acesso à informação e satisfação de seus usuários através de instrumentos que organizam e recuperam a informação.

**Palavras-Chave:** Organização do conhecimento. Recuperação da informação. Estudos de usuário.



**Abstract:**

*Discuss the Organization of Knowledge and information retrieval with a focus on user access to archives and libraries. It covers the concepts of Knowledge Organization (KO) and user study. Lists the activities of archives and libraries as essential for accessing information and satisfying their users through instruments that organize and retrieve information.*

**Keywords:** Knowledge organization. Information retrieval. User studies.

## 1 INTRODUÇÃO

A reflexão sobre a qual dissertaremos versa sobre o estudo da busca de informações pelos usuários em arquivos e bibliotecas com a ótica da Organização do Conhecimento (OC). O foco apresentado aqui visa inserir, no debate, o tema da recuperação da informação e sua relação com o usuário no contexto da OC, nas duas áreas: Arquivologia e Biblioteconomia.

Durante o período de 1972 a 2021, apenas 34 trabalhos foram realizados tratando, no título, Arquivologia e Biblioteconomia, nas pesquisas realizadas na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci). Vê-se que, em 49 anos, não chega a um trabalho científico por ano sobre a temática das duas áreas juntas. Nesse sentido, percebemos a importância e o alerta de que precisamos trabalhar mais em conjunto para minimizar as dificuldades, trocar experiências e refletir mais sobre a ótica da Organização do Conhecimento.

Salientamos a importância dos usuários nos processos de decisão ao procurar estratégias e formas de acessar, através da Organização do Conhecimento, informações pertinentes ao contexto dos arquivos e bibliotecas. Desta forma, práticas arquivísticas e biblioteconômicas são apropriadas ao estudo e desenvolvimento de estratégias para melhorar a recuperação da informação nestas instituições.

A reflexão aqui proposta tem como papel fundamental a abordagem de temas como difusão em arquivos e disseminação da informação nas bibliotecas, na qual pretende-se aproximar os estudos de recuperação da informação e estudos de usuários. Visando contribuir para os estudos da Organização do Conhecimento e seus efeitos sobre o acesso de usuários a informações em arquivos e bibliotecas, este trabalho apresenta, além da Introdução, uma seção destinada à Metodologia, bem como a Recuperação da Informação em arquivos e bibliotecas:

desafios, Organização do Conhecimento, na Arquivologia e na Biblioteconomia, Acesso, usuários e a OC, considerações finais e referências.

## 2 METODOLOGIA

A pesquisa aqui apresentada motivou-se a partir de estudos no Grupo de Pesquisa em Organização e Representação do Conhecimento – Abordagens Linguísticas em Arquivos e Bibliotecas – OrcaLab, da Universidade Federal do Rio Grande do Sul (UFRGS). Através de pesquisas em artigos relacionados à Organização do Conhecimento, tivemos a possibilidade de refletir sobre a importância de estudos na área.

A metodologia empregada esteve centrada na pesquisa bibliográfica, subsidiada por livros e artigos disponíveis de forma on-line em sua maioria. As fontes de informação utilizadas na pesquisa foram: Scientific Electronic Library Online - SciELO Brasil, Google Acadêmico, LUME Repositório Digital – Universidade Federal do Rio Grande do Sul (UFRGS).

A terminologia nas buscas se deteve em: Organização do Conhecimento, Recuperação da Informação, Usuário, Difusão Arquivística, Disseminação da Informação, Sistemas de Recuperação da Informação, Arquivologia e Biblioteconomia. A partir dos resultados obtidos, foram selecionados os textos que mais se adequaram à proposta da pesquisa.

## 3 A RECUPERAÇÃO DA INFORMAÇÃO EM ARQUIVOS E BIBLIOTECAS: DESAFIOS

Para que a recuperação da informação pesquisada pelo usuário faça sentido, é necessário que algumas medidas sejam adotadas dentro dos arquivos e bibliotecas. O conceito de recuperação da

informação aqui estudado é o de Belkin; Croft (1987)<sup>69</sup> *apud* Laipelt e Monteiro-Krebs (2021, p. 13):

A recuperação da informação efetiva-se em uma solicitação do usuário ao SRI, que, então, compara o que se solicitou com a representação dos documentos armazenados, por meio de um conjunto de procedimentos.

Entende-se SRI como Sistemas de Recuperação da Informação, que são “[...] aqueles que, entre outras funções, visam dar acesso às informações contidas em documentos neles registrados.” (ARAUJO, 2020, p. 1).

Na Biblioteconomia, quando precisamos orientar a indexação com vistas a atender os usuários do centro de informação (LANCASTER, 2004), precisamos também saber como o SRI captará as expressões de busca dos usuários. Podemos perceber que, além das bibliotecas, publicações periódicas e base de dados são orientadas a refletir sobre a terminologia usada e seus usuários. Ao estudarmos a recuperação da informação, vemos que o “todo” é muito mais do que apenas inserir uma palavra e recuperar o que se quer.

O “processo de recuperação da informação” envolve a representação dos documentos, a formulação da consulta, a computação da resposta e a comunicação da resposta. A representação de documentos – por meio de termos indexadores, descritores ou palavras-chave – pode ser feita por especialistas ou extraída automaticamente do texto integral do documento (LAIPELT; MONTEIRO-KREBS, 2021, p. 13).

Na Arquivologia, a tipologia documental se faz necessária nos registros das descrições dos documentos, pois reflete as atividades desenvolvidas de uma instituição. Nesse sentido, termos como certidão de nascimento, ata de reunião, plantas arquitetônicas, entre outros, são necessários ao elaborar um Sistema de Organização do

69 BELKIN, J. N.; CROFT, W. B. Retrieval techniques. *Annual Review of Information Science and Technology* (Arist), Silver Spring, v. 22, p. 109-145, 1987.

Conhecimento (SOC) para arquivos. Vemos, em Smit (2017, p. 20), que este nosso questionamento é considerado quando ela nos diz:

Embora não muito discutida, a indexação de conteúdo revela-se em certos casos determinante para evitar que o usuário deva pesquisar uma informação contida num respeitável volume de documentos, representantes da mesma tipologia documental, mas contendo informações variáveis.

Os SOCs precisam aparecer mais nos estudos da área para que o usuário tenha mais autonomia em arquivos Menne-Haritz (2001) apud. Smit (2017, p. 22) nos elucida neste sentido:

[...] a recuperação da informação e dos documentos, ou seja, o acesso a eles pelos usuários, somente adquire sua plena significação quando inclui uma preocupação didática e tradutora: não se trata de interpretar os documentos para o usuário, mas fornecer ao mesmo todas as informações que preservem o potencial informacional “original”, ou seja, o contexto original de produção do documento, abrindo caminho para que o usuário os interprete.

Percebe-se, então, que os arquivos precisam considerar mais as buscas dos usuários e suas necessidades, trazendo sua representatividade aos arquivos, para que, enfim, o arquivo tenha o seu papel social e de difusão da informação com verdadeira propriedade. Na busca de informações em bibliotecas, os estudos recentes mostram que a autonomia de pesquisa dos usuários pode ser alcançada com a mediação midiática, através de tecnologias da informação como *smartphones*, internet e redes sociais.

Portanto, é considerado o meio utilizado para dar visibilidade à Instituição, no sentido de apresentar sua função e conduzir a sociedade à ampliação de sua cultura. Desta forma, visa o acesso às informações mediante a elaboração de estratégias que criem uma aproximação dos indivíduos com informações referentes à sua história ou a conhecimentos específicos (ROCKEMBACH; DIAS, p. 337, 2018).

Cada vez mais necessitamos de estudos em Ciência da Informação e Tecnologia, visto que a interdisciplinaridade da área é evidenciada. Nela, podemos relacionar a Ciência da Informação com tantas outras teorias e práticas, tais como a Informática, a Biblioteconomia, a Arquivologia, a Museologia, a Documentação, a Ciência da Computação, a Comunicação, entre outras. Neste contexto, percebe-se como uma Ciência Contemporânea multifacetada permitindo integração com outras práticas, voltadas à interdisciplinaridade dos estudos. Nesse sentido, entende-se que se faz cada vez mais necessário adotar recursos que permitam um diálogo adequado às linguagens utilizadas pelos usuários.

Partimos do pressuposto que as organizações apresentam, cada vez mais, novas práticas informacionais, que requerem estratégias de recuperação da informação mais eficientes e eficazes. Portanto, é difícil ancorar a recuperação da informação contida nos documentos de arquivo apenas no instrumento de classificação. Precisamos incorporar nesse esforço novos instrumentos que possibilitem a recuperação da informação de modo compatível com as demandas informacionais atuais (SOUSA; ARAÚJO JÚNIOR, 2013, p. 133).

É preciso refletir sobre os recursos empregados para a recuperação da informação, pois a funcionalidade dos instrumentos deve ser um pré-requisito. Um instrumento que não é compreendido pelo usuário acaba dificultando a recuperação da informação, resultando no total oposto do que é esperado.

Segundo Costa<sup>70</sup> (2011 *apud* VAZ; ARAÚJO, 2015, p. 11), um fator a ser discutido

[...] é a falta de habilidade do usuário em utilizar os instrumentos de recuperação da informação nos arquivos: guias, catálogos, inventários e índices. Esses instrumentos de pesquisa são obras complexas, especializadas e de difícil elaboração, a ponto de não serem inteligíveis aos usuários de arquivo. É por isso que o arquivista deve se sentir responsabilizado por essa

70 COSTA, Marli Guedes da. Acesso aos arquivos públicos: aspectos jurídicos e práticos. **Cenário Arquivístico**, Brasília, DF, v. 4, n. 1, p. 22-31, jan./jun., 2011.

incompreensão das ferramentas que elabora. Cabe ao profissional de arquivo o treinamento do seu usuário para possibilitar que ele tenha acesso pleno à todas as informações disponíveis.

Na Arquivologia, os instrumentos tradicionais de pesquisa precisam de uma revisão teórica, pois, como apontam os autores, acabam dificultando a pesquisa e a recuperação da informação. Os instrumentos de busca e recuperação da informação precisam ter uma linguagem que converse com os usuários, bem como o uso deve ser intuitivo, poupando tempo de esforço a fim de tornar a pesquisa mais atrativa aos pesquisadores. Nesse sentido, estudos na área de recuperação da informação, sistemas de informação, disseminação da informação em bibliotecas e difusão em arquivos são temas em constante debate e percepção por parte da Ciência da Informação (CI).

## 4 ORGANIZAÇÃO DO CONHECIMENTO (OC)

Inicia-se esta seção com a Organização do Conhecimento, campo interdisciplinar teórico e metodológico relacionado à Ciência da Informação. Conforme definido por Hjørland (2016), a OC busca fazer representações de documentos, assuntos e conceitos, através de descrição e organização, por exemplo. Essa organização pode ser feita de forma manual ou por programas de computador. Assim, vislumbra-se aproximar a Organização do Conhecimento à Arquivologia e à Biblioteconomia para facilitar a construção de instrumentos de pesquisa e representação visando ao acesso à informação.

Ao construir metodologias para isso, a Organização e Representação do Conhecimento tem como propósito desenvolver ferramentas capazes de propiciar ao usuário a capacidade de acessar às informações arquivísticas de forma empírica e independente da ajuda do arquivista. Assim, a coleta de dados abre caminhos para a reinserção de algumas tendências na Ciência da Informação, como, por exemplo,

as linguagens, estruturas e classificações (BARROS; SOUSA, 2019). Inclusive estes são aspectos importantes para a estruturação de metodologias mais aplicáveis às necessidades de recuperação da informação e ao ímpeto investigativo do usuário. Essas necessidades e formas de interação entre usuário e informação são recentes. Estima-se que são tendências que surgiram nos últimos vinte anos e estão relacionadas à forma como ocorre a gestão administrativa que, consequentemente, tem impacto sobre como as pessoas constroem o conhecimento e buscam informações (BARROS; SOUSA, 2019).

Como uma tendência que se revela diante da Ciência da Informação e busca facilitar o acesso à informação de maneira dinâmica e centrada nas necessidades do usuário, os estudos relacionados à Organização do Conhecimento implicam a verificação de novos conceitos e, também, uma revisão dos processos que se apresentam desatualizados diante de um contexto totalmente tecnológico, digital e físico. De fato, a pesquisa de informação está carente de recursos sem a inserção de novos conceitos para que não precise estigmatizar o processo do acesso do usuário à informação.

As já mencionadas tendências de interação entre usuários e informação fomentam o desenvolvimento de todas as áreas que subsidiam essa relação. Arquivos e bibliotecas buscam, na Organização do Conhecimento, os subsídios e ferramentas para suprir as necessidades de usuários e tornar o acesso à informação dinâmico e independente.

#### **4.1 ORGANIZAÇÃO DO CONHECIMENTO NA ARQUIVOLOGIA**

Sob a perspectiva da organização e acesso, observam-se pontos em comum entre a Arquivologia e a Organização do Conhecimento. Suas relações são recentes e se estreitam a partir dos sistemas baseados em características do contexto de produção dos documentos,

que refletem na gestão, classificação, controle e acesso (BARROS; SOUSA, 2019; BARROS; TOGNOLI, 2015).

Dessa forma, a Organização do Conhecimento pode contribuir intimamente com as atividades desenvolvidas por arquivistas, principalmente no que tange à gestão documental. A base do tratamento arquivístico, dispensado a documentos e a informações, está pautada em seus contextos particulares de produção e utilização. Com isso, as metodologias e estudos do campo da Organização do Conhecimento podem contribuir para os arquivos, desde que respeitando tais contextos e particularidades.

Percebe-se que a Organização do Conhecimento trouxe conceitos que permitiram refletir melhor sobre a aplicação de metodologias advindas de outras áreas nas ferramentas arquivísticas, como, por exemplo, a linguística, a taxonomia, os tesouros, entre outros. Com esse efeito, os documentos foram inseridos em uma nova metodologia de classificação que permitiu que ferramentas de pesquisa e banco de dados tivessem estruturas dinâmicas e de fácil uso. Embarcamos em uma proposta de estudo que permite estender o campo de pesquisa da Arquivologia em relação aos estudos técnicos, pois vemos que alguns temas ainda não estão sendo explorados o suficiente em relação aos avanços dos arquivos, suas tecnologias, interação com os usuários, recuperação da informação e interdisciplinaridade da área.

Os documentos de arquivo registram ações que são representadas na organização destes, a partir da classificação, utilizada para evidenciar o vínculo arquivístico. Na Arquivologia, esse vínculo revela informações contextuais a respeito da produção desses documentos, tornando-se essencial para a recuperação das informações. Sendo assim, a classificação é uma das funções matriciais da Arquivologia, compondo parte fundamental do processo de gestão documental nos arquivos. Nesse processo de organização e representação, ressalta-se

a relação orgânica dos documentos, tão cara à Arquivologia (ROUSSEAU; COUTURE, 1998; SOUSA, 2009; SCHMIDT; SMIT, 2015).

No contexto atual, os documentos de arquivo são produzidos de forma híbrida propondo um novo cenário, no qual documentos físicos e digitais compartilham o espaço e, nesse sentido, os processos de representação e organização do conhecimento no contexto dos arquivos são estudados na Arquivologia de forma compartimentada e dissociada de uma visão sistêmica (BARROS, 2019, p. 77). Com isso, propõe-se uma relação mais estreita aos métodos da Organização do conhecimento.

Complementa-se a isso, o fato de que a classificação é o ponto de partida para os processos de organização dos arquivos (BARROS; TOGNOLI, 2015). Logo, é a partir dessa função que são pautados os demais procedimentos técnicos envolvidos na gestão documental. Ainda,

As atividades de descrição e classificação são, para os arquivos, o que são as atividades de classificação, catalogação e indexação para as bibliotecas, ou seja, a organização do conhecimento arquivístico em si (BARROS; TOGNOLI, 2015, p. 95).

Isso reforça as relações entre Organização do Conhecimento e Arquivologia, principalmente porque menciona etapas fundamentais do tratamento dispensado aos arquivos e que fazem parte da gestão documental. Além disso, ressalta processos-chave para a Organização do Conhecimento, que são a organização e representação de documentos, e traça um paralelo entre atividades desempenhadas em arquivos e bibliotecas.

## 4.2 ORGANIZAÇÃO DO CONHECIMENTO NA BIBLIOTECONOMIA

Há algum tempo, a Organização do Conhecimento, dentro da literatura especializada, vem trazendo diversas relações com a Biblioteconomia, quando busca, por meio de Sistemas de Organização do

Conhecimento – SOCs, refletir sobre os usos de instrumentos que qualificam a atividade do bibliotecário. Temos, na Biblioteconomia, vocabulários controlados, lista de assuntos, tesouros, estes sendo Sistemas de Organização do Conhecimento que pretendem auxiliar os usuários em suas buscas em adequação também aos Sistemas de Recuperação da Informação, aliado na disseminação da informação. Podemos refletir que, assim, os Sistemas de Organização do Conhecimento abrangem todos os tipos de esquemas de organização no que tange à informação, tais como: esquemas de classificação e cabeçalhos de assuntos, arquivos de autoridades, estando estes em arquivos, bibliotecas e museus (HODGE, 2001).

Nesse sentido, os Sistemas de Organização do Conhecimento precisam ser avaliados, estudados, percebidos com grande importância no papel de acesso à informação, visto estarem ligados também à disseminação da informação e aos usuários que os utilizam. Assim, a Organização do Conhecimento pelas linguagens documentárias são instrumentos privilegiados que podem apresentar dupla função: representar o conhecimento e promover interação entre o usuário e o conteúdo, promovendo, assim, o uso cada vez mais efetivo de bibliotecas. Contribuições como esta podem ser utilizadas pela Arquivologia para refletir sobre usuários e acesso, tema de estudos na Biblioteconomia.

Percebe-se que glossários, vocabulários controlados e tesouros auxiliam na revocação dos Sistemas de Recuperação da Informação (GABRIEL JÚNIOR; LAIPELT, 2019). Estudos desses SRIs, na Biblioteconomia, vêm avançando com, inclusive, novas formas de estudar as expressões de busca feitas nesses sistemas por usuários, através da metodologia de análise de *logs*, por exemplo, estuda por Laipelt (2015). Para minimizar problemas de comunicação entre bibliotecários e usuários, a autora escolheu

[...] como objeto de estudo as lexias de buscas contidas nos “logs” de pesquisa dos usuários do Portal LexMI do Senado

Federal Brasileiro, visto que, toda pesquisa realizada no Portal fica registrada em um arquivo log armazenado no servidor Web da instituição e, através de sua análise, é possível verificar a linguagem utilizada pelos usuários para a recuperação de informação (LAPELT, 2015, p. 16).

Assim, os estudos de Biblioteconomia e Organização do Conhecimento trazem grande discussão sobre como podemos utilizar os estudos há tempos realizados nessa área e refletir sobre como podemos, dentro da interdisciplinaridade, adentrar na organização da produção intelectual humana e também nos documentos oriundos de instituições, como na Arquivologia. Desse modo, Brascher e Carlan (2010) afirmam ser essenciais, em um Sistema de Organização do Conhecimento, os conceitos, que são as unidades básicas do conhecimento.

#### 4.3 ACESSO, USUÁRIOS E ORGANIZAÇÃO DO CONHECIMENTO

Pesquisas sobre estudo de usuários e acesso à informação na Arquivologia são recentes, devido a maior exigência dos usuários. Em outras áreas, como, por exemplo, na Biblioteconomia, são temas que possuem mais *know-how* sobre práticas, métodos e regras para o acesso a documentos e como podem ser desenvolvidas novas ferramentas que permitam que a informação possa ser recuperada de forma adequada. Um dos conceitos estudados foi o de Cavalcante, Galo, Vendramini (2017, p. 388-389):

[...] os estudos de usuários surgiram com o objetivo de mapear características de determinada população para que assim fosse possível planejar as informações mais adequadas a serem oferecidas, desenvolvendo estudos dirigidos no intuito de perceber se as necessidades de determinados grupos de usuários estavam sendo satisfeitas adequadamente, buscando oferecer um feedback aos sistemas.

Em bibliotecas especializadas, nas quais os usuários são mais exigentes e possuem, de regra, um maior conhecimento do assunto, o uso de vocabulários controlados traz um tratamento mais diferenciado a estes. A Linguagem Documentária, construída e utilizada para caracterizar o conteúdo dos documentos (MAIMONE, 2006), permite que toda informação gerada possa ter seu próprio código e, dessa forma, parametrizando formas de busca mais acessíveis aos usuários.

Vejamos que, na Arquivologia, Santos (2009) traz uma concepção atual sobre as funções arquivísticas<sup>71</sup>, reorganizando-as seguindo as tendências de um profissional dinâmico e centrado na informação como objeto de estudo e trabalho. Outros pesquisadores veem que os processos de organização arquivística geram produtos (BARROS, 2020), ou seja, que o resultado da organização do conhecimento permite a construção de ferramentas capazes de refletir a organização documental da instituição.

É importante observar que o tratamento das informações nos arquivos é um todo indissociável. Todos os produtos gerados na aplicação das funções são ferramentas que permitem dar continuidade à gestão. No caso da difusão e acesso, que são as últimas etapas do processo, ainda assim, não vemos um fim em si mesmos, já que é através deles que a informação ganha dinamização e visibilidade perante seu usuário. Nesse sentido, entende-se difusão como práticas exercidas no arquivo, perpassando todas as outras funções, ou seja, divulgando e propagando as ideias do arquivo (SANTOS, 2009).

Seguindo essa linha, Rockembach (2015) acredita que a difusão precisa ser realizada com uma abordagem interdisciplinar, abarcando algumas temáticas específicas, como a acessibilidade e transparência, estudo de usuários, comportamento informacional, mediação da

71 Os autores canadenses Rousseau e Couture apresentaram, em 1998, no livro *Os fundamentos da disciplina arquivística*, as funções: criação, avaliação, aquisição, conservação, classificação, descrição e difusão dos arquivos.

informação e literacia informacional. Percebemos aqui a necessidade constante de estudos de usuários, nas duas áreas, para melhorar o atendimento e vislumbrar novos serviços.

Ressalta-se que o grande trunfo da difusão é a aproximação que esta realiza entre o acervo e o usuário, residindo nisso o principal motivo para se difundir acervos com informações arquivísticas. Sendo assim, Barros (2020, p. 80) define, em seu texto, que “os processos de difusão e acesso lidam com o que existe de mais sensível na missão dos arquivos no cumprimento de servir à sociedade e dar acesso aos conjuntos documentais.”

Com base nas recentes pesquisas e encabeçando uma proposta de inclusão dos usuários, tem-se como pressuposto entender as necessidades destes, baseando-se nessa realidade e propiciando acessibilidade às instituições de arquivo. O estreitamento de relações entre o arquivo e seu usuário dá credibilidade à instituição na comunidade e visibilidade quanto ao seu potencial informacional e de pesquisa. Tudo isso colabora para a criação de uma identidade do arquivo com a sociedade ao seu redor, transformando-o em uma parte do cotidiano dessas pessoas.

A difusão propõe, ao arquivista, o papel de priorizar e instigar a comunidade a perceber o envolvimento e a influência sociocultural do arquivo na sociedade, tanto no seu aspecto administrativo quanto no aspecto histórico. Somente com projetos de difusão, é possível essa integração, além de propor uma nova relação do arquivista com a comunidade, que antes era de ser o guardião da informação para o facilitador ao acesso à informação.

Nas bibliotecas, o sentido de disseminação ou, mais precisamente, a disseminação da informação recai sobre a utilidade da informação e de como podemos transferi-la. A disseminação assume variadas formas que geram inúmeros produtos e serviços, dependendo do enfoque e da prioridade conferida (LARA; CONTI, 2003).

As formas de pesquisa abrangidas na parte inicial das telas de catálogos on-line de bibliotecas são importantes de se comunicar ao usuário, para que ele se sinta confiante em suas pesquisas. Muitos se distanciam dos catálogos e de sua própria pesquisa porque não conhecem e nem sabem como procurar no sistema. O treinamento de diferentes usos de tecnologias, assim como o uso dos catálogos de bibliotecas, por exemplo, é de extrema importância.

Devemos entender que o digital precisa de uma relação diálogo-perceptiva de ferramentas intuitivas, mas também de textos que sejam simples e fáceis de entender. Mais do que disseminar a informação ao usuário, é necessário refletir como faremos esta disseminação, quais meios atenderão suas expectativas e, também, torná-lo capaz de acessar independentemente de qualquer profissional, caso assim o queira.

## 5 CONSIDERAÇÕES FINAIS

A proposta que aqui deixamos é a de incentivar arquivistas e bibliotecários a aprofundarem estudos interdisciplinares, para terem mais proximidade com os usuários e adentrar mais na área de desenvolvimento de ferramentas para a recuperação da informação. Capacitar profissionais no desempenho de suas atividades e propor novos paradigmas para a arquivística na busca de evoluir o estudo teórico e profissional. Partindo desse ponto, verificar quais ferramentas podem ser aplicadas para melhorias do que está sendo discutido.

Muitas bibliotecas, desde o advento das Tecnologias de Informação e Comunicação (TICs), buscam utilizar catálogos eletrônicos para melhorar recuperar seus acervos. Estes não são apenas locais de descrição de documentos ali existentes, transportam a utilização de instrumentos da organização do conhecimento, como os vocabulários controlados e os tesouros para localizar diversos assuntos.

Deve-se refletir, no âmbito da Organização do Conhecimento, que, no campo de estudo da Biblioteconomia, o uso de forma mais reflexiva desses instrumentos pode trazer ao usuário uma maior capacidade de encontrar aquilo que necessita, de uma forma menos complicada. Vemos sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, tesouros, ontologias, entre outros que exigem normas e metodologias adequadas de construção e atualização para finalidades e aplicações específicas, mas que precisam sim, refletir as necessidades desses usuários. Assim, também influencia na visibilidade de acervos, pois estes precisam pensar sobre as necessidades dos usuários para que se indique a biblioteca, utilize-se da biblioteca e ela possa ter significado como tal.

Dessa forma, também percebemos que a Arquivologia busca visibilidade em seus arquivos a partir da função de difusão/acesso. Ou seja, após a aplicação de outras seis funções, dentre elas a classificação e avaliação, é que se tem a possibilidade de difundir acervos e instituições arquivísticas.

Nos arquivos, a Organização do Conhecimento se faz presente em alguns instrumentos, como nos planos de classificação e instrumentos de pesquisa, que são justamente os produtos da aplicação das funções arquivísticas. Logo, percebe-se que a Organização do Conhecimento aplicada à Arquivologia perpassa essas funções e reside na última delas, o conector entre arquivos e usuários, promovendo, dessa maneira, o acesso às informações.

A partir disto, entende-se que é compromisso dos arquivos estipular uma métrica para utilização dos documentos, ou seja, uma gestão documental capaz de proporcionar ao usuário formas padronizadas de recuperação à informação e não o contrário. Vemos que o usuário tem tomado o controle em atribuir a significância aos documentos, causando um mar de possibilidades de organicidade aos documentos e, com isso, prejudicando a recuperação da informação.

O relacionamento disciplinar das áreas da Arquivologia e Biblioteconomia é concernente à utilização de instrumentos de organização do conhecimento para difundir e disseminar essas instituições. Para a Biblioteconomia, os catálogos mostram uma relação interessante nesse sentido. Vida escrita (pessoal ou institucional – como nos arquivos), pesquisada, impressa e disponibilizada por diferentes aspectos de um material informacional, considerando instrumentos que trazem ao usuário maior credibilidade ao serviço, é o que bibliotecas e arquivos estão buscando, como mostra esse trabalho, de forma conjunta e reflexiva, para o desenvolvimento da Organização do Conhecimento.

## 6 REFERÊNCIAS

ARAUJO, Vania Maria Rodrigues Hermes de. Sistemas de recuperação da informação: a metodologia da tese. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 15, n. 4, p. 1-15, 2020. Disponível em: <https://brapci.inf.br/index.php/res/download/151941>. Acesso em 22 nov. 2021. DOI: 10.22478/ufpb.1981-0695.2020v15n4.56535 Acesso em: 22 nov. 2021.

BARROS, Thiago Henrique Bragato. Sem gestão não há difusão e acesso: aspectos histórico-conceituais da arquivística canadense e brasileira. **Acervo**, Rio de Janeiro, v. 33, n. 3, p. 68-85, set-dez. 2020.

BARROS, Thiago Henrique Bragato; SOUSA, Renato Tarciso Barbosa de. Organização do conhecimento e Arquivologia: abordagens metodológicas. **Informação & Informação**, Londrina, v. 24, n. 2, p. 76-92, maio/ago. 2019.

BARROS, Thiago Henrique Bragato; TOGNOLI, Natalia Bolfarini. Os processos de representação do conhecimento arquivístico: elementos históricos e conceituais da classificação e descrição. *In*: GUIMARÃES, J. A. C.; DODEBEI, V. **Organização do conhecimento e diversidade cultural**. Marília: Fundepe, 2015, v. 3, p. 94-99.

BRASCHER, Marisa.; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. *In*: ROBREDO, J.; BRASCHER, M. (Org.). **Passeios pelos bosques da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília: IBICT, 2010. p. 147-176. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso em: 10 jun. 2013.

CAVALCANTE, Luciane F. B.; GALO, Regina; A. C.; VENDRAMINI, Virgínia. **Estudos de usuário da Arquivologia**: reflexões. VII SECIN, UEL, 2017. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2017/secin2107/paper/viewFile/461/285>. Acesso em: 21 nov. 2021.

GABRIEL JUNIOR, Rene F.; LAIPELT, Rita do Carmo. F. Descrição das relações semânticas para aplicação em KOS: uso do Tesouro Semântico Aplicado (THESA). **P2P & Inovação**, Rio de Janeiro, v. 6 n. 1, p.117-135, Set.2019/Fev. 2020. Disponível em: <http://revista.ibict.br/p2p/article/view/4946>. Acesso em: 7 out. 2021.

HJØRLAND, Birger. Knowledge organization. **Knowledge Organization**, Toronto, v. 43, n. 6, p. 475-484, 2016. Disponível em: [http://www.isko.org/cyclo/knowledge\\_organization](http://www.isko.org/cyclo/knowledge_organization). Acesso em: 12 de jun. de 2021.

HODGE, Gail. **Systems of knowledge organization for digital libraries**: beyond traditional authority files. Washington, D.C.: 2001. The Digital Library Federation Council on Library and Information. Disponível em: [www.clir.org/pubs/abstract/pub91abst](http://www.clir.org/pubs/abstract/pub91abst). Acesso em: jun. 2021.

LAIPELT, Rita do Carmo Ferreira. **Metodologia para seleção de termos equivalentes e descritores de tesouros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. São Leopoldo, 2015. Tese (Doutorado) – Programa de Pós-graduação em Linguística, Universidade do Vale do Rio dos Sinos – UNISINOS. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2015. Disponível em: [http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/4853/RITA+DO+CARMO+FERREIRA+LAIPELT\\_.pdf?sequence=1](http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/4853/RITA+DO+CARMO+FERREIRA+LAIPELT_.pdf?sequence=1). Acesso em: 22 nov. 2021.

LAIPELT, Rita do Carmo Ferreira; MONTIERO-KREBS, Luciana. **Termos sob a superfície**: elementos teóricos, metodológicos e terminológicos para a representação do conhecimento. Rio de Janeiro: Interciência, 2021. 159 p.

LANCASTER, Frederick W. **Indexação e resumos**: teoria e prática. 2. ed. rev. atual. Brasília, DF: Briquet de Lemos, 2004. 452 p.

LARA, Marilda Lopes Ginez de; CONTI, Vivaldo Luiz. Disseminação da informação e usuários. **São Paulo em Perspectiva**, São Paulo, 2003, v. 17, n. 3-4, p. 26-34. Disponível em: <https://doi.org/10.1590/S0102-88392003000300004>. Acesso em: 20 nov. 2021.

MAIMONE, Giovana D. **Glossário de indexação e resumos**. 2006.

ROCKEMBACH, Moisés. Difusão em arquivos: uma função arquivística, informacional e comunicacional. **Informação Arquivística**, Rio de Janeiro, v. 4, n. 1, p. 98-118, jan-jun. 2015.

ROCKEMBACH, Moisés; DIAS, Eliane Carniel. Difusão em arquivos na Cinemateca Capitólio: Um Estudo de Caso. **Informação & Informação**, Londrina, v. 23, n. 1, p. 335 – 350, jan./abr. 2018. Disponível em: <http://www.uel.br/revistas/informacao/>. Acesso em: 20 nov. 2021.

ROUSSEAU, Jean-Yves; COUTURE, Carol. **Os fundamentos da disciplina arquivística**. Lisboa: Dom Quixote, 1998.

SANTOS, Vanderlei B. A prática arquivística em tempos de gestão do conhecimento. *In*: SANTOS, V. B. dos; INNARELLI, H. C.; SOUSA, R. T. B. (Org.). **Arquivística**: temas contemporâneos: classificação, preservação digital, gestão do conhecimento. 3. ed. Distrito Federal: SENAC, 2009.

SCHMIDT, Clarissa; SMIT, Johanna. Organização e representação da informação em arquivos: uma análise a partir da função classificação. *In*: XII Congreso ISKO España y II Congreso ISKO, Organización del conocimiento para sistemas de información abiertos, 19-20 nov. 2015, Murcia. **Anais** [...]. Murcia: Universidad de Murcia, 2015.

SMIT, Johanna. W. Recuperação, acesso e uso dos documentos arquivísticos. **Ciência da Informação**, Brasília, v. 42, n. 1, 2013. DOI: 10.18225/ci.inf..v42i1.1391 Disponível em: <https://brapci.inf.br/index.php/res/download/56102>. Acesso em: 19 nov. 2021.

SOUSA, Tarciso B. S. A classificação como função matricial do que-fazer arquivístico. *In*: SANTOS, V. B. dos; INNARELLI, H. C.; SOUSA, R. T. B. (Org.). **Arquivística**: temas contemporâneos: classificação, preservação digital, gestão do conhecimento. 3. ed. Distrito Federal: SENAC, 2009.

SOUSA, Tarciso B. S.; ARAÚJO JÚNIOR, Rogério Henrique. A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. **Ciência da Informação**, Brasília, v. 42, n. 1, 2013. DOI: 10.18225/ci.inf..v42i1.1400 Acesso em: 19 nov. 2021.

VAZ, G. A.; ARAÚJO, Carlos A. V. A importância dos estudos de usuários na formação do arquivista. **Informação Arquivística**, Rio de Janeiro, v. 4, n. 2, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/41831>. Acesso em: 18 nov. 2021.

# 12

Camila Monteiro de Barros

Edgar Bisset Alvarez

Desdobramentos da transcrição  
colaborativa nas abordagens  
teórico-metodológicas  
da organização do conhecimento

*Developments of collaborative transcription  
in knowledge organization  
theoretical-methodological approaches*

DOI: 10.31560/pimentacultural/2022.95613.12

### Resumo:

Este capítulo tem como objetivo discutir os desdobramentos teórico-metodológicos dos projetos de transcrição e descrição colaborativa de documentos para a área da Organização do Conhecimento (OC). Alguns exemplos desse tipo de projeto são o *Anti-Slavery Manuscripts* e *Scribes of the Cairo Geniza*, ambos hospedados na plataforma Zooniverse. Esses projetos contam com a participação de voluntários na tarefa de classificação e transcrição de documentos. No primeiro caso, trata-se principalmente de correspondências de cunho abolicionista no contexto norte-americano do século XIX. No segundo, fragmentos de documentos judaicos do período entre os séculos X e XIII d.C., escritos em hebreu e árabe formal e informal, estão à disposição dos usuários para que os transcrevam. As transcrições são realizadas com uma participação cultural diversificada e até diversa daquela do contexto do documento. A possibilidade de representação colaborativa da cultura global por meio do uso de plataformas tecnológicas nos aproxima das abordagens das Humanidades Digitais, que prevê o emprego de tecnologia no questionamento de temas humanísticos. Das temáticas que compõem domínios às temáticas abordadas em documentos, daquilo que será transcrito e reconstruído colaborativamente, serão derivadas interpretações de interesse da OC, sendo que as escolhas linguísticas e significativas realizadas pelos colaboradores influenciarão nessas interpretações. Dessa perspectiva, observamos alguns desdobramentos teórico-metodológicos para a OC: gênese da noção de “OC colaborativa”; ampliação do espaço para linguagens não especializadas, fomentando as discussões sobre pluralidade de saberes; perspectiva das humanidades digitais como ferramental tecnológico para a efetiva objetivação da pluralidade de saberes; para além da instrumentalização dos fazeres relacionados à OC, a área pode se apropriar dos impactos teóricos e vinculações sociais; consolidação das humanidades digitais como dispositivo potencializador do estudo das questões culturais humanas que são tão relevantes para a OC.

**Palavras-Chave:** Organização do conhecimento; representação do conhecimento; transcrição colaborativa; humanidades digitais.

**Abstract:**

*This chapter aims to discuss the theoretical-methodological developments of documental collaborative transcription projects for the area of Knowledge Organization (KO). Some examples of this type of project are Anti-Slavery Manuscripts and Scribes of the Cairo Geniza, both hosted on the Zooniverse platform. These projects have the participation of volunteers in the task of classifying and transcribing documents. In the first case, the project deals with North American anti-slavery documents, mainly correspondences, from the 19th century. In the second, fragments of Jewish documents from the period between the 10th and 13th centuries AD, written in formal and informal Hebrew and formal and informal Arabic, are available for users to transcribe. The transcriptions are carried out with a diversified cultural participation and are even different from that of the document's context. The possibility of collaborative representation of global culture through the use of technological platforms brings us closer to the approaches of Digital Humanities, which foresees the use of technology in questioning humanistic themes. From the themes that make up domains to the themes addressed in documents, from what will be transcribed and reconstructed collaboratively, interpretations of interest to KO will be derived, and the linguistic and significant choices made by the collaborators will influence these interpretations. From this perspective, we observe some theoretical-methodological developments for KO: genesis of the notion of "collaborative KO"; expansion of the space for non-specialized languages, fostering discussions on plurality of knowledge; perspective of the digital humanities as a technological tool for the effective objectification of the plurality of knowledge; in addition to the instrumentalization of actions related to the KO, the area can appropriate theoretical impacts and social bonds; consolidation of the digital humanities as a potentiating device for the study of human cultural issues that are so relevant to KO.*

**Keywords:** Knowledge organization; knowledge representation; crowdsourcing transcription; digital humanities.

## 1 INTRODUÇÃO

Entre a Organização do Conhecimento (OC), os aspectos culturais do conhecimento e o âmbito digital, deparamo-nos com as possibilidades ferramentais das Humanidades Digitais, que perpassa por todos esses aspectos e, nessa discussão, será cotejada com as discussões específicas da OC, permitindo trazer elementos de grande importância para o desenvolvimento, evolução e processamento do conhecimento nas ciências humanas, sociais e culturais, fortemente mediados por recursos digitais.

De acordo com alguns levantamentos, as abordagens que já vêm sendo desenvolvidas na relação entre OC e as possibilidades das humanidades digitais parecem dar maior foco ao desenvolvimento de sistemas de organização do conhecimento e acesso à grande quantidade de dados e documentos (ALBUQUERQUE; QUINTANA, 2021) e às questões de natureza aplicada, com menor atenção à teorização (ANDRADE, DAL'EVEDOVE, 2021). Nesse contexto, propomos uma reflexão sobre os desdobramentos teórico-metodológicos para a OC dos projetos de transcrição e descrição colaborativa de documentos que se pautam no uso de tecnologias e se aproximam, inclusive, da ideia de ciência cidadã.

As transcrições e descrições colaborativas dessa natureza são realizadas em linguagem natural e seu conteúdo tem relação não só com o conteúdo do próprio documento, mas também com as experiências das pessoas engajadas nos projetos. Portanto, a transcrição é realizada com uma participação cultural diversificada e até diversa daquela em que o documento está contextualizado. Nesse contexto, a organização social e a organização intelectual do conhecimento (HJØRLAND, 2008) se entrelaçam inadvertidamente.

Esse quadro abre espaço para diferentes interpretações a respeito, por exemplo, de textos religiosos, políticos e imagens arquitetônicas.

Julgamos que essa discussão é de fundamental importância para a OC, pois tem relação direta com a manutenção das heranças culturais, com questões de pluralidade linguística e de saberes, e, conseqüentemente, na análise de assunto, representação de domínio, construção de sistemas de organização do conhecimento.

Tendo como foco a Organização do Conhecimento na concepção de Hjørland (2008), em que essa área engloba perspectivas teóricas, passa pelo desenvolvimento e análise de instrumentos e abarca também o âmbito do próprio documento, sabemos que as questões de transcrição têm ligações com outras áreas. A Paleografia e a Diplomática, por exemplo, concatenadas à Arquivologia e aos estudos históricos têm como foco o estudo das diferentes formas de escrita, verificação da veracidade de documentos, sua legitimidade jurídica, entre outros aspectos. (BREWANGER; LEAL, 2020). Também estão relacionados com a temática que propõe a área da Linguística, com os estudos de tradução, a área de História e os estudos voltados à própria estrutura tecnológica de plataformas de colaboração.

Sem adentrar na especialidade dessas áreas, mas sim da OC, nosso foco é a flutuação da noção de domínio, assunto, suas várias possibilidades de contextualização e reconhecimento, oportunizadas pela ampla participação de pessoas, aproximando a discussão do fundamento cultural de variação de significados. Assim, questionamos: Como a OC é impactada por esse quadro? Como a área pode pensar a questão do domínio, das suas teorias fundantes, da representação, beneficiando-se das ferramentas tecnológicas na perspectiva das humanidades digitais?

Para tanto, trazemos alguns exemplos de projetos de transcrição colaborativa de documentos, contextualizamos os princípios das humanidades digitais e discorremos sobre as configurações da OC nas suas interseções com o debate da representatividade cultural, da variação das linguagens e dos significados. Para além da tecnologia

como finalidade e mantendo-nos alertas com relação “à crença excessiva no potencial e neutralidade dos algoritmos para identificar informações e fazê-las circular” (MOURA, 2018, p. 119), centramos os temas justamente na participação das pessoas no contato com documentos, sem definições algorítmicas prévias ou automatizadas. Pautamos a reflexão em um formato ensaístico, mas que aponta, nas suas conclusões, desdobramentos que impactam diretamente a OC e que são passíveis de vinculação teórica e social para a composição da área.

## 2 CONTEXTO DOS PROJETOS DE TRANSCRIÇÃO E DESCRIÇÃO COLABORATIVA

Silva e Borges (2018) demonstraram que, associados às ciências documentais, os projetos de transcrição de documentos antigos são tradicionalmente relacionados a instituições de memória. As autoras ponderam que, mais recentemente, os projetos dessa mesma natureza têm sido estruturados diretamente relacionados com projetos de Humanidades Digitais, englobando maior diversidade de temas, documentos e pessoas envolvidas, aproximando-se, em alguns casos, dos princípios da ciência cidadã.

Alguns exemplos desse tipo de iniciativa são o *Scribes of the Cairo Geniza* e *Anti-Slavery Manuscripts*. Ambos os projetos estão hospedados na plataforma Zooniverse<sup>72</sup>, considerada a maior no que se refere a projetos colaborativos de ciência cidadã e humanidades digitais, em diversas áreas. Criada em 2007, com o projeto *Galaxy Zoo* para identificação morfológica de galáxias (SIMPSON; PAGE; ROURE, 2014), atualmente, a plataforma abriga mais de 50 projetos ativos com cerca de 1,6 milhão de usuários inscritos (HERODOTOU

72 Diapornível em: <https://www.zooniverse.org/>. Acesso em: 19 jul. 2022.

*et al.*, 2020). Mais informações sobre a história e a estrutura da plataforma podem ser encontradas em Simpson, Page e Roure (2014) e informações sobre o considerável engajamento de voluntários e os tipos de tarefas por eles desempenhadas podem ser acessadas em Herodotou *et al.* (2020) e no site da própria plataforma.

No projeto *Scribes of the Cairo Geniza*<sup>73</sup>, ainda em andamento no momento de escrita deste texto, fragmentos de documentos judaicos, na sua maior parte com origem entre os séculos X e XIII d.C., estão à disposição dos usuários para que os classifiquem e transcrevam. Na classificação, que é a primeira parte do projeto, solicita-se aos voluntários que identifiquem a língua do documento, podendo ser hebreu formal ou informal e/ou árabe formal ou informal. As estruturas formais das línguas correspondem ao nível “fácil” da segunda etapa do projeto, a transcrição, enquanto a escrita informal, corresponde ao nível “desafiante”. Os documentos são originalmente provenientes da genizá (sala de armazenagem documentos que, pelo seu caráter religioso, devem ser queimados após a morte do dono, mas que curiosamente não o foram) da Sinagoga de Ben Ezra, em Fustat (Cairo antiga), no Egito. Diferente do que se esperava encontrar em uma genizá, foram mapeados documentos religiosos e documentos do dia a dia, como contratos e petições. No século XIX, quando da descoberta da genizá do Cairo, os documentos ficaram dispersos em coleções particulares de pesquisadores e colecionadores. Pouco a pouco, ao longo dos anos, foram incorporados aos acervos de várias bibliotecas universitárias e especializadas. Atualmente, os documentos que fazem parte do projeto *Scribes of the Cairo Geniza* (figura 1) são disponibilizados por instituições parceiras que têm a guarda desses acervos, como a biblioteca da Universidade da Pensilvânia (EUA), biblioteca do Seminário Teológico Judaico (EUA), biblioteca da Universidade de Cambridge (Inglaterra), entre outras (SCRIBES OF THE CAIRO GENIZA, 2021).

73 Disponível em: <https://www.scribesofthecairogeniza.org/>. Acesso em: 19 jul. 2022.

Figura 1 – Exemplos de Fragmentos de documentos disponíveis para transcrição no projeto *Scribes of the Cairo Geniza*



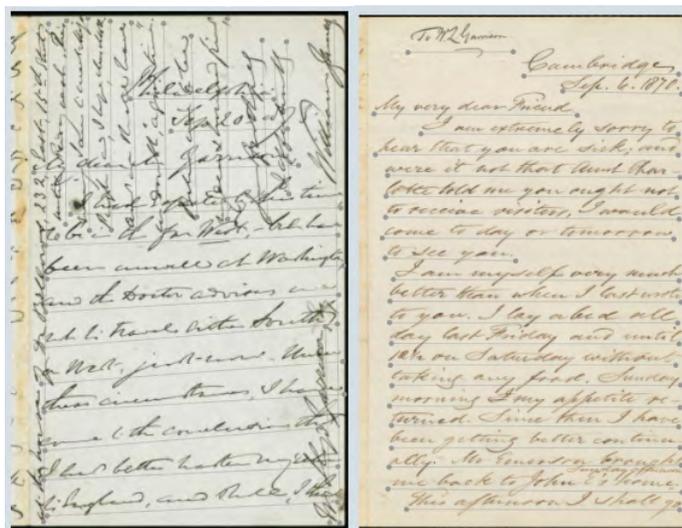
Fonte: *Scribes of the Cairo Geniza* (2021).

O projeto inclui designação de *tags* para os documentos e fóruns de discussão a respeito das escolhas de tradução, correspondência semântica das palavras adotadas atualmente com aquelas do período de produção do documento, prováveis usos dos documentos e outras questões. O projeto dos escritos do Cairo lida não só com a transcrição mas também com a tradução. As discussões são realizadas em inglês, assim como a identificação de palavras e frases. Dada a complexidade das tarefas, grande parte dos documentos encontra-se na fase de reconhecimento de caracteres do hebreu e do árabe para, a partir disso, poder-se trabalhar na transcrição e tradução de fato como, por exemplo, nos comentários a seguir extraídos da plataforma do projeto: “*This looks like liturgical composition with each line ending in the #letter\_beyt*”; “*The cursive Hebrew is hard to decipher. Has the purpose of this been identified??*”, “*Can someone confirm the character at the end of the fifth line from the bottom (image 2)?*”. Com a participação da equipe

que gerencia o projeto (na sua maior parte composta por bibliotecárias e bibliotecários) e dos voluntários, o projeto também desenvolve conjuntos de referências alfabéticas e de frases-chave para auxiliar no reconhecimento dos caracteres dos documentos. Essas espécies de dicionários são um produto secundário, ou seja, não são o objetivo do projeto, mas geram bases de conhecimento que podem permanecer após o término do projeto (SCRIBES OF THE CAIRO GENIZA, 2021).

O projeto *Anti-Slavery Manuscripts*<sup>74</sup>, finalizado em agosto de 2020, é composto por documentos abolicionistas norte-americanos, principalmente correspondências (figura 2), do período entre 1830 e 1870, que compõem a Coleção Abolicionista da Biblioteca Pública de Boston (EUA). O objetivo do projeto é realizar a transcrição integral dos documentos para, posteriormente, convertê-los em documentos legíveis por máquina e possibilitar a busca em texto completo.

**Figura 2 – Exemplos de documentos disponíveis para transcrição no projeto *Anti-Slavery Manuscripts***



Fonte: *Anti-Slavery Manuscripts* (2021).

74 Disponível em: <https://www.antislaverymanuscripts.org/>. Acesso em: 19 jul. 2022.

Hospedado na estrutura do Zooniverse, o projeto *Anti-Slavery Manuscripts* também disponibiliza tutoriais, possibilidade de designação de *tags* e fóruns de discussão. No caso desse projeto, trata-se somente da transcrição, o que não deixa de abrir espaços para variações linguísticas. Algumas discussões podem ser vistas em comentários como: “*Sometimes the paper they wrote on was very thin and there is bleed through from the verso. Tough to distinguish sometimes. Is that the case here?*”; “*We took a closer look at this and have concluded that March 21 is likely the correct date*”; “*I would agree with Holly and transcribe as ‘is!’*”.

Entre projetos das áreas de biologia, artes, física, natureza, espaço, o Zooniverse oferece uma estrutura para a participação colaborativa de pessoas nas mais diversas formas: identificação de ocorrência de espécies de animais e plantas, análise de mapas, análise de imagens microscópicas, transcrição de imagens arquitetônicas, catalogação de objetos museais etc. Alguns projetos visam utilizar a análise humana como *input* para aprendizagem de máquina com base em padrões.

No presente texto, escolhemos os dois projetos descritos anteriormente para que figurem como ilustração da discussão que segue nas próximas seções. O principal motivo da escolha é o foco na participação humana pela sua própria natureza de subjetividade. As escolhas, as discussões, dúvidas e sugestões sobre conteúdo, uso e contexto dos documentos são partes efetivamente integrantes do desenvolvimento desses projetos. Parece, então, que essas iniciativas, apesar de apenas ilustrativas da discussão, colocam em evidência que práticas no âmbito das humanidades digitais impactam no que se refere a pensar o conhecimento coletivamente.

### 3 POSSIBILIDADES NAS HUMANIDADES DIGITAIS

O surgimento e evolução da cultura digital têm facilitado a abertura, acesso e democratização do conhecimento e sua produção, no contexto digital. Segundo Risam (2018), “[...] a produção de conhecimento digital está envolvida em uma política de conhecimento da era da informação que não apenas reproduziu, mas também ampliou os valores culturais dominantes”. A evolução e o avanço acelerado das tecnologias excederam muito as expectativas para o futuro da humanidade, o que abre um desafiante campo de estudos, as “Humanidades Digitais”.

A cultura do digital modificou de maneira significativa as formas como entendemos e interagimos com a realidade. Esse fenômeno, sustentado pelos avanços e desenvolvimento cada vez mais acelerado das TICs, tem modificado as formas de comunicação, as formas de relacionamento social e até a forma de construção e gerenciamento de conhecimento, tendo a web como seu principal aliado nessa desconstrução do tradicional. Nesse sentido, Llanes-Padrón e Jorente (2017, p. 104, tradução nossa) afirmam que

As formas hipertextuais de apresentação da informação interferem na estética das interações e abrem novos conhecimentos em um redesenho (representação) da cultura. Com isso, as representações modificadas por essa estética – vinculadas à mídia web – e utilizadas como meio de difusão de informações e conhecimentos de muitos para muitos – criam novos mapas de contatos na rede de percepções e na cognição dos indivíduos.

O desenvolvimento e aperfeiçoamento de tecnologias voltadas para a organização, representação, visualização, acesso e apropriação da informação lançaram novas formas de pesquisa nas ditas ciências humanas e sociais, cobrando novas abordagens teórico-metodológicas, que acabaram produzindo um modelo interpretável e interativo de

uso, posse e apropriação do conhecimento. Segundo Pimenta (2020, p. 7-8) “[...] os recursos digitais tornaram-se ferramenta desejável – e caminho quase que compulsório para algumas pesquisas – na produção, na comunicação e na divulgação do conhecimento não somente para as ciências humanas, mas em todo cenário científico planetário”.

Dessa forma, estudos sobre humanidades digitais, considerados por Albuquerque e Hernández-Quintana (2021, p. 729, tradução nossa) como “[...] um conjunto de práticas ou uma comunidade de práticas de pesquisa onde a ligação é o uso de tecnologias digitais.”, apresentam trabalhos que juntam o conhecimento das áreas das ciências humanas, sociais, artísticas e linguísticas, sendo abordadas mediante o uso de técnicas, métodos e práticas advindas das ciências computacionais. Segundo Ortega (2013, p.18, tradução nossa),

[...] as Humanidades Digitais podem definir-se como o espaço de convergência entre ciências da computação, mídia digital e disciplinas humanísticas na busca de novos modelos interpretativos e novos paradigmas de conhecimento de acordo com as transformações operadas no seio da sociedade digital.

Por sua vez, autores como Pimenta (2020) declaram que as humanidades digitais podem ser consideradas um campo transdisciplinar, em que podem ser encontradas ações vinculadas às ciências humanas e sociais, com recursos e práticas oriundos das ciências computacionais, na busca por respostas aos novos cenários que vêm se apresentando produto da evolução da humanidade e na forma em como o digital tem influenciado a produção do conhecimento. Paletta (2018, p. 2) argumenta que “uma característica distintiva das Humanidades Digitais é o cultivo de uma relação bidirecional entre as humanidades e o digital: o campo emprega tecnologia na pesquisa e no questionamento de temas humanísticos”. As humanidades digitais, seguindo a linha de pensamento dos autores acima relacionados, podem ser consideradas um fenômeno interdisciplinar capaz de interligar e dar abrigo aos novos conhecimentos gerados na busca por soluções

aos problemas das ciências humanas e sociais, produto da inserção das tecnologias digitais.

Esse tipo de abordagem tem permitido um avanço nas formas de gerar, disseminar e compartilhar os conhecimentos gerados por pesquisadores, profissionais e especialistas das diferentes áreas de conhecimento. Esse cenário se constitui um desafio para os profissionais da informação, visto que, considerando a proximidade existente entre as humanidades digitais e a Ciência da Informação, entende-se que “[...] ambas podem ser consideradas ciências interessadas na informação, como, disciplinas interessadas no estudo do paradigma da pesquisa científica dos dados” (RÍO RIANDE, 2018, p. 2, tradução nossa). A forma como se organiza o conhecimento, com o objetivo de recuperá-lo, disseminá-lo e o torná-lo acessível para a satisfação das necessidades informacionais da sociedade, precisa de novos olhares, novos objetivos e novos enfoques epistemológicos, teóricos e práticos.

Nas palavras de Hjørland (2008), o processo de organização da informação e os sistemas de organização do conhecimento podem ser considerados um domínio de pesquisa, ensino e prática, fortemente vinculado à Ciência da Informação. Entretanto, sob o olhar das humanidades digitais, seria necessário entender as pesquisas de fronteira que permitam a abordagem interdisciplinar, híbrida e, principalmente, de trabalho em equipe, envolvidas no desenvolvimento de projetos como os descritos na seção anterior, voltados para garantir a organização, disseminação e acesso ao conhecimento resultado das atividades do ser humano.

Segundo Golub, Kamal e Vekselius (2021, p. 26, tradução nossa), “[...] o caminho para o avanço das humanidades digitais é um envolvimento mais profundo com a disciplina da organização do conhecimento”. Nesse sentido, as autoras ainda ilustram que, considerando a organização da informação e do conhecimento um dos principais

campos de estudo, tornou-se universal a necessidade cada vez maior de organizar sistematicamente as informações na era digital.

Esse cenário de profunda transformação nas práticas socioculturais, provocadas pelos avanços tecnológicos a nível global, provoca que projetos como *Anti-Slavery Manuscripts* e *Scribes of the Cairo Geniza* precisem de uma maior atenção em questões vinculadas à organização do conhecimento, para evitar problemas de sustentabilidade e escalabilidade. Por isso, profissionais, pesquisadores e especialistas que se dedicam às Humanidades Digitais, acostumados a encarar problemas como estes, conseguem trabalhar e desempenhar em plataformas híbridas, juntando o tradicional com as tecnologias, o especializado com cidadãos comuns.

## 4 ORGANIZAÇÃO DO CONHECIMENTO E LINGUAGENS

O exercício de objetificação do conhecimento traz consigo a possibilidade da manutenção de sua existência temporal e certa permanência espacial. No século XX, sedimenta-se o entendimento de que os estudos da linguagem, combinados à noção de estruturação do pensamento – ou arquitetura das ideias, como preferia Paul Otlet – poderiam revelar lógicas de percepção de mundo. Tal abordagem da filosofia da linguagem, embora não seja única, continua fortemente concatenada aos estudos da área de Organização do Conhecimento.

Entretanto, os significados, as referencialidades, as nomações expressivas das emoções, a semiótica e a fenomenologia do discurso, as “armadilhas da linguagem”, como registrou Marcondes (2017), desconhecem a ânsia da existência e permanência. No movimento incessante da língua e das ideias por meio dela expressas, a condição

humana de interpretação tem papel, senão responsabilidade, na constituição do conhecimento. Para Smiraglia (2014), as palavras fazem sentido àquele que as emprega e, uma vez socializado, os registros do conhecimento passam a compor cânones que, para o autor, são *corpora* de obras (*works*, na acepção de Smiraglia) que compõem significados culturalmente referenciados.

Nessa discussão, temos como foco os registros do conhecimento, individualmente compreendidos como documentos, mas que igualmente demonstram “uma impermanência a ser saboreada pelos seus receptores, volátil segundo a percepção decorrente dos fins divergentes para os quais podem ser consultados” (SMIRAGLIA, 2001, p. 196, tradução nossa). Hjørland (2008) afirma que os documentos são contextuais e seus significados estão ligados ao uso para os quais serão empregados.

A partir da diferença entre linguagem (conformação abstrata) e língua (fenômeno empírico) (MARCONDES, 2017), cabe retomar Foucault (2020) a respeito da ordem imposta pela estrutura da língua ao ato cultural do discurso. A língua se conforme como regra de estrutura do discurso, mas é preciso atentar-se à sua apropriação social para a emergência da perspectiva do acontecimento que permeia o discurso (FOUCAULT, 2020).

A transcrição colaborativa de documentos encontra-se nesse contexto complexo em que se tem à disposição a estrutura da língua dos documentos antigos (que pode ser mais ou menos conhecida) e as possibilidades de interpretação do discurso ali expresso, cujas especificidades precisam ser desvendadas. Defendemos que o tecido de significados (BARTHES, 1987) desses documentos precisa ser reconstruído no bojo dialógico entre passado, presente, cultura, estudo, especificidade, universalidade e engajamento plural.

De fato, não é possível submeter o arcabouço cultural unicamente à sua expressão linguística, no entanto, é necessário admitir que há uma conexão filosófica entre o conhecimento e a linguagem que o representa, sendo que sua compreensão pode ser inferida a partir de um âmbito cultural (SMIRAGLIA, 2014). Essa mesma situação se coloca com relação aos produtos das transcrições. Estes são também (novos) documentos, cuja estrutura da língua é empregada para expressar alguma interpretação.

Trazendo a reflexão para o âmbito mais particular da OC, Szostak (2014) questiona o significado de “universal”, pensando se seriam várias perspectivas particulares em consenso, se um sistema universal seria “um lugar para tudo”<sup>75</sup> e, em particular, se as realidades digitais contemporâneas requerem universalidade. A universalidade, então, descolada da noção de consenso, vem acompanhada da noção de pluralidade. A participação colaborativa de pessoas no desvelamento dos discursos de documentos e fragmentos, como nos projetos descritos anteriormente, é uma forma de pluralidade que traz desdobramentos importantes para a OC.

Ao sugerirem traduções e termos, as pessoas trazem suas próprias referências para o âmbito do documento e, da colaboração, surgem reconstruções dos significados dos saberes que chegaram até nós. A partir disso, derivam também reconstruções históricas que conferem certas delimitações às temáticas e aos domínios de conhecimento ali envolvidos. Sabe-se, na OC, que tanto as estruturas objetivas da sociedade quanto as estruturas subjetivas das relações e do pensamento humano competem para que se compreenda e se represente um domínio. Sobre a representação, temos o universo dos acontecimentos que somente o documento, no seu percurso, conhece e temos aquilo que se torna visível e acessível à análise da perspectiva da OC. Trata-se de um conhecimento “passado” reconstituído por outros autores/leitores,

75 “a place for everything”

com base nos parâmetros atuais da língua e do entendimento. Guedes, Moura e Dias (2011) demonstram que o uso da língua é a condição para o estabelecimento do diálogo (sincrônico ou diacrônico), já que é uma forma de linguagem pactuada entre os interlocutores.

Trazer à discussão da OC essa variação no entendimento do domínio é se apropriar do que, de fato, as Humanidades Digitais oferecem, ou seja, a vinculação social, humana e teórica da interação cultural. É importante registrar que reconhecemos as injustiças sociais que habitam a humanidade e que a representatividade cultural, nos mais diversos setores, ainda está longe de ser alcançada de fato. Distante de qualquer apelo “democrático” em relação à tecnologia e mais distante ainda de admitir a possibilidade de uma representação plural de forma acrítica, lançamos essa reflexão com vistas a trazer mais complexidade à abordagem teórica da OC que, muitas vezes, compreende a noção de “conceito” como um tipo de fenômeno homogêneo, sendo que os conceitos têm distintas naturezas (BARROS, CAFÉ, LAPLANTE, 2016).

As teorias socioculturais da OC, identificadas também como teorias críticas da OC, vêm fundamentando – ainda que sob outros aspectos – o foco que trazemos na presente discussão. No que se refere às metodologias das teorias críticas, Martínez-Ávila, Semidão e Ferreira (2016, p. 119, tradução nossa) afirmam que “na sua forma mais rica, combinam elementos/valores éticos, de busca por justiça social, pragmáticos, a partir da posição e ponto de vista de alguns grupos sociais que não são considerados por sistemas universais”. Garcia Gutierrez (2018, p. 56, tradução nossa) clama que “na era da transcultura digital, a verdade se reduzirá, de fato, à interação de valores, a uma verdade como processo, como trânsito, como precariedade, como impostura, como pós-verdade”. No âmbito da Epistemologia Social, os aspectos culturais aparecem principalmente a partir da ideia de organização social do conhecimento (VERONEZ JÚNIOR *et al.*, 2021).

Das temáticas que compõem domínios às temáticas abordadas em documentos, daquilo que será transcrito, traduzido e reconstruído serão derivadas interpretações de interesse da OC, sendo que as escolhas linguísticas e significativas realizadas pelos colaboradores influenciarão nessas interpretações e na “compreensão dos meandros discursivos próprios ao domínio no qual incide” (MOURA, 2018, p. 122).

No que tange à noção de representação do conhecimento, algumas questões permanecem contemporâneas: As temáticas são definidas pela linguagem do documento? O assunto está na gênese do documento, no contexto atual, na visão crítica daquele que analisa ou na revisão contextualizada de seu conteúdo? Está o assunto implicado na concepção (se não no direito) de escolha daquele que lê? Rondeau (2014) demonstra que a Ciência da Informação adota posições conflitantes sobre essas questões, mas que, no entanto, nunca estão completamente dissociadas daquele que lê o documento.

É no bojo das humanidades digitais que se pode exercitar um entendimento fenomenológico das ferramentas tecnológicas que permitem a interseção dessa miríade de possibilidades. Esses projetos colaborativos permitem que pessoas com diferentes perspectivas exponham suas interpretações. Se antes a linguagem especializada dava suporte à compreensão cultural de documentos, agora soma-se a isso a valorização de pontos de vista de não especialistas. Para Moura (2018, p. 128), “ao admitirmos outros contextos e fontes para corroborar os termos que constituirão as linguagens de indexação, por exemplo, torna-se possível retirar das sombras do rareamento inúmeras temáticas e agendas sociais.”

No fluxo entre os registros trazidos nos documentos e os novos documentos produzidos como fruto de transcrição e tradução, há, de fato, interferências culturais que trazem modificações substanciais pela ótica da terminologia (mais obviamente na tradução) e pela ótica das relações dos elementos contextuais que caracterizam aquele universo

de conhecimento. Por exemplo: ao estarem disponíveis para acesso, os documentos transcritos e, portanto, legíveis, tornam-se fontes para reconhecimento de domínios de conhecimento, levantamento terminológico, desenvolvimento de SOCs etc. Nesse contexto, de que domínio se ocupa a OC, de fato? Do documento histórico ora representado ou da expressão atualizada daquela representação? A “agenda social” que dali se desprende corresponde a que universo simbólico e linguístico? O que a OC organizará ou representará nesse caso? Se, por um lado, a tecnologia não representa, por si só, o desenvolvimento da área de OC (como muitas pesquisas fazem parecer especialmente no que se refere ao desenvolvimento de SOCs e tarefas automatizadas de análise de informação), por outro lado, a área de OC se desenvolve também no bojo das reconfigurações sociais, em que novos questionamentos de temas humanísticos se revelam no emprego da tecnologia.

Pode-se tecer uma aproximação com a teoria da desclassificação de Garcia Gutiérrez que prevê a desconstrução das referências hegemônicas fundamentais de visão de mundo. Para o autor, “desclassificação é um sistema aberto que instala pluralismo lógico no núcleo do entendimento e processos de enunciação, através de ferramentas metacognitivas”. (GARCIA GUTIÉRREZ, 2011, p. 5, tradução nossa). A partir do momento em que as referências culturais acabam se tornando preceitos que guiam a estruturação hierárquica de conhecimentos (LARA; MENDES, 2017), a participação colaborativa pode tecer gatilhos de desclassificação dado que as bases culturais seriam, a princípio, mais diversificadas que o fechamento do âmbito especializado em estudos disso ou daquilo. Além disso, por meio dos fóruns de discussão das plataformas, que são públicos, tem-se uma “fagulha” da necessária contrariedade na construção do conhecimento (GARCIA GUTIÉRREZ, 2011).

Da transcrição, depende grande parte da descrição dos documentos, causando impactos relevantes para a OC na perspectiva teórica e metodológica. É o caso na seleção e definição de termos e expressões representativas tanto da cultura trazida no documento quanto

das culturas da atualidade, colocando como elemento de extrema importância a possibilidade de representação coletiva da cultura global.

## 5 DESDOBRAMENTOS

A partir da reflexão – ainda inicial – exposta no texto, podem-se depreender desdobramentos à OC que, com vistas à provocação, apresentamos a seguir com um pouco mais de objetividade. O primeiro impacto diz respeito à construção colaborativa do próprio domínio de conhecimento e não somente da sua representação. Na sua esfera ideal e simbólica, as práticas sociais de um contexto passado, quando ressignificadas em outro contexto, podem gerar uma referência totalmente nova do que se entende daquela cultura. A partir de linguagens e interpretações não especializadas, os novos documentos figuram como os novos registros daquele conhecimento, fazendo-nos perceber a emergência da OC colaborativa.

No caminho da pluralidade, tem-se um desdobramento teórico que chama a OC a se apropriar do impacto das práticas de transcrição no seu âmbito de estudo. Nesse sentido, é interessante incluir o tema da colaboração especialmente nos estudos críticos da OC.

A partir daí, já podemos vislumbrar as humanidades digitais como potencialidade metodológica para algumas teorias críticas da OC. Essa perspectiva localiza a estrutura tecnológica no *status* de ferramenta para as humanidades na medida em que, a princípio, aumentam as chances de visibilidade de diferentes pontos de vista. No entanto, para além da questão instrumental, as humanidades digitais fazem também emergir questões sociais importantes para a OC.

Finalmente, sem intenção de esgotar o tema, percebe-se que existem inúmeras oportunidades na produção e compartilhamento do

conhecimento, a partir da adoção das técnicas, práticas e métodos propostos nas humanidades digitais, em que a organização do conhecimento ocupa um papel de suma importância, como mostrado nos projetos acima mencionados, delineando, de forma transparente, uma parceria entre ambas disciplinas que amplia a transdisciplinaridade e colaboração, que é defendida no presente trabalho.

## 6 REFERÊNCIAS

ALBUQUERQUE, Ana Cristina de; QUINTANA, Ania Rosa Hernandez. Sistemas de organização do conhecimento e humanidades digitais: possíveis interlocuções a partir da abordagem da análise do domínio. *In*: SILVA, Carlos Guardado da; REVEZ, Jorge; CORUJO, Luís (Eds.). Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde. CONGRESSO ISKO ESPANHA-PORTUGAL, 5, 2021, Lisboa. **Anais** [...]. Lisboa: Centro de Estudos Clássicos, Colibri, 2021. Disponível em: <https://repositorio.ul.pt/handle/10451/50067>. Acesso em: 19 jul. 2022.

ANDRADE, Laura Mariane de; DAL'EVEDOVE, Paula Regina. Aproximações entre organização do conhecimento e humanidades digitais. *In*: SILVA, Carlos Guardado da; REVEZ, Jorge; CORUJO, Luís (Eds.). Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde. CONGRESSO ISKO ESPANHA-PORTUGAL, 5, 2021, Lisboa. **Anais** [...]. Lisboa: Centro de Estudos Clássicos, Colibri, 2021. Disponível em: <https://repositorio.ul.pt/handle/10451/50067>. Acesso em: 19 jul. 2022

ANTI-SLAVERY MANUSCRIPTS. 2021. Disponível em: <https://www.antislaverymanuscripts.org/> Acesso em: 19 jul. 2022.

BARROS, Camila Monteiro; CAFÉ, Lúcia Maria Arruda; LAPLANTE, Audrey. Emotional Concepts in Music Knowledge Organization. *In*: CONGRESSO ISKO-BRASIL, 14, 2016, Rio de Janeiro. **Anais** [...]. São Paulo, Baden-Baden (Alemanha): Verlag, 2016, p. 164, 2016. Disponível em: <https://www.nomos-e-library.de/10.5771/9783956504389/knowledge-organization-for-a-sustainable-world-challenges-and-perspectives-for-cultural-scientific-and-technological-sharing-in-a-connected-society?page=1> . Acesso em: 19 jul. 2022.

BARTHES, Roland. **O prazer do texto**. São Paulo: Perspectiva, 1987.

BERWANGER, Ana Regina; LEAL, João Eurípedes Franklin. **Noções de paleografia e de diplomática**. Santa Maria: Fundação de Apoio a Tecnologia e Ciência, Editora UFSM, 2020.

FOUCAULT, Michel. **A arqueologia do saber**. 8. ed. Rio de Janeiro: Forense, 2020.

GARCÍA GUTIÉRREZ, Antonio L. **En pedazos**: el sentido de la desclasificación. Madrid: ACCI, 2018.

GARCÍA GUTIÉRREZ, Antonio. Desclassification in knowledge organization: a post-epistemological essay. **TransInformação**, Campinas, v. 23, n. 1, p. 5-14, 2011. Disponível em: <https://www.scielo.br/j/tinf/a/89vfv6PdSjGkRMr56Gqv-J/?format=pdf&lang=en>. Acesso em: 19 jul. 2022.

GOLUB, Koraljka; KAMAL, Ahmad M.; VEKSELIUS, Johan. Knowledge organisation for digital humanities. In: Golub, K., & Liu, Y.-H. **Information and Knowledge Organisation in Digital Humanities: Global Perspectives**. Routledge, 2021. Disponível em: <https://doi.org/10.4324/9781003131816>. Acesso em: 19 jul. 2022.

GUEDES, Roger de Miranda; MOURA, Maria Aparecida; DIAS, Eduardo José Wense. Indexação social e pensamento dialógico: reflexões teóricas. **Informação e Informação**, Londrina, v. 16 n. 3, p. 40 – 59, jul./ ago. 2011. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/10477/10642>. Acesso em: 19 jul. 2022.

HERODOTOU, Christothea; ARISTEIDOU, Maria; MILLER, Grant; BALLARD, Heidi; ROBINSON, Lucy. What Do We Know about Young Volunteers? An Exploratory Study of Participation in Zooniverse. **Citizen Science: Theory and Practice**, v. 5, n. 1, p. 2, 2020. Disponível em: <http://doi.org/10.5334/cstp.248>. Acesso em: 19 jul. 2022.

HJØRLAND, Birger. What is knowledge organization (KO)? **Knowledge Organization**, Baden-Baden (Alemanha), v. 35, n. 2-3, p. 86-101, 2008. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-86/what-is-knowledge-organization-ko-volume-35-2008-issue-2-3?page=1>. Acesso em: 19 jul. 2022.

LARA, Marilda Lopes Ginez de; MENDES, Luciana Cortes. Referências socio-culturais nos Sistemas de Organização do Conhecimento. **Iris – Informação, Memória e Tecnologia**, Recife, v. 3, n. especial, p. 26-44, 2014/2017. Disponível em: <https://periodicos.ufpe.br/revistas/IRIS/article/view/236185/28898>. Acesso em: 19 jul. 2022.

LLANES PADRÓN, Dunia; JORENTE, Maria José Vicentini. Complejidad, Representación y acceso a la Información archivística en los nuevos contextos paradigmáticos. In: JORENTE, Maria José Vicentini; PADRÓN, Dunia Ilanes (Orgs).

**Una Mirada a la ciencia de la información desde los nuevos contextos paradigmáticos de la posmodernidad.** São Paulo: Cultura Acadêmica, 2017.

MARCONDES, Danilo. **As armadilhas da linguagem.** Rio de Janeiro: Zahar, 2017.

MARTÍNEZ-ÁVILA, Daniel; SEMIDÃO, Rafael; FERREIRA, Marcio. Methodological aspects of critical theories in knowledge organization. **Knowledge Organization**, Baden-Baden (Alemanha), v. 43, n. 2, p. 118-125, 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/161468> Acesso em: 19 jul. 2022.

MOURA, Maria Aparecida. Organização social do conhecimento e performatividade de gênero: dispositivos, regimes de saber e relações de poder. **Liinc em Revista**, Rio de Janeiro, v. 14, n. 2, p. 118-135, nov. 2018. Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/46587> . Acesso em: 19 jul. 2022.

ORTEGA, Nuria Rodríguez. Humanidades Digitales, Digital Art History y cultura artística: relaciones y desconexiones. **Artnodes**, Málaga, n. 13, p. 16-25, 2013. Disponível em: <https://www.raco.cat/index.php/Artnodes/article/download/285089/372946>. Acesso em: 01 out. 2021.

PALETTA, Francisco Carlos. Ciência da Informação e Humanidades Digitais – uma reflexão. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 19., 2018, Londrina. Anais... Londrina: ANCIB, 2012. Disponível em: <https://brapci.inf.br/index.php/res/download/124826>. Acesso em: 01 out. 2021.

PIMENTA, R. M. Por que Humanidades Digitais na Ciência da Informação? Perspectivas pregressas e futuras de uma prática transdisciplinar comum. *Informação & Sociedade: Estudos*, [S. l.], v. 30, n. 2, 2020. DOI: 10.22478/ufpb.1809-4783.2020v30n2.52122. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/52122>. Acesso em: 01 out. 2021.

RÍO RIANDE, Gimena del. Humanidades Digitales CAICYT: Las Humanidades Digitales en la agenda de investigación de las Ciencias de la Información. **8ª JoBiCyT del congreso de ABGRA**. 24-26 abr 2018. Disponível em: [doi:10.5281/zenodo.3233995](https://doi.org/10.5281/zenodo.3233995). Acesso em: 15 set. 2021.

RISAM, Roopika. **New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy.** Illinois: North-western University Press, 2018.

RONDEAU, Sophie. The life and times of aboutness: a review of the library and information science literature. **Evidence Based Library and Information Practice**, Alberta, v. 9, n.1, p. 14-35, 2014. Disponível em: <https://journals.library.ualberta.ca/ebliip/index.php/EBLIP/article/view/19091/16224>. Acesso em: 01 out. 2021.

SCRIBES OF THE CAIRO GENIZA. 2021. Disponível em: <https://www.scribethecairogeniza.org/>. Acesso em: 05 out. 2021.

SILVA, Ana Margarida Dias da; BORGES, Leonor Calvão. A transcrição e a leitura de manuscritos entre o crowdsourcing e a participação cidadã. *In*: CONGRESSO BAD, 13, Lisboa, 2018. **Anais** [...]. Lisboa: Associação portuguesa de bibliotecários, arquivistas, profissionais da informação e documentação, 2018. Disponível em: <https://eg.uc.pt/bitstream/10316/81196/3/A%20transcri%c3%a7%c3%a3o%20e%20a%20leitura%20de%20manuscritos.pdf>. Acesso em: 18 set. 2021.

SIMPSON, Robert; PAGE, Kevin R.; ROURE, David. Zooniverse: observing the world's largest citizen science platform. *In*: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 14, Seul, 2014. **Anais** [...]. Seul: International World Wide Web Committee. <https://dl.acm.org/doi/pdf/10.1145/2567948.2579215>

SMIRAGLIA, Richard P. **The elements of knowledge organization**. Londres: Springer, 2014.

SMIRAGLIA, Richard P. Works as signs, symbols, and canons: the epistemology of the work. **Knowledge Organization**, Baden-Baden (Alemanha), v. 28, n. 4, p. 192-202, 2001. Disponível em: [doi.org/10.5771/0943-7444-2001-4-192](https://doi.org/10.5771/0943-7444-2001-4-192). Acesso em: 01 out. 2021.

SZOSTAK, Rick. How universal is universality? **Knowledge Organization**, Baden-Baden (Alemanha), v. 41, n. 6, p. 468-470, 2014. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2014-6-468/how-universal-is-universality-volume-41-2014-issue-6?page=1>. Acesso em: 30 set. 2021.

VERONEZ JÚNIOR, Wilson Roberto; ALMEIDA, Carlos Cândido de; MARTÍNEZ-ÁVILA, Daniel; RODRIGUEZ, Sonia Maria Troitino. Epistemologia social e organização do conhecimento: contribuições as abordagens culturais. *In*: SILVA, Carlos Guardado da; REVEZ, Jorge; CORUJO, Luís (Eds.). Organização do Conhecimento no Horizonte 2030: Desenvolvimento Sustentável e Saúde. CONGRESSO ISKO ESPANHA-PORTUGAL, 5, 2021, Lisboa. **Anais** [...]. Lisboa: Centro de Estudos Clássicos, Colibri, 2021. <https://repositorio.ul.pt/handle/10451/50067>

ZOONIVERSE. S./., 2021. Página inicial. Disponível em: <https://www.zooniverse.org/>. Acesso em: 19 jul. 2022.

# 13

Adriana Carla Ribeiro dos Santos

Raí Rocha Costa

Marcos Oliveira da Costa

*O software* Thesa como SOC  
no contexto arquivístico:  
uma proposta  
de modelo teórico

*Thesa software as SOC  
in the archival context*

*a proposal for  
a theoretical mode*



### Resumo:

Esta pesquisa tem como objetivo desenvolver uma proposta teórica para a adaptação do *software* Thesa (Tesauro Semântico Aplicado), aplicando-o para o contexto dos arquivos, com base no estudo das relações semânticas dos sistemas de organização do conhecimento (SOC). Entende-se como importante o desenvolvimento deste estudo, pois a problemática do acesso e utilização da informação é um dos fenômenos investigados pela Ciência da Informação (CI). A utilização da Organização do Conhecimento (OC) por meio de SOCs como o Thesa é necessária para as instituições de forma geral, e principalmente em arquivos, tendo em vista que o controle semântico auxilia em uma melhor recuperação da informação no âmbito destes. O estudo define-se como pesquisa exploratória, quanto a abordagem do problema, configura-se como qualitativa e, para discutir os marcos teóricos da área e aplicá-los a realidade que será estudada, o instrumento de coleta de dados foi o bibliográfico. Por fim, baseando-se nos dados que foram demonstrados, identificou-se as funcionalidades do Thesa e apontou-se o este como um possível SOC eficaz no processo de organização do conhecimento arquivístico.

**Palavras-Chave:** Organização do conhecimento; sistema de organização do conhecimento; Thesa; arquivos.

**Abstract:**

*This research aims to develop a theoretical proposal for the adaptation of Thesa (Applied Semantic Thesaurus) software, applying it to the context of archives, based on the study of the semantic relationships of knowledge organization systems (SOC), it is understood the development of this study is important, as the issue of access and use of information is one of the phenomena investigated by Information Science (CI). The use of the Knowledge Organization (CO) through SOCS such as Thesa is necessary for institutions in general, and especially in archives, considering that semantic control helps in a better information retrieval in their scope. The study is defined as exploratory research, as the approach to the problem is configured as qualitative and to discuss the theoretical frameworks of the area and apply them to the reality that will be studied, the data collection instrument was the bibliographic. Finally, based on the data that were demonstrated, the functionalities of Thesa were identified and it was pointed out as a possible SOC effective in the process of organizing archival knowledge.*

**Keywords:** Knowledge organization; knowledge organization system; Thesa; archives.

## 1 INTRODUÇÃO

Os estudos sobre Organização do Conhecimento (OC) expandem-se cada vez que surgem novas discussões sobre a sistematização e organização da informação para promover a ação do conhecimento. Nesse processo, acabamos encontrando com a Organização e Recuperação da Informação Arquivística (ORIA), um campo ainda pouco explorado e estudado que surge com as iniciativas de sistematização e organização da informação em arquivo. Na arquivologia, o objeto de estudos são as informações orgânicas registradas, em que seu conteúdo informacional apresenta o processo comunicativo humano, para, assim, promover o processo comunicativo da informação arquivística, pensando na padronização da linguagem por meio da adoção do instrumento terminológico, o tesouro.

Este artigo aborda temas referentes a OC no domínio da arquivologia, propondo uma nova perspectiva para o acesso da informação arquivística. Objetiva-se desenvolver uma proposta teórica para a adaptação do software Thesa (Tesouro Semântico Aplicado), aplicando-o para o contexto dos arquivos, com base no estudo das relações semânticas dos sistemas de organização do conhecimento (SOCs).

A justificativa dessa pesquisa está relacionada à importância do desenvolvimento prático das ações de organização e sistematização da informação nos arquivos, considerando suas normatizações, pois a problemática se dá pelo acesso e utilização da informação que é considerado um dos fenômenos de investigação da Ciência da Informação (CI). Portanto, para entendermos melhor a influência da OC na arquivologia, faremos uma contextualização da construção do conceito e de sua definição durante os últimos anos, ligado ao SOC – que são estruturas organizadas que objetivam a construção de padrões abstratos da realidade, representando os conceitos de um domínio.

Há poucas referências relacionado ao tema de tesouro para o uso em arquivos na Ciência da Informação. O Thesa foi criado como um instrumento para a elaboração de tesouros, a fim de reduzir o trabalho operacional e priorizar o trabalho cognitivo e conceitual referente a modelagem do domínio (GABRIEL JUNIOR; LAIPELT, 2017). Esse *software* é indispensável para as instituições de forma geral, e, principalmente, em arquivos, tendo em vista que o controle semântico auxilia em uma melhor recuperação da informação no âmbito destes.

## 1.1 METODOLOGIA

No que diz respeito aos aportes metodológicos, essa pesquisa define-se como de natureza exploratória. De acordo Gil (2002, p. 41), isso se dá pois esta tem por objetivo “proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses. Pode-se dizer que estas pesquisas têm como objetivo principal o aprimoramento de ideias ou a descoberta de intuições”. Levando em consideração que essa pesquisa tem como problemática central criar uma proposta teórica (hipótese) para o *software* Thesa, visando a sua aplicação em arquivos, entende-se que se enquadra como um estudo exploratório, uma vez que analisa um objeto que já existe e pretende criar hipóteses com o objetivo de aprimorar ideias e descobertas, ou seja, é condizente com a definição acima exposta.

Quanto a abordagem do problema, define-se como qualitativa, na medida em que analisa os dados de forma comparativa e sistemática, mas sem a necessidade de se quantificar os resultados por meio de cálculos e métricas (MINAYO, 2002, p. 22). Esse tipo de abordagem é importante para pesquisa, pois, para Prodanov e Freitas (2013), o pesquisador mantém contato direto com a sua “fonte de estudo”, podendo analisá-la e criar hipóteses a partir disso. No caso da pesquisa em questão, essa abordagem é pertinente, levando em consideração

que os pesquisadores mantiveram contato com o objeto de estudo (Thesa) e suas funcionalidades básicas, para que se pudesse criar sugestões sobre a adaptação deste para o universo do arquivo.

No que concerne ao método de coleta de dados, este se deu por meio de pesquisa bibliográfica. De acordo com Marconi e Lakatos (2003), por mais simplório que possa parecer, esse método requer detalhamento no que concerne a suas fases de elaboração, que vai desde a coleta dos dados, a seu fichamento análise, comparação com a realidade que se está investigando e, por fim, a produção textual.

## 2 ORGANIZAÇÃO DO CONHECIMENTO E DA INFORMAÇÃO

Para que a informação documental seja entendida, ela precisa ser apreendida, registrada e armazenadas em sistema de informação, para, assim, cumprir seu objetivo de recuperação e uso. A organização acontece quando reunimos o que conhecemos em uma estrutura sistematicamente organizada, de forma que possa se representar a informação. Entende-se o conceito de Organização do Conhecimento (OC), segundo Dahlberg, como:

A ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento (características) inerentes e a aplicação esses conceitos e classes de conceitos ordenados a objetos/assuntos (DAHLBERG, 2006, p. 12).

A OC conseguiu se inserir na área da ciência mediante estudos realizados por pesquisadores como Dahlberg e pelos membros da *International Society of Knowledge Organization* (ISKO), e, a partir de 1973, decidem reconhecer a OC como área de estudo. Desenhando cada vez mais a OC, Barité (2001, p. 42-53) apresenta dez princípios

básicos que justificam o *status* científico e intelectual da Organização do Conhecimento (OC):

1. o conhecimento é um produto social, uma necessidade social e um dínamo social;
2. o conhecimento se realiza a partir da informação e ao socializar se transforma em informação;
3. a estrutura e a comunicação do conhecimento formam um sistema aberto;
4. o conhecimento deve ser organizado para seu melhor aproveitamento individual e social;
5. existem muitas formas possíveis de organizar o conhecimento;
6. toda organização do conhecimento é artificial;
7. o conhecimento registra sempre em documentos, como conjunto organizado de dados disponíveis, e admite usos indiscriminados;
8. o conhecimento se expressa em conceitos e se organiza mediante sistemas de conceitos;
9. os sistemas de conceitos se organizam para fins específicos, funcionais ou de documentação;
10. as leis que regem a organização de sistemas de conceitos são uniformes e previsíveis, e se aplicam por igual a qualquer área disciplinar.

Dessarte, a OC por muitos anos estava presente apenas nas discussões da biblioteconomia que a viam como uma prática operacional interna das bibliotecas para a classificação bibliográfica, atividade essa que possui processos em que há criação de instrumentos para representar conteúdos temático de documentos e por resultado aumentando a eficácia da recuperação (SILVA, 2021).

Hjørland (2008) comenta sobre atividades que a OC inclui como descrição de documentos e indexação em bibliotecas, bancos de dados bibliográficos, arquivos e desenvolvimento de algoritmos de computador realizados por bibliotecários, arquivistas, especialistas em informação, especialistas em assuntos e leigos. Institucionalizando essas atividades em sua organização, Smit (2012), discorre que “para poder cumprir sua missão, ao selecionar, organizar e disponibilizar a informação, as instituições atribuem um selo de qualidade àquela informação” e, assim, gera ciclos infinitos de informação. De acordo com Guimarães (2008), um conhecimento que é produzido, registrado, organizado, socializado e apropriado, gerará um novo conhecimento que também passará pelo mesmo movimento, constituindo um fluxo helicoidal de informações infinito.

A aproximação da OC com a Arquivística veio acontecer algum tempo depois. A arquivística, ainda no século XX, estava se redescobrimo e seus conceitos aprimorando-se. Ao entrar no século XXI, novos desafios são propostos pelas novas tecnologias de informação e novas formas de produção documental se estabelecem. O arquivista da “era da informação” deve estar atento a essas novas construções que circundam e implementam as atividades arquivísticas, pois novas teorias chegam para repensar e aprimorar as funções bases da arquivística. Como Terry Cook (1997) comenta:

Na introdução de sua análise sobre a história do pensamento arquivístico ao longo de um século, retoma que a mudança de postura do arquivista com relação à intervenção nos processos de produção e de recuperação do documento demonstraram sua capacidade de compreender a natureza do documento de arquivo, interpretando sua função, estrutura, processo administrativo e contexto documental, indo além do conteúdo histórico, alterando, assim, o valor de imparcialidade atribuído ao arquivista e aos documentos de arquivo, por exemplo. (COOK, 1997, p.19-20, tradução nossa).

Quando o objeto de estudo da arquivologia passou a ser a informação arquivística e de modo sistematizado, o propósito era organizar o conhecimento específico produzido e registrado pela sociedade e que

tivesse como resultado a sua recuperação e o acesso. Diferente da biblioteconomia, na arquivologia, a organização dos documentos se estabelece através do contexto, e não do conteúdo, respeitando sua proveniência e ordem original, que são princípios fundamentais da arquivística. Assim, as atividades e funções desempenhadas pelo arquivista se dão quando há o reconhecimento do fundo do arquivo para operacionalização (ROUSSEAU; COUTURE, 1998, p. 79).

Para poder apresentar ao usuário a informação produzida pelo fundo, é necessário um conjunto de atividades arquivísticas. Destacamos aqui a classificação e a descrição arquivística, consideradas funções primordiais para a organização e representação do conhecimento arquivístico. As duas funções são parte de processo intelectual que também é materializado por meio dos instrumentos de pesquisa resultados dos processos de representação do conhecimento arquivístico e entendidos como uma ponte entre o usuário e o documento (TOGNOLI, 2012, p. 89).

Analisando a aplicação dos sistemas e processos de OC sob o ângulo da arquivística, vemos que a preocupação se dá em cima do contexto de produção. Evidenciando o contexto de produção documental, “podemos compreendê-lo enquanto a própria essência da Arquivologia o que a torna singular dentro da OC”, como explica Silva (2021, p. 35). Os estudos de vocabulários controlados e a elaboração de tesouros funcionais, para a Organização do Conhecimento Arquivístico, são introduzidos pelas autoras Alencar e Cervantes:

Verifica-se que o documento de caráter arquivístico encontra sua razão de existência nas diferentes finalidades funcionais do órgão produtor. Assim, nota-se que um vocabulário controlado representando, portanto, as funções, irá ser coerente com o fazer arquivístico, uma vez que, pelo princípio da proveniência, o caráter norteador para organizar e representar, está no sentido que originou a ação (ALENCAR; CERVANTES, 2017, p. 02).

Segundo as autoras, a produção desses instrumentos possibilita entender a função que deu origem ao documento, uma estratégia

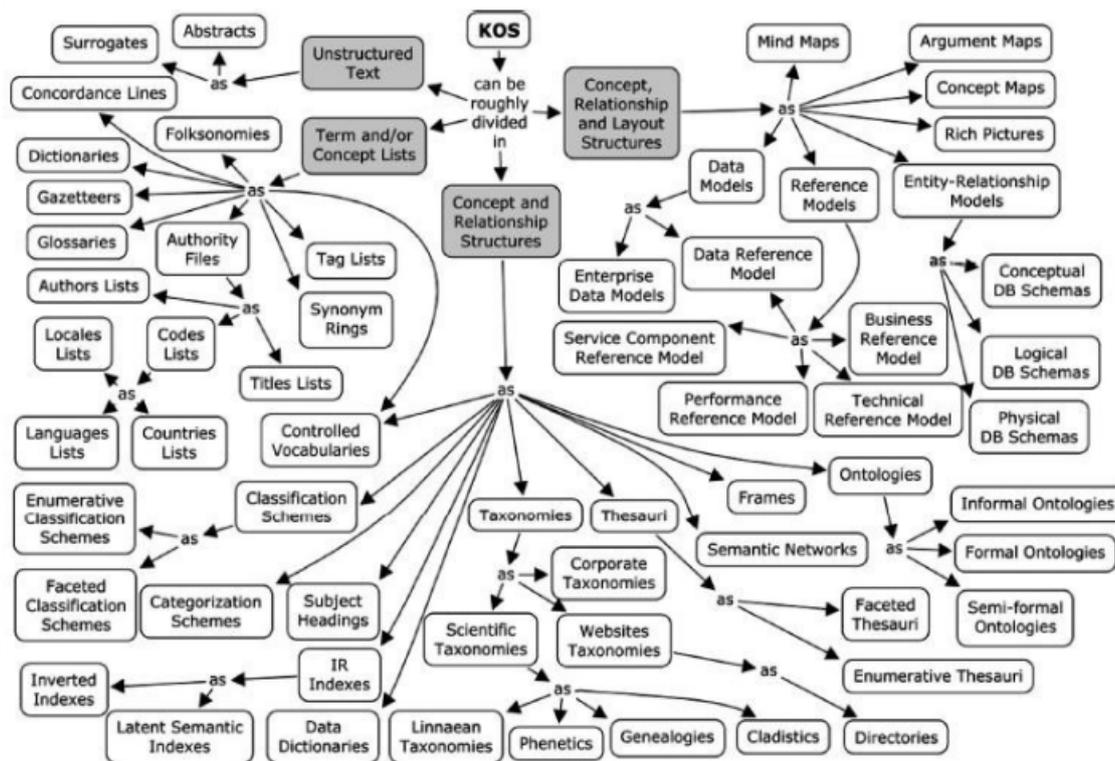
do acesso à informação. O vocabulário controlado é o instrumento criado para normalizar os termos e vocábulos que representam as informações, com a finalidade agilizar a busca e o acesso, o controle deste tem o objetivo de recuperar e organizar as informações, aprimorando, assim, o acesso. Para Smit e Kobashi (2003), o vocabulário controlado é criado para padronizar e facilitar a entrada e a saída de dados de um sistema de informações, promovendo maior precisão e eficácia na comunicação entre os usuários e os sistemas de informações. Então, a apresentação dos termos padronizados representam uma nova perspectiva para a área arquivística.

### 3 SISTEMA DE ORGANIZAÇÃO DO CONHECIMENTO

Os Sistemas de Organização do Conhecimento (SOCs) são instrumentos de representação do conhecimento que foram estabelecidos pelo *Networked Knowledge Organization Systems Working Group*, em uma Conferência da *ACM Digital Libraries* em 1998, *Pittsburgh na Pennsylvania*. O termo (SOC) é do original inglês “*Knowledge Organization System*” (KOS). A sigla KOS é aplicada com frequência na literatura, em português, usa-se a sigla SOC. Hjørland (2008) refere-se aos SOCs como ferramentas que colaboram na interpretação da Organização do Conhecimento, conhecidas como ferramentas semânticas.

Segundo Hodge (2000), os SOCs são todos os tipos de instrumentos cujo objetivo é organizar a informação e promover a ação do conhecimento, os quais incluem os sistemas de classificação que organizam materiais, os cabeçalhos de assunto para fornecer o acesso mais detalhado, os catálogos que controlam versões variantes de informação, como nomes geográficos ou nomes de pessoas e outros esquemas, as redes semânticas, tesouros, taxonomias e ontologias. Como mostra a figura 1:

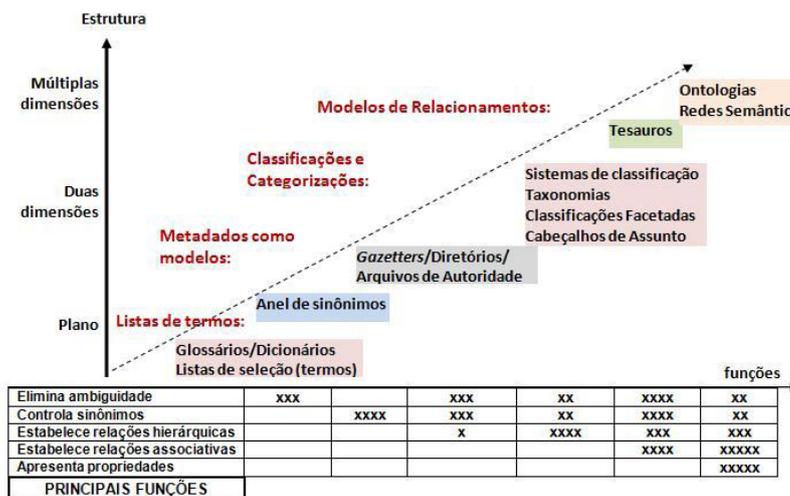
Figura 1 – SOC



Fonte: Souza *et al.*, (2012).

Assim, organizar a informação é a função de um SOC, por isso, são denominados como o “coração” dos Sistemas de Recuperação da Informação (SRI) de museus, bibliotecas e arquivos, tanto em ambiente físico quanto na web, segundo Hodge (2000). Já a classificação de Zeng (2008), apresentada na figura 2, demonstra dois grupos separados, ou seja, Listas de Termos e Modelos do tipo Metadados, que são diferentes das propostas por Hodge e Souza *et al.* (2012), que formam um único grupo.

Figura 2 – Os SOCs



Fonte: Zeng, 2008.

Ressalta-se, portanto, que os SOCs são estruturas organizadas que objetivam a construção de padrão abstratos da realidade, representando os conceitos de um domínio. Geralmente as ferramentas semânticas são usadas para o tratamento da informação, possibilitando a recuperação da informação, em ambiente informatizado ou tradicional.

Na Web Semântica, as ferramentas para desenvolvimento de SOCs estão se popularizando, principalmente os tradicionais, devido à necessidade de compartilhamento de padrões orientados por ontologias. Atualmente, os sistemas de busca são os principais meios de localização de informação na web. Porém, há uma certa confiabilidade nesses resultados de busca, o que significa que nem sempre encontramos aquilo que realmente desejamos.

### 3.1 TESAURO APLICADO AO ARQUIVO

A Lei 12.527, de 18 de novembro de 2011, comumente conhecida como Lei de Acesso à Informação (LAI), foi promulgada diretamente para regulamentar o acesso à informação no Brasil (BRASIL, 2011). Para garantir o acesso informação, cabe ressaltar, é fundamental que os documentos estejam organizados e representados de forma que possam ser utilizados. No entanto, para que os documentos possam estar acessíveis aos usuários/pesquisadores, faz-se necessário o uso de instrumentos que contribuam, consecutivamente, na relação entre os usuários/pesquisadores e as instituições mantenedoras de arquivos.

Ademais, o processo acelerado dos recursos físicos tem gerado uma crescente produção do conhecimento arquivístico, no que diz respeito ao contexto das Tecnologias de Informação e Comunicação (TIC). Muito embora, o campo digital seja um ambiente complexo, a necessidade de representar tal conhecimento com instrumentos terminológicos é imprescindível para o acesso informacional.

Para tanto, os principais instrumentos de organização nos arquivos são o Plano de Classificação e a Tabela de Temporalidade. A partir dessas funções, o tesauro funcional, o mais apropriado para arquivos, é desenvolvido levando em conta as atividades que são desempenhadas pelos arquivos. Assim:

Um tesauro funcional controla o vocabulário que designa a razão de ser do documento (sua função) e não o que ele contém (assunto). A sua construção partirá do plano de classificação e o reordenará. Nunca é demais salientar que um tesauro funcional não substitui tal plano (SMIT; KOBASHI, 2003, p. 45).

O tesauro é um instrumento de recuperação da informação que oferece subsídios para o relacionamento entre os termos (hierarquia, associação e equivalência), uma forma mais detalhada da informação

orgânico-funcional. As relações associativas são aquelas dependentes dos conceitos, porém, não estabelecem uma super ordenação. Essas relações são divididas em dois grupos: relação sequencial, que é a relação extrínseca entre conceitos, que faz referência a objetos que têm relação de unidade espaço-temporal, como, por exemplo, “causa-efeito”, “produtor x produto”, “etapas de um processo”; e a relação pragmática, caracterizada como a relação entre conceitos que podem ser estabelecidos por conexões temáticas. Segundo Cintra *et al.* (2002, p. 54), as relações que não se submetem a uma hierarquia são aquelas que apresentam uma proximidade espacial ou temporal entre termos, que não possuem relacionamento hierárquico e nem de equivalência. Para o controle do vocabulário, as relações de equivalência são usadas em uma linguagem documentária, isso porque operam no nível de sinonímia e da polissemia, assim determinam as remissivas para conduzir o usuário para os termos preferidos pelo sistema.

Lembrando que, na linguagem natural, vários termos, palavras ou expressões podem designar uma única ideia ou conceito, porém, na Linguagem Documentária, apenas uma denominação é selecionada para designar como descritor. Assim, os outros termos são considerados como não descritores, tornando-se uma remissiva (GOMES, 1990, p. 47).

Quando se usa a sinonímia na construção de tesouros, verificamos que sinônimos são dois termos que tem a possibilidade funcional de serem trocados um pelo outro. Cintra *et al.* (2002, p. 77) destacam que pode “compreender tanto a sinonímia absoluta como a quase-sinonímia” e falam, ainda, que a sinonímia nas linguagens documentárias “é de caráter eminentemente preferencial e visa remeter o usuário de um termo não preferencial, para um termo selecionado [...]” (CINTRA *et al.*, 2002, p. 77).

Na Arquivologia, o tesouro funcional pode contribuir de várias formas, respeitando os princípios da área. É produzido na perspectiva de funcionamento de índice de termos do plano de classificação;

como índice para encontrar séries documentais a serem eliminadas; provê pontos de acesso ao usuário, permitindo o uso de sinônimos ou outros termos não adotados para a terminologia, mas adotado pelo arquivo; pode ser usado como ferramenta para ajudar o funcionário a encontrar documentos necessários para as tarefas do dia a dia (SMIT; KOBASHI, 2003, p. 45).

Considerando tais fatores, Smit e Kobashi (2003) destacam que os tesouros funcionais ofereceram subsídios à Arquivologia no que se refere ao tratamento da informação, ou seja, existe uma diferença entre uma informação arquivística e a biblioteconômica, a primeira é avaliada por sua função e, a segunda, por seu assunto. Cabe ressaltar que um quadro de classificação em Arquivo se refere a representação da estrutura orgânico/funcional do fundo ao qual se aplica, porém, é completamente diferente, da sua origem, dos sistemas de classificação de bibliotecas, cujas rubricas representam assuntos. Portanto, a estrutura de um quadro de classificação em Arquivos possibilita o acesso por proveniência, não especificamente recuperação.

Com isso, o tesouro tradicional não seria indicado para organizar arquivos, mas, é possível encontramos tesouros para arquivos, especificamente os históricos, que privilegiam o conteúdo temático. Podemos citar: o espanhol *Tesouro del Patrimonio Histórico Andaluz*, o inglês *United Kingdom Archival Thesaurus (UKAT)*, e os franceses *Thesaurus W: vocabulaire normalisé pour la description des archives administratives locales contemporaines* e o *Archives de Cannes*.

Os tesouros citados desenvolvem basicamente três listas: assuntos, instituições e localidades geográficas (GIMÉNEZ-CHORNET; ESCRIG-GIMÉNEZ, 2011). Porém, tais tesouros, não consideraram a orientação da norma internacional para gestão de arquivos (ISO 15.489), ao hierarquizarem assuntos e não funções (GIBBONS; SHENTON, 2003).

A importância de um tesauro aplicado a arquivos está relacionada ao fato de poder ser modelado a documentos e registros eletrônicos, tanto quanto aos documentos físicos. Estes são ligados à construção de estruturas de diretórios nomeadas com as palavras-chave do tesauro, permitindo que os documentos sejam definidos por metas aos objetivos da instituição, contribuindo na eficiência e na rapidez da recuperação da informação (NATIONAL ARCHIVES OF AUSTRALIA, 2003).

### 3.2 THESA – TESAURO SEMÂNTICO APLICADO

O Thesa é um *software* gratuito, que funciona em ambiente Web, e foi desenvolvido como um instrumento para os estudantes de graduação de biblioteconomia na disciplina de Linguagens Documentárias para a elaboração de tesauros, a fim de reduzir o trabalho operacional e priorizar o trabalho cognitivo e conceitual referente a modelagem do domínio (GABRIEL JUNIOR; LAIPELT, 2017). Na versão 0.20.05.18 do Thesa, é possível criar um número ilimitado de tesauros, ou seja, tem uma concepção de múltiplos tesauros, e os usuários desses tesauros podem configurá-lo para uso público ou privado (GABRIEL JUNIOR; LAIPELT, 2017). Além do mais, tem como base as normas internacionais de construção de tesauros da ISO e NISO vigentes, assim como, as diretrizes e os requisitos semânticos dos SOCs e da Web Semântica.

A escolha desse *software* se fez pela sua estrutura, que é baseada na concepção das relações entre os conceitos, ou seja, nele, o conceito pode ser representado por um termo, uma imagem, um som, um link ou qualquer outra forma que possa ser explicitada, como descreve seus produtores. Os organizadores destacam que o Thesa é um *open source* (código aberto), podendo ser modificado ou aperfeiçoado, desde que mantendo os créditos, e, ainda, aceita contribuições de melhoramentos pela comunidade.

## 4 O THESA APLICADO NO CONTEXTO DOS ARQUIVOS

Visto que os tesouros podem ser instrumentos eficazes no processo de organização do conhecimento arquivístico de uma instituição, nesse subcapítulo, explana-se sobre a possível utilização deste por meio do *software* Thesa. No entanto, sugerem-se pequenas adaptações para que esse *software* trabalhe plenamente com alguns conceitos e tipologias documentais requeridas pela documentação arquivística. Essas “sugestões” fazem-se necessárias, uma vez que o Thesa não foi pensado, inicialmente, para o ambiente do arquivo, mas, por ter certa flexibilidade, sua adaptação se torna possível para outras áreas do conhecimento (GABRIEL JUNIOR; LAIPELT, 2017).

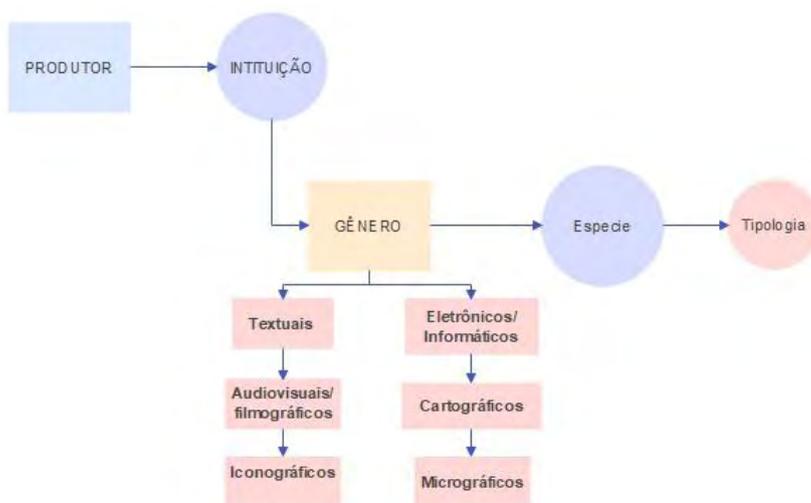
Dessa forma, reforça-se que, para que se elaborassem as sugestões em questão, utilizou-se a própria estrutura do Thesa e pensou-se, no âmbito desta, as modificações necessárias para que esse *software* comporte, de maneira mais eficaz, a organização de um tesouro voltado para arquivos. Ainda sobre isso, outro ponto a ser ressaltado é que, para intuito didático e meramente ilustrativo, foi utilizado uma série documental do arquivo central da Universidade Federal do Pará (UFPA)<sup>76</sup>, relativa aos contratos, acordos e convênios, visando a uma melhor compreensão da proposta e sugestões feitas nesse trabalho.

A série documental foi escolhida para essa proposta teórica, para além do atributo didático, por se entender que, para um *software* de relações semânticas como o Thesa, ao menos inicialmente, trabalhar a proposta baseando-se em uma série documental traria maior facilidade de compreensão do objetivo dessa pesquisa e se enquadraria mais nos campos e funcionalidades já existentes no *software*. Neste

<sup>76</sup> Ressalta-se que nenhum dado sigiloso ou delicado foi utilizado na demonstração. Por se tratar de documentação permanente, muitas das nomenclaturas utilizadas são comuns aos mais diversos arquivos, sendo usadas aqui como forma de ilustrar didaticamente a proposta.

sentido, um dos pontos iniciais que foi refletido sobre o *software*, para uma possível mudança, diz respeito principalmente ao ato da criação do Tesouro no Thesa. Após uma análise inicial, percebeu-se que, para melhor atender o âmbito do arquivo, alguns elementos voltados aos conceitos arquivísticos devem estar presentes no *software*. Dessa forma, segue-se uma figura para melhor ilustrar tal necessidade:

Figura 3 – Elementos conceituais arquivísticos



Fonte: Elaborado pelos autores (2021).

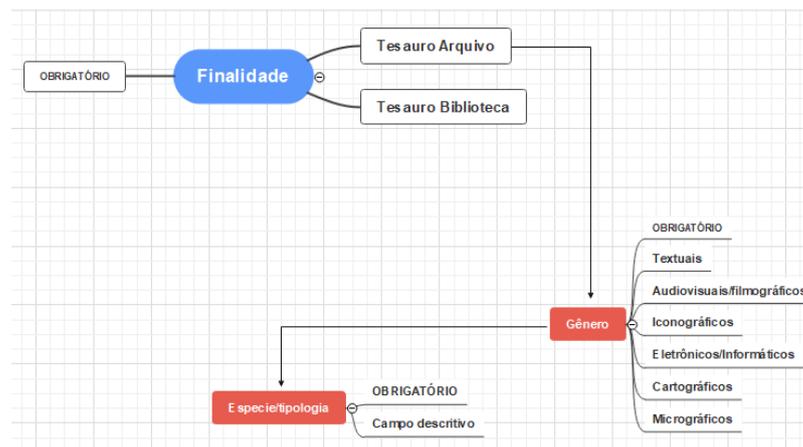
Nota-se que, para melhor organização do conhecimento arquivístico, é necessário que alguns elementos conceituais da área estejam presentes no processo inicial de criação do tesouro no Thesa. São eles: o produtor da série documental em que se embasará a criação do Tesouro; a descrição desta instituição produtora; informações como objetivos, propósitos e finalidade e breve histórico; a seleção do gênero documental do tesouro que se está organizando, delimitando-se, como proposta ilustrativa, aos gêneros textuais, audiovisuais/filmográficos,

iconográficos, eletrônicos/informáticos, cartográficos, micrográficos<sup>77</sup>; por fim, a descrição de espécie e tipologia documental.

Um aspecto a ser ressaltado é com relação à escolha do gênero. Os tipos de gêneros documentais apontados por este estudo são meramente ilustrativos, em um esforço de englobar todos os possíveis gêneros trabalhados em um arquivo. Cabe ao usuário selecionar o que melhor se enquadra com a documentação com a qual ele está trabalhando na criação do seu tesauro.

Dessarte, após análise, no que concerne ainda ao processo de criação de um tesauro voltado para arquivo no âmbito do Thesa, no que diz respeito às funcionalidades práticas e posições dos campos sugeridos para o *software*, identificou-se três mecanismos importantes a serem inseridos neste, baseado no esquema conceitual apontado anteriormente, o quais seriam: finalidade, gênero, espécie/tipologia. Para melhor compreensão da disposição desses dados no *software*, segue uma imagem ilustrativa:

Figura 4 – Modelo teórico de metadados para o THESA



Fonte: Elaborado pelos autores, 2021.

<sup>77</sup> Ressalta-se que, atualmente, seja provável que o Thesa não comporte todos os gêneros documentais citados. Uma saída para o criador do tesauro, caso isso ocorra, seria a criação de links que redirecionem o usuário para o arquivo armazenado em um serviço de nuvem, por exemplo.

Nota-se, de acordo com o modelo acima, que poderiam ser inseridos no *software* três campos obrigatórios no momento da criação do tesouro, seguindo a ordem mostrada na figura 4, em que a opção “finalidade” seria a primeira a ser selecionada, antes mesmo do campo já existente no Thesa, “nome do tesouro”. Sendo assim, o usuário escolheria qual o propósito da criação do seu tesouro, ou seja, sua finalidade, tendo duas opções: 1) **tesouro voltado para arquivo**; ou a opção convencional, 2) **voltado para a biblioteca**. Caso decida-se pela opção 2, que seria o tesouro convencional, os campos do Thesa permaneceriam os mesmos sem necessidade de mudanças. Porém, se o usuário seleciona-se a opção 1, relativa ao tesouro voltado para documentação arquivística, surgiriam mais dois campos obrigatórios como forma de metadados da opção 1: o campo gênero que, na estrutura de edição/criação do Thesa, encontrar-se-ia logo após o campo da **metodologia**, o qual estaria na forma de campo-seleção, em que o usuário selecionaria por meio de seu cursor a opção de gênero (opções essas citadas anteriormente) que mais se enquadre em sua documentação; após isso, outro campo que se encontraria logo depois do campo de gênero documental seria o de espécie/tipologia, o qual estaria no formato descritivo, ou seja, em caixa de texto, em que o usuário descreveria quais as espécies e tipologias da série documental para qual está elaborando seu tesouro. Em vista disso, observa-se, na figura 5, o aproveitamento dos campos existentes do Thesa para adaptá-los ao âmbito do arquivo:

Figura 5 – Modelo teórico de preenchimento dos campos Thesa para arquivo

Home Configurações Ferramentas **adicionar termo** buscar termos pesquisar

---

## Thesaurus

Nome do tesouro \*

Descrição do vocabulário / tesouro

### Descrição do vocabulário / tesouro

Introdução

Público alvo

Metodologia

Idioma preferencial \*

Fonte: página web do software Thesa<sup>78</sup>, 2021.

Dessa forma, viu-se a necessidade de reaproveitamento dos campos para sua utilização e sentido no contexto do arquivo. No campo **1**, referente ao nome do tesouro, colocar-se-ia o nome da série documental e do produtor da série em que se baseará a construção do

78 Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa>. Acesso em: 05/12/2021.

tesauro; em seguida, no campo **2**, alimentar-se-ia com as informações relativas à instituição produtora da série, como, propósito, breve histórico e rápida explanação sobre a série; no campo **3**, introdução, far-se-ia uma apresentação do propósito do tesauro e contextualização deste; no campo **4**, público alvo, e no campo **5**, metodologia da criação do tesauro, alimentar-se-ia com dados de acordo com a necessidade do usuário.

Assim, entende-se como necessário demonstrar breve relação semântica já existentes por meio das funcionalidades do Thesa que podem auxiliar na organização do conhecimento arquivístico, no âmbito desse *software*, de acordo com a figura 6:

Figura 6 – Funcionalidades do THESA

The screenshot shows the THESA web interface. At the top, there is a navigation bar with 'Home', 'Configurações', and 'Ferramentas'. A search bar contains 'adicionar termo' and 'buscar termos'. Below the navigation bar, the main content area displays the search results for 'Contratos imobiliários (por)'. The URL is 'https://www.ufpa.br/tesauros/index.php/thesa/c/28078'. The term is identified as 'TG: CONTRATOS (por)'. A definition box titled 'Definição do conceito' states: 'Contratos celebrados por partes para regular operações que envolvam bens imóveis'. To the right, there is a form titled 'CONTRATO PARTICULAR DE INCORPORAÇÃO IMOBILIÁRIA' with various fields for completion. At the bottom left, a small diagram shows 'Contratos imobiliários' and 'CONTRATOS' connected by a 'TG' relationship.

Fonte: página web do *software* Thesa.

Nota-se que o Thesa possui mecanismos que auxiliam na compreensão da relação semântica existente entre os termos do tesauro que se cria. Uma das relações semânticas que podem ser de grande utilidade no contexto arquivístico é a de Hiperonímia, que diz respeito

à relação hierárquica existente entre os termos, na imagem acima. Isso fica claro na relação de hiperonímia existente entre o termo “contratos imobiliários” e o termo “contratos”, em que o segundo seria o termo geral (TG) dessa série documental e o primeiro seria um termo específico (TE) da mesma série. Para além disso, existem as possibilidades das relações associativas, o que viabiliza ao usuário fazer maior gama de associações entre suas séries documentais, caso necessário. Por fim, percebe-se que outras duas funcionalidades do Thesa são demasiadamente úteis no contexto do arquivo: a definição do conceito relacionado ao termo que se alimenta no tesauro; e a possibilidade de inserir documentos iconográficos, como mostra a figura acima, na qual existe um exemplo do tipo documental representado pelo termo em questão.

## 5 CONSIDERAÇÕES FINAIS

Neste capítulo, buscou-se fazer breve explanação sobre os conceitos de sistemas de organização do conhecimento (SOCs) e apontou-se os tesouros, como parte integrante desses sistemas, tendo, com isso, potencial para auxiliar na organização do conhecimento arquivístico. Dentro desse escopo, identificou-se o Thesa, *software* voltado à criação de tesouros, como potencial instrumento auxiliador na criação de tesouros direcionado à organização do conhecimento arquivístico.

Desse modo, apontou-se as funcionalidades do Thesa e elaborou-se uma proposta teórica de adição de campos e metadados nesse *software* para comportar, de maneira mais eficaz, a documentação arquivística. Para tal, descreveu-se também suas funcionalidades atuais e como estas poderiam contribuir sem necessidade de uma alteração de grande proporção para a organização da documentação arquivística.

Por fim, ressalta-se que esta proposta teórica é extremamente delicada, visto que o Thesa foi desenvolvido inicialmente pensando

no contexto da Biblioteconomia, tornando essa proposta um desafio de adaptação do *software* para o âmbito da Arquivologia. Entretanto, espera-se que essa investigação possa servir de base para possíveis pesquisas relacionadas a essa temática, e que as propostas teóricas aqui mencionadas possam servir de base para o desenvolvimento de estudos mais profundos – com relação a semântica e a linguagem de programação – relativos à eficácia e funcionalidade do Thesa como *software* voltado para o conhecimento arquivístico.

## 6 REFERÊNCIAS

ALENCAR, Maíra Fernandes, e CERVANTES, Brígida Maria Nogueira. **Em Questão, Organização e representação do conhecimento arquivístico:** e busca de um método para construção de um tesauro funcional. Porto Alegre, v. 25, n. 1, p. 368-389, jan./abr. 2019. Disponível em: <http://dx.doi.org/10.19132/1808-5245251.368-389>. Acesso em: 05/12/2021.

BARITÉ, Mário. **Organización del Conocimiento:** un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA K. (Org.). Educação, Universidade e pesquisa. Marília: Unesp Marília, 2001. p. 35-60.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Brasília, DF: CGU, 2011.

CINTRA, A. M. M. *et al.* **Para entender as linguagens documentárias.** 2. ed. rev. e ampl. São Paulo: Polis, 2002. 92.

COOK, Terry. 1997. What is past is prologue: a history of archival ideas since 1898, and the Future Paradigm Shift. **Archivaria** 43, Feb. 1997. p. 19-20. Disponível em: <https://archivaria.ca/index.php/archivaria/article/view/12175>. Acesso em: 05/12/2021.

DAHLBERG, I. **Knowledge organization:** a new science? Knowledge Organization, Frankfurt, v. 33, n. 1, p. 11-19, 2006.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. Thesa: ferramenta para construção de tesauro semântico aplicado interoperável. **Revista P2P e INOVAÇÃO**, v. 3, n. 2, p. 124-145, 2017. Disponível em: 10.21721/p2p.2017v3n2.p124-145. Acesso em: 10 dez. 2021.

GIBBONS, Paul; SHENTON, Caroline 2003 Implementing a records management strategy for the UK Parliament: the experience of using Keyword AAA. *Journal of the Society of Archivists*. Acesso em 23/05/2022. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/0037981032000127016>.

GIMÉNEZ-CHORNET, Vincent; ESCRIG-GIMÉNEZ, Mercedes 2011 Designing a thesaurus to give visibility to the historical archives in Archivo del Reino in Valencia. *Knowledge Organization*. Acesso em 23/05/2022. Disponível em: <https://www.isko.org/ko382toc.pdf>.

GOMES, Hagar Espanha. **Manual de elaboração de tesouros monolíngües**. Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

GUIMARÃES, J. A. C. **A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO)**. *Revista Ibero-americana de Ciência da Informação*, v. 1, n. 1, p. 77-99, 2008.

HJØRLAND, B. What is Knowledge Organization (KO)? *Knowledge Organization*, v. 35, n. 2-3, p. 86-101, 2008.

HODGE, G. **Systems of knowledge organization for digital libraries: beyond traditional authority files**. Washington, D.C.: The Digital Library Federation Council on Library Information Resources, 2000.

MARCONI, Maria de Andrade; LAKATOS, Eva Maria. **Fundamentos da metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

MINAYO, Maria Cecília de Sousa *et al.* (Org.). **Pesquisa social: teoria, método e criatividade**. 21. ed. Petrópolis: Vozes, 2002.

NATIONAL ARCHIVES OF AUSTRALIA. **Developing a Functions Thesaurus: Guidelines for Commonwealth Agencies**, Canberra: National Archives of Australia, 2003.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Editora Feevale, 2013.

DE LIMA, Gercina Ângela. **Organização e representação do conhecimento e da informação na web: teorias e técnicas**. *Perspectivas em ciência da informação*, p. 57-97, 2020. Disponível em <https://periodicos.ufmg.br/index.php/pci/article/view/22283/17900>. Acesso em: 05/12/2021.

SMIT, Johanna W.; KOBASHI, Nair Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Imprensa Oficial do Estado de São Paulo, 2003.

SMIT, J. W. A Informação na ciência da informação. InCID: **Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 3, n. 2, p. 84-101, jul./dez. 2012.

SILVA, Amanda Marissa Soares da. **A Arquivologia na organização do conhecimento: uma análise de domínio**. 2021. 100 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-graduação em Ciências da Informação, Universidade Federal Fluminense. Niterói: Universidade Federal Fluminense, 2021.

TOGNOLI, Natália Bolfarini. A representação na Arquivística contemporânea. *Revista Ibero-Americana de Ciência da Informação*, v. 5, n. 2, p. 79-92, 2012.

ZENG, Marcia Lei. Knowledge organization systems (KOS). **KO Knowledge Organization**, v. 35, n. 2-3, p. 160-182, 2008.

# 14

Lucas Andrade Sá Corrêa

Natália Bolfarini Tognoli

A justiça social  
na organização  
do conhecimento

*Social justice  
in knowledge organization*

### Resumo:

Este capítulo tem como objetivo discutir como o tema da justiça social vem sendo abordado no domínio da Organização do Conhecimento (OC). Para tanto, adotamos uma abordagem de pesquisa qualitativa e exploratória, que usa a pesquisa bibliográfica para a coleta de dados associada a uma análise interpretativa de textos recuperados nas bases *Web of Science* e *EBSCO*, com vistas a identificar os autores que discutem o tema, os subtemas ligados à justiça social e à OC, os conceitos que os baseiam e as teorias que subsidiam essas discussões. Os resultados demonstram que as discussões acerca da justiça social na OC estão inseridas no contexto das teorias críticas, com forte influência das obras de Berman e Olson. Ao final, recomenda-se uma aproximação às contribuições da filósofa Nancy Fraser, para quem a justiça social é vista como paridade de participação, amparada na distribuição e no reconhecimento, fornecendo elementos para um aprofundamento no debate em direção a uma Organização do Conhecimento crítica como subsídio à justiça social em arquivos e bibliotecas.

**Palavras-Chave:** Justiça social; organização do conhecimento crítica; sistemas de organização do conhecimento hegemônicos.

**Abstract:**

*This chapter aims to discuss how social justice has been addressed in the Knowledge Organization (KO) domain. For this purpose, we adopted a qualitative and exploratory research approach using bibliographic research for data collection associated with an interpretative analysis of papers from Web of Science and EBSCO to identify the authors who discuss the subject, the subtopics linked to social justice and KO, the concepts that underlie them, and the theories that support these discussions. The results show that the discussions about social justice in OC are rooted in the context of critical theories, with strong influence from Berman's and Olson's works. In the end, we recommend an approach to the contributions of the philosopher Nancy Fraser, for whom social justice is defined as parity of participation, supported by distribution and recognition, providing elements for a deeper debate towards a critical Knowledge Organization as a support for social justice in archives and libraries.*

**Keywords:** *Social Justice; critical knowledge organization; hegemonic knowledge organization systems.*

## 1 INTRODUÇÃO

Nas últimas décadas, a Organização do Conhecimento (OC), especificamente em sua dimensão social e política, tem discutido a universalidade dos processos de organização do conhecimento, emergindo um conjunto de teorias, práticas e movimentos profissionais que promovem debates críticos à lógica hegemônica dos sistemas de organização do conhecimento (SOCs). Esses debates incluem a defesa da não neutralidade e a crítica à universalidade presentes nos principais sistemas de organização do conhecimento, encontrando um marco teórico na obra de Berman (1971), *Prejudices and Antipathies*, e ecoando nos estudos que se seguiram sobre os problemas, preconceitos e desvios no tratamento temático e descritivo nas bibliotecas, com especial destaque para a obra de Hope Olson.

O tema, embora em debate há pelo menos 40 anos, está longe de se esgotar e, com o passar do tempo, a confirmação de que os processos de organização precisam levar em conta a desigualdade econômica e informacional, o multiculturalismo e a diversidade da sociedade – para além do conteúdo dos documentos e de seu produtor – se faz cada vez mais presente nas discussões da área. Especificamente, no que toca aos produtos e instrumentos de OC, seja nos arquivos, seja nas bibliotecas, os desafios recaem sobre o reconhecimento dessas estruturas enquanto construções sociais que, como tal, são moldadas por um contexto de espaço e tempo que acaba privilegiando e destacando alguns aspectos, conceitos, pontos de vista e experiências, enquanto marginaliza ou desloca seu olhar de outros (MAI, 2013).

A importância do reconhecimento dessas estruturas, enquanto construções em um contexto de espaço e tempo, encontra fulcro no olhar sociocognitivo e cultural, destacado, também, a partir da perspectiva da análise de domínio preconizada na OC por Hjørland e Albrechtsen (1995), as discussões sobre as comunidades discursivas

envolvidas no domínio. Nesse contexto, considerando a análise de domínio a partir de um viés sociológico-epistemológico, o foco deslocar-se dos paradigmas físico e cognitivo – com ênfase no objeto e no indivíduo – para o paradigma social, enxergando o sujeito contextualizado em sua relação com o objeto-informação (SMIT, 2012; HJØRLAND, 2002). Assim, é necessário compreender que cada domínio possui necessidades específicas e que a universalidade deve ser repensada na construção dos sistemas de organização do conhecimento.

Frequentemente, ao debate da não neutralidade e da crítica à universalidade nos sistemas de organização do conhecimento, é incorporada a discussão sobre a justiça social. Esta é compreendida como um imperativo ético que se deve buscar nos processos de organização e representação do conhecimento.

Face ao exposto, o presente trabalho pretende iniciar as reflexões sobre como o tema da justiça social vem sendo abordado no domínio da Organização do Conhecimento. Para tanto, realizamos uma revisão de literatura com vistas a identificar os autores que discutem o tema, os subtemas ligados à justiça social e à OC, os conceitos que os baseiam e as teorias que subsidiam essas discussões no âmbito da OC. Cumpre destacar que o trabalho é fruto das discussões e pesquisas desenvolvidas no âmbito do Grupo de Pesquisa do CNPQ/UFF, “Arquivos, Bibliotecas e Organização do Conhecimento”, que vem abordando o tema da justiça social no contexto da organização do conhecimento na Arquivologia e Biblioteconomia.

Além dessa Introdução, o presente capítulo divide-se em quatro seções. Na segunda seção, trazemos a metodologia do trabalho, elencando os artigos recuperados e as categorias para análise dos textos. Na terceira seção, apresentamos dois autores centrais para a discussão que se pretende trazer nesse trabalho: Sanford Berman, autor de *Prejudices and Antipathies* (1971), e Hope Olson, autora de *Power to Name* (2002). A introdução desses autores e de seus respectivos trabalhos é

necessária por eles ocuparem posição central entre as referências dos artigos analisados. Na quarta seção, debatemos as ideias centrais dos textos recuperados, relacionando-os e buscando compreender suas propostas e questões que incidam sobre a justiça social e sua relação com a OC. Por fim, apresentamos as considerações finais do trabalho.

## 2 METODOLOGIA

O presente trabalho teórico adota uma abordagem de pesquisa qualitativa e exploratória, que usa a pesquisa bibliográfica para a coleta de dados associada a uma análise interpretativa, utilizando-se da técnica do *close reading* (MCCLENNEN, 2001). O objetivo da aplicação dessa técnica é a observação dos fatos e detalhes de um texto ao “lê-lo mais de perto”.

Assim, para o cumprimento dos objetivos do trabalho, realizamos buscas nas bases de dados *Web of Science* e *EBSCO* pelos termos combinados “*knowledge organization*” AND “*social justice*”, em texto completo. A busca recuperou 73 resultados e, após análise prévia com base nos critérios preestabelecidos, foram selecionados 21 artigos para leitura e análise.

Em um primeiro momento, foram observadas “[...] todas as características marcantes do texto, incluindo os recursos retóricos, os elementos estruturais, as referências culturais” (KAIN, 1998). Em um segundo momento, essas observações foram interpretadas utilizando o raciocínio indutivo com o auxílio da Análise de Conteúdo de Bardin (2009), isto é, passamos da observação de fatos particulares e de detalhes para uma conclusão ou interpretação baseada nas categorias de análise preestabelecidas.

### 3 ORGANIZAÇÃO DO CONHECIMENTO CRÍTICA: AS CONTRIBUIÇÕES DE SANFORD BERMAN E HOPE OLSON

As origens de uma prática em organização do conhecimento voltada para a diferença social e preocupada com questões de justiça são encontradas nos trabalhos de Sanford Berman e Hope Olson (LEAZER; MONTOYA, 2020). Nesses trabalhos, os autores apresentam uma crítica aos Sistemas de Organização do Conhecimento (SOCs) hegemônicos a partir da análise da representação (ou da ausência desta) dos grupos marginalizados.

Berman, considerado o catalogador radical mais conhecido entre os bibliotecários (WATSON, 2020), observou, desde o início de sua trajetória profissional, influenciado fortemente pela cultura da Guerra Fria, a centralidade das questões que perpassam problemas da justiça social, assim como a necessidade de uma biblioteconomia engajada e comprometida. Em sua obra, *Prejudices and Antipathies* (1971), o autor apresenta uma lista de 225 propostas de mudanças, inclusões e exclusões à Lista de Cabeçalhos de Assunto da Biblioteca do Congresso dos Estados Unidos da América (*Library of Congress Subject Headings*), doravante LCSH, publicado e utilizado como referência desde pelo menos 1909, exercendo uma influência hegemônica, sendo usado integralmente ou de forma adaptada em diversos países. Para o autor, apesar da pretensão de universalidade do LCSH, apenas um pequeno grupo sentir-se-ia representado pela sua lista: homens europeus e norte-americanos nacionalistas e brancos, cristãos – especialmente protestantes – possuidores de rendas média e alta, imbuídos da glória da civilização ocidental (BERMAN, 1971).

Os princípios que guiaram ou justificaram a reprodução do pensamento dessa elite no LCSH foram apresentados por David J. Haykin,

chefe da divisão de Catalogação de Assunto da Biblioteca do Congresso, em seu livro, *Subject Headings: A Practical Guide*. Segundo ele, “o leitor é o foco de todo princípio e prática de catalogação”, dessa forma a terminologia é determinada por uma aproximação à provável psicologia da maioria dos leitores (HAYKIN, 1951 apud MARSHALL, 1977).

Berman cita, ainda, a resposta de A.C. Foskett às críticas com relação ao viés racista ou colonialista da LCSH. Segundo ele, esta teria sido “projetada para o uso em bibliotecas ocidentais” e, portanto, refletiriam “o viés (bias) histórico dessas bibliotecas e de seus usuários”. A crítica de Berman a essa resposta é importante e será considerada no conjunto da nossa análise. Para o autor, uma vez reconhecida a parcialidade, o caráter equivocado e preconceituoso de alguns termos, estes devem ser substituídos, pois não podem ser vistos como censuráveis apenas pelas tradições não ocidentais, mas em qualquer lugar (BERMAN, 1971).

Para Berman, a hegemonia do LCSH não deveria ser ignorada, pois se é utópico pensar que este poderia se tornar um dia um instrumento desinteressado, seria igualmente utópico acreditar que ele seria facilmente substituível por uma alternativa ainda não existente (BERMAN, 1971). Ainda, para ele, mais do que escolher entre o ponto de vista local ou universal, o bibliotecário ou o catalogador deve responder à pergunta: “Qual lado você está?” (BERMAN, 1971, p. 16). Em seu trabalho crítico, orientado “para ampliar, humanizar e simplificar o acesso aos recursos e informações das bibliotecas” (BERMAN; GROSS, 2017), o autor aconselha os colegas catalogadores a serem criativos, abandonando a conformidade, exigindo um papel na tomada de decisões, priorizando a imaginação e a integridade profissional.

Além da sua importância como um dos precursores de uma Biblioteconomia e de uma Organização do Conhecimento orientada para a justiça social, o trabalho de Berman foi eficiente também do ponto de vista aplicado. Das 225 sugestões de mudanças apresentadas no livro,

abrangendo questões como raça, nacionalidades, fé, terceiro mundo, política, classe, trabalho, sexo, mulheres e homens, crianças, entre outros, 88 – isto é, 39% – foram mudadas quase exatamente como sugeridos por Berman (KNOWTON, 2005).

Outro importante marco teórico nos estudos de uma Organização do Conhecimento crítica, sem o qual muitas instâncias de opressão permaneceriam, até hoje, despercebidas (FOX, 2016), é a autora Hope Olson, professora emérita na *School of Information Studies* da Universidade de Wisconsin–Milwaukee. A sua obra influenciou sobremaneira os trabalhos encontrados em nossa busca, representando inovações não apenas nas críticas feitas aos SOCs tradicionais, mas, também, no instrumental teórico e epistemológico empregado por ela ao trazer as teorias pós-estruturalistas e feministas para o debate.

Em 2002, Olson publica *Power to Name: Locating the Limits of Subject Representation in Libraries*, a sua obra mais celebre e a mais citada entre os artigos estudados nesse trabalho. O livro traz uma análise crítica contextualizada de dois autores – Melvil Dewey (1851-1931) e Charles Cutter (1837-1903) – e de dois instrumentos de referência hegemônicos na Biblioteconomia e Ciência da Informação: a Classificação Decimal de Dewey (CDD) e o LCSH.

Em sua obra, Olson (2002) declara ter como objetivo alcançar tanto os interessados em organização do conhecimento, ou seja, bibliotecários, cientistas da informação, *web developers* e gestores de conhecimento, quanto os interessados nos estudos teóricos sobre a representação, incluindo neste grupo as feministas e os teóricos pós-estruturalistas e pós-coloniais interessados nas “construções de significado”. Os pressupostos teóricos de Olson estão ancorados nas teorias pós-estruturalistas e feministas, ocupando lugar central na arquitetura lógica e metodológica de sua obra, a feminista Drucilla Cornell e sua “filosofia do limite”, assim como a sua releitura do método de desconstrução de Derrida, que dá centralidade ao conceito de iterabilidade e iteração.

A iterabilidade é uma palavra formada pelo latim “iter”, que significa “de novo” e o sânscrito “itara”, “outro”. Como conceito, estaria ligado à característica de se repetir, como essência de uma “existência absoluta”. Em relação com essa característica, a iteração seria uma técnica de desconstrução da iterabilidade, negando o caráter universal pela sua rearticulação em contextos diferentes.

Olson adota a iteração para criticar a noção de “público” implícita nas regras de Cutter. A autora considera que o “público”, ao qual se destina o catálogo de Cutter, é um “público singular”, isto é, uma parte da humanidade, separada artificialmente e homogeneizada (OLSON, 2002). Nesse contexto, uma vez identificado o procedimento que cria um “público singular” e exclui o “outro” de sua lógica, Olson propõe um caminho para o enfrentamento dessa realidade, recuperando uma imagem proposta por Audre Lorde, a qual avaliamos ser central não apenas para as contribuições trazidas por Olson, mas para o conjunto de problemas enfrentado por quem busca uma Organização do Conhecimento crítica e orientada para a justiça social. Para Lorde (2019, p. 137),

“[...] as ferramentas do senhor nunca derrubarão a casa-grande. Elas podem possibilitar que os vençamos em seu próprio jogo durante certo tempo, mas nunca permitirão que provoquemos uma mudança autêntica”.

A resposta de Olson a esse desafio reside em subverter os SOCs existentes, provocando reformas estruturais na casa-grande, fazendo dela um espaço permeável às vozes dos excluídos (OLSON, 2002) ao incorporar uma perspectiva mundana, cotidiana. Entretanto, a autora parte de uma crítica a qualquer pretensão de universalidade, o que recoloca o desafio de Lorde sobre a casa e os instrumentos do senhor.

Ao defender que “soluções universais não são viáveis” e que sacrificam aspectos significativos das culturas locais, Olson propõe a

combinação de ações globais com ações locais. Se seria possível melhorar instrumentos como o LCSH e a CDD e torná-los mais permeáveis aos grupos marginalizados, a autora conclama os bibliotecários e catalogadores locais a assumirem a responsabilidade para o atendimento dos seus usuários (OLSON, 2002). Dessa forma, se a compreensão das deficiências do sistema hegemônico nos possibilita desenvolver seus potenciais (OLSON, 2002), a crítica ao universalismo impõe limites às propostas globais que abre ao poder de nomear, desde o “local”, como a estratégia possível para as transformações.

Como veremos, os trabalhos de Berman e Olson podem ser considerados marcos de uma importante agenda de pesquisa em Organização do Conhecimento que busca revelar os processos, escolhas, projetos, interesses e disputas. Nestes, escondem-se os termos e conceitos presentes nos sistemas de organização do conhecimento globais, os quais frequentemente levam à perda dos significados e práticas locais a favor de uma universalização e padronização.

Uma breve análise de citação nos textos analisados nos permitiu identificar os autores Olson e Berman como os mais citados, sendo que a primeira é mencionada em 18 dos 21 artigos analisados, enquanto o segundo em 11 deles. Ao compararmos esses dados com aqueles levantados por Castanha e Wolfran (2018) em um artigo que identificou os autores mais citados no campo da OC, é possível percebermos um contraste relevante com relação aos estudos sobre justiça social. No referido artigo, os autores identificaram Hjørland, Mai e Tennis como os três autores mais citados na OC, com Olson ocupando a oitava posição e sem menção a Berman. Em contraposição, no pequeno universo analisado aqui, Olson e Berman ocupam posição central.

## 4 A JUSTIÇA SOCIAL NA ORGANIZAÇÃO DO CONHECIMENTO

Para a análise dos artigos selecionados neste trabalho, utilizamos o método da análise de conteúdo de Bardin (2009). Este consiste em um conjunto de técnicas de análise das comunicações executada em três pólos: pré-análise (constituição do corpus e elaboração das categorias), exploração do material e tratamento dos resultados (inferências e interpretações).

A fase da pré-análise consistiu na escolha do campo empírico – as bases de dados internacionais – e nas buscas pelos artigos nos títulos, resumos e palavras-chaves. A fase de formulação das hipóteses e objetivos foi realizada *a posteriori*, em consonância com Bardin (1977), que clarifica não ser obrigatória a existência de um *corpus* de hipótese para a realização da análise. Segundo Bardin (1977), para a execução da segunda fase, é necessário proceder-se à constituição de um *corpus*, reunindo os documentos a serem submetidos à análise. As categorias que guiaram a leitura e interpretação dos textos selecionados foram: 1) natureza do estudo; 2) teoria/epistemologia abordada; 3) processo de organização do conhecimento; 4) sistema de organização do conhecimento; 5) conceito de justiça social.

Essa pesquisa é composta por 21 artigos internacionais que abordam a justiça social em conjunto com a Organização do Conhecimento, recuperados a partir das buscas pelos termos combinados “*social justice*” AND “*knowledge organization*”, nas bases *EBSCO* e *Web of Science*, conforme demonstrado no Quadro 1.

Quadro 1 – Artigos analisados

Título	Autor	Periódico e ano de publicação
Indigenization of knowledge organization at the Xwi7xwa library	DOYLE, A. M., LAWSON, K., & Dupont, S	Journal of Library & Information Studies, 2015
How Subjects Matter: The Kinsey Institute's <i>Sexual Nomenclature: A Thesaurus</i> (1976)	DRUCKER, D. J.	Information & Culture, 2017
Contemplating Co-creator Rights in Archival Description	GILLILAND, A. J.	Knowledge Organization, 2012
"Priorities of Arrangement" or a "Hierarchy of Oppressions?": Perspectives on Intersectionality in Knowledge Organization	FOX, M. J.	Knowledge Organization, 2016
Race and Ethnicity in Classification Systems: Teaching Knowledge Organization from a Social Justice Perspective	ADLER M.; HARPER, L. M.	Library Trends, 2018
Shifting sands and the prophet's dream: Exploring the future of information organization education	SNOW, K.	Journal of education for Library & Information Science, 2019
The case for taxonomic reparations	ADLER, M.	Knowledge Organization, 2016
Methodological aspects of critical theories in knowledge organization.	MARTINEZ-ÁVILA, D.; SEMIDÃO, R; FERREIRA, M.	Knowledge Organization, 2016
The Politics of Knowledge Organization: Introduction to the Special Issue	LEAZER, G. H.; MONTOYA, R.	Knowledge Organization, 2020
Knowledge Organization as Knowledge Creation: Surfacing Community Participation in Archival Arrangement and Description	BAK, G.; ALLARD, D.; FERRIS, S.	Knowledge Organization, 2019
Dewey Deracialized: A Critical Race-Theoretic Perspective	FURNER, J.	Knowledge Organization, 2007
Declassifying Knowledge Organization	GARCIA GUTIERREZ, A.	Knowledge Organization, 2014
Classifying for Social Diversity	SZOSTAK, R.	Knowledge Organization, 2014

Just KOS! Enriching Digital Collections with Hypertexts to Enhance Accessibility of NonWestern Knowledge Materials in Libraries	THARANI, K.	Knowledge Organization, 2020
Neoliberalism and Library & Information Science: Using Karl Polanyi's Fictitious Commodity as an Alternative to Neoliberal Conceptions of Information	COPE, J.	Progressive Librarian, 2014-2015
'There was Sex but no Sexuality*': Critical Cataloging and the Classification of Asexuality in LCSH	WATSON, B.W.	Cataloging and Classification Quarterly, 2020
Feminism and the Future of Library Discovery	SADLER, B.; BOURG, C.	Code-4Lib Journal, 2015
Transcending Library Catalogs: A Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in LibraryThing for Transgender Books	ADLER, M.	Journal of web librarianship, 2009
Transformative praxis – building spaces for Indigenous self-determination in libraries and archives	THORPE, K.	In the library with the lead Pipe, 2019
Transforming Subject Access: Some Critical Issues for South African Information Professionals	DICK, A.L.; BURGER, M.	South African Journal of Library & IC, 1995
A Library Matter of Genocide: The Library of Congress and the Historiography of the Native American Holocaust	DUDLEY; M.Q.	The International Indigenous Policy Journal, 2017

Fonte: Elaborado pelos autores, 2022.

Dos 21 trabalhos selecionados, 10 foram escritos por autores vinculados a universidades dos Estados Unidos, 6 do Canadá, 1 da Espanha, 1 da Alemanha e 1 do Brasil. Os artigos foram publicados em 12 periódicos diferentes, sendo que apenas um deles se repete, o *Knowledge Organization* (KO), no qual estão publicados 10 dos 21 artigos, o que demonstra a grande centralidade desse periódico nos debates sobre o tema. A única autora que se repete no universo estudado é Melissa Adler, com 3 artigos.

No tocante à natureza do trabalho, dos 21 artigos analisados, 7 são de natureza aplicada (DOYLE, 2015; DRUCKER, 2017; BAK, ALLARD e FERRIS, 2019; FURNER, 2007; THARANI, 2020; WATSON, 2020; ADLER, 2009). Isso demonstra uma preocupação ainda mais teórica no âmbito da OC quando tratamos da justiça social.

Sobre as teorias/epistemologias abordadas, 11 trabalhos (DRUKER, 2017; BAK *et al.*, 2019; WATSON, 2020; ADLER, 2009; 2016; ADLER e HARPER, 2018; FOX, 2016; MARTINÉZ-ÁVILA *et al.*, 2016; GARCIA GUTIÉRREZ, 2014; SZOSTAK, 2014; SADLER E BOURG, 2015) trazem discussões que perpassam as questões de gênero e sexualidade. Destes, 6 (DRUKER, 2017; FOX, 2016; ADLER, 2019; 2016; WATSON, 2020; SADLER E BOURG, 2015) tratam especificamente dessas questões, tendo, nas epistemologias feministas e na Teoria Queer, seus principais fundamentos teóricos.

Em evidente continuidade com o trabalho de Berman e Olson, o artigo de Melissa Adler, *Transcending Library Catalogs: A Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in LibraryThing for Transgender Books* (2009), faz uma investigação dos termos referentes a temas relacionados a transgêneros, comparando o LCSH e as *tags* criadas pelos usuários do site *Librarything*. A autora inicia citando a proposta de Berman (2008) de inclusão dos termos “*Drag queens*” e “*Genderqueers*” aos LCSH. Apesar dos seus esforços, até a publicação do artigo, os termos não haviam sido incluídos. Em contraste com esse fato, a autora chama a atenção para a existência desses termos na experiência de folksonomia das *tags* da *LibraryThing*, em que os usuários catalogam seus próprios livros. Adler propõe um estudo de caso comparativo entre os usos do cabeçalho na *Worldcat* – onde esse termo também existe – e no *Librarything*, onde existem e são atribuídos pelos próprios usuários. Essa comparação é a base principal para a defesa de Adler da

importância do poder de uma comunidade de usuários de nomear seus próprios “recursos” (ADLER, 2009).

A discussão sobre a inclusão de termos no LCSH e a sua relação com grupos marginalizados foi estudada também por Watson (2020) ao discutir a disputa política em torno da tentativa de substituição do termo imigrante ilegal (*illegal alien*) no LCSH. Após um amplo movimento de diversos setores da sociedade e dos catalogadores radicais, o termo estava para ser mudado, quando foi impedido pela bancada republicana. O caso gerou o questionamento da deputada democrata Debbie Wasserman Schultz: Se a Biblioteca do Congresso mudava centenas de termos por ano sem a interferência do Congresso, “por que este? Por que agora?” (WATSON, 2016). Watson defende a importância dos catalogadores críticos (*Critical Catalogers*), nas leituras dos catálogos e sistemas de classificação orientada para a justiça social, expondo e desafiando “as formas como as bibliotecas e a profissão apoiam consciente e inconscientemente os sistemas de opressão” (WATSON, 2016).

A discussão sobre a organização e representação do conhecimento no contexto dos povos indígenas está presente em 10 artigos (DOYLE, 2015; GILLILAND, 2012; ADLER, 2016; 2018; MARTINÉZ-ÁVILA *et al.*, 2016; LEAZER E MONTOYA, 2020; BAK *et al.*, 2019; THARANI, 2020, THORPE, 2019; DUDLEY, 2017). Como demonstra Dudley (2017) a omissão e o eufemismo que marcam o holocausto indígena – no caso estudado, da América do Norte, mas que, certamente, poderia ser generalizado por outros territórios – apontam para o fato de que o imperativo de conceber e instituir assuntos e classificações para o acesso não é uma questão de historiografia, mas de justiça social (DUDLEY, 2017, p. 24).

Os “fracassos da teoria e da prática da biblioteca e do arquivo em apoiar as prioridades indígenas” são o alvo da denúncia do arquivista indígena australiano, Kirsten Thorpe (THORPE, 2019). Thorpe parte das suas próprias experiências profissionais para demonstrar o sentimento de “insegurança cultural” (*culturally unsafe*) provocada pelos arquivos e

bibliotecas aos usuários e profissionais indígenas. O autor propõe, entre outras medidas, “uma agenda de pesquisa (...) em parceria com povos e comunidades indígenas”, reforçando os apelos por participação dos grupos excluídos presentes em grande parte da bibliografia abordada.

A forma como povos e nações podem ser excluídos ou sub-representados nos sistemas de organização do conhecimento hegemônicos pode ser verificado também pelo exemplo sul-africano, discutido por Dick e Burger (1995). Os autores demonstram como a busca por uma “neutralidade” pode omitir e reproduzir relações de poder na sociedade, colaborando para a exclusão e marginalização de alguns grupos. O exemplo da inclusão do termo *apartheid* no LCSH apenas em 1986, demonstra o quanto os SOCs podem prejudicar e enviesar o autoconhecimento de um povo.

O racismo, a questão racial e a Teoria Crítica da Raça são temas abordados em 6 artigos (FOX, 2016; ADLER, 2016; MARTINÉZ-ÁVILA *et al.*, 2016; LEAZER E MONTOYA, 2020; GARCIA GUTIÉRREZ, 2014; FURNER, 2007). Destes, em três, é o foco específico (ADLER E HARPER, 2018; ADLER, 2016; FURNER, 2007).

Furner (2007), ao perceber a necessidade do método ontogenético para a compreensão de como a questão racial vem sendo tratada na Classificação Decimal de Dewey, corrobora com Tennis (2002) sobre a necessidade de uma ontogenia dos assuntos. O autor demonstra como a CDD buscou superar as estruturas racistas do seu sistema por um processo de desracialização, embora a erradicação do racismo, não seja o mesmo que erradicar a termo “raça”. Ao propor a Teoria Crítica da Raça como fundamento teórico para uma classificação que não reproduza nem omita as estruturas racistas da sociedade, Furner defende que é preciso reconhecer a realidade da “raça” como construção social e o seu papel na manutenção de opressões.

Um movimento parecido é proposto por Fox (2016), partindo das teorias da interseccionalidade, para lidar com as inter-relações entre opressões e como estas não costumam estar representadas nos sistemas de classificação. Fox acompanha de perto a abordagem de Olson e enfatiza a importância das mudanças locais para a solução de problemas específicos, reduzindo a universalidade e aumentando o controle local. Fox recupera a metáfora de Audre Lorde e corrobora com a solução encontrada por Olson: se não há modelo para demolir a casa-grande, seriam possíveis ainda “sérias redecorações”. O caminho para isso, partindo do local, seria o de ampliação da participação dos grupos excluídos nos processos de criação e de tomada de decisão em temas relacionados a eles. Conceitos produzidos no âmbito dos debates da interseccionalidade, como o de solipsismo branco (*white solipsism*) – isto é, uma incapacidade epistemológica dos brancos de perceberem o outro – demonstram a necessidade de participação dos grupos excluídos na crítica e criação de alternativas aos sistemas de organização do conhecimento hegemônicos (FOX, 2016).

Para além destes, foram recuperados artigos cujos temas englobam a decolonialidade e a crítica aos SOCs centrados no ocidente e que trazem as teorias críticas para discutir as questões pedagógicas, epistemológicas e conceituais (MARTINÉZ-ÁVILA *et al.*, 2016; GILLILAND, 2012; SNOW, 2019; LEAZER E MONTOYA, 2020, BAK *et al.*, 2019; GARCIA GUTIERREZ, 2014; COPE, 2014), sem focar especificamente em comunidades específicas. Estão presentes, ainda, as discussões sobre a educação e o ensino em organização do conhecimento voltados para uma perspectiva crítica em todos os níveis de formação (FOX, 2016; ADLER, 2018; SNOW, 2019).

Adler, a partir de revisão bibliográfica e exemplos de aplicação em sala de aula, defende importância de discussões sobre diversidade, inclusão e justiça social. Para a autora, essas questões não são apenas importantes em si mesmas, mas também por permitirem “que

os alunos compreendam os aspectos epistemológicos e os fundamentos ontológicos sobre os quais os sistemas de organização do conhecimento se baseiam” (ADLER, 2018, p. 67). Snow (2019), ao discutir as mudanças no currículo nos cursos de “Organização da Informação” – a qual a autora aproxima da OC –, defende uma ênfase maior nos impactos éticos da Organização da Informação, além de uma articulação entre teoria e prática como um elemento importante no ensino para uma aproximação daqueles que estão atuando na linha de frente.

Além da crítica aos SOCs hegemônicos, parte dos artigos recuperados em nossa busca apresentam estudos de caso discutindo alternativas locais. Estas, de diferentes formas, pretendem orientar-se para a justiça social, resolvendo as lacunas e superando os preconceitos percebidos por especialistas e usuários.

Gilliland (2012) defende o conceito de co-produtor (co-creator) na Arquivologia, a partir do estudo do *Aboriginal and Torres Strait Islander Data Archive* (ATSIDA). Doyle, Lawson e Dupont (2015) apresentam um estudo de caso sobre a biblioteca indígena *Xwi7xwa Library*, propondo elementos para a indigenização da Organização do Conhecimento. Drucker (2017) analisa o tesouro do Instituto Kinsey, o *Sexual Nomenclature*, criado na década de 1970. Bak, Allard e Ferris (2019) examinam dois projetos arquivísticos baseados em experiências comunitárias, no Canadá: O *Project Naming* da *Library and Archives Canada*, que trabalha com a comunidade *Inuit* e o *The Sex Work Database* da Universidade de Manitoba, que trabalha com ativistas profissionais do sexo. Por fim, Tharani (2020) estuda o desenvolvimento de um Sistema de Organização de Conhecimento para garantir a acessibilidade dos *ginans*, hinos devocionais dos muçulmanos Ismailis.

No que tange aos processos de organização e representação do conhecimento abordados nos trabalhos, a classificação é mencionada na maioria dos artigos, seguido da catalogação. Sobre os

sistemas de organização do conhecimento, a CDD e o LCSH aparecem com mais frequência.

No tocante à última categoria de análise, o conceito de justiça social, observamos que embora grande parte dos artigos trate da justiça social a partir do recorte de um ou mais grupos marginalizados, notadamente nos SOCs, poucos enfrentam explicitamente a questão: o que é Justiça Social? Os trabalhos são unânimes em considerar que a justiça social está ligada à inclusão de grupos excluídos e marginalizados, além do consenso ao diagnosticar o fracasso dos SOCs hegemônicos em produzir justiça social.

Justiça Social é um termo em disputa. Diferentes interesses, abordagens e autores lhe darão diferentes contornos e definições. Como nos lembra Furner (2007), apenas a afirmação de estar orientado para a justiça social não diz muito, já que poucos afirmariam defender uma posição contrária – a injustiça. Dessa forma, mais que defender uma Organização do Conhecimento orientada para a justiça social, é necessário conceituá-la.

Os artigos analisados, em sua maioria, partem da necessidade de se “historicizar o assunto”, de criticar os termos estabelecidos, de mapear as classes e demonstrar a não neutralidade das categorias. Apesar disso, muitas vezes o conceito de justiça social aparece como dado, sem que ele mesmo seja historicizado, criticado, mapeado.

Podemos nos perguntar se a “centralidade do local”, a influência do método da desconstrução e a crítica às pretensões de universalidade e objetividade não poderiam ter contribuído para a não definição de um conceito de justiça social. Afinal, a conceituação de justiça social não poderia apontar para uma teleologia? Para o “ponto de chegada”, desacreditado por Olson?

Ao analisarmos os únicos dois trabalhos (FURNER, 2007; LEAZER E MONTOYA, 2020) recuperados em nossa busca que discutem

e propõem explicitamente um conceito de justiça social, pretendemos demonstrar que não há oposição entre a crítica e a sistematização. Uma dialética entre esses dois polos pode não ser um trabalho fácil, mas é o desafio posto pelo problema que enfrentamos quando questionamos os SOCs hegemônicos e seu “universalismo” excludente.

Para Furner (2007) a justiça social é equivalente à justiça distributiva. Isto é, a justiça estaria relacionada à forma como uma sociedade distribui bens e serviços, não apenas materiais, mas também culturais. O autor apresenta duas formas de compreensão da justiça social (distributiva). A primeira, é a contratualista, fundamentada na clássica obra de John Haws, *Uma Teoria da Justiça* (1971). Nesta, Haws parte do individualismo metodológico para propor um modelo em que indivíduos optariam, por um cálculo racional, a entrar em um contrato que seria benéfico para todos (FURNER, 2007, p. 149). As teorias contratualistas, entretanto, poderiam, segundo o autor, servir como justificativa implícita para situações injustas surgidas não por interferência direta do Estado, mas pela relação entre indivíduos na esfera privada. Dessa forma, o monopólio de bens e serviços materiais e culturais permaneceria intacto.

Em oposição às teorias contratualistas, Furner apresenta as teorias comunitárias. Para estas, “uma distribuição justa é aquela que não viole direitos ou liberdades” de grupos ou comunidades, e seus membros, em especial aquelas que estão alijadas do poder e/ou são vítimas de opressão (FURNER, 2007, p. 150). Essa definição de justiça social, ainda segundo o autor, permitir-nos-ia avaliar sistemas de classificação como “geradores de distribuição de bens conceituais”. Dessa forma, entendendo cada classe como unidade a ser distribuída poderíamos nos perguntar, por exemplo, “Quão justo seria alocar, digamos, 70% das subclasses disponíveis da classe principal ‘Religião’ para assuntos relacionados ao cristianismo?” (FURNER, 2007, p. 155).

Leazer e Montoya (2020) partem do trabalho de Berman e Olson para apontar que as diferenças sociais criticadas por eles nos SOCs não podem ser reduzidas à questão de distribuição. Os autores adotam a definição da cientista política Iris Marion Young, para quem a “justiça social significa a eliminação de dominação e opressão institucionalizada” (YOUNG 1990, p.15). Para Young (1990), as teorias distributivas têm dois problemas principais: elas tendem a focar na distribuição de bens materiais, “ignorando a estrutura social e o contexto institucional que muitas vezes ajudam a determinar os padrões distributivos”. O segundo problema se dá justamente a partir da tentativa de incluir na teoria distributiva bens sociais imateriais, reduzindo relações e processos sociais a “coisas” ou mercadorias (YOUNG, 1990, p. 16).

A autora pretende mudar o foco dos padrões distributivos para processos de participação, em que todas as pessoas tenham direito e oportunidade de participar, deliberar e tomar decisões em todas as instituições da vida coletiva da qual fazem parte ou por qual são afetadas. Assim, para Young, “a democracia é elemento e condição para a justiça social” (YOUNG, 1990, p. 91).

A partir da definição de Young, Leazer e Montoya consideram que a Organização do Conhecimento é ferramenta e produto de hegemonia cultural e, ao reproduzir preconceitos e desigualdade sociais, os sistemas de organização do conhecimento também podem servir – quando objetos de análises críticas – para revelar o sistema de relações entre os conceitos que sustentam a dominação e opressão social. Ao invés de instrumentos neutros de representação, os autores propõem que os sistemas de organização do conhecimento possam ser lidos como fonte de evidência do desenvolvimento de conceitos social e politicamente opressores (LEAZER, MONTOYA, 2020).

Como vimos, poucos trabalhos discutiram o conceito de justiça social. Apesar disso, em sua crítica aos termos, às classes e à própria lógica dos SOCs que reproduzem a opressão e a dominação

institucionalizada, os autores apontaram caminhos que não podem ser ignorados e devem ser incluídos na busca por uma definição de justiça social para a Organização do Conhecimento.

## 5 CONSIDERAÇÕES FINAIS

Buscamos, nesse trabalho, analisar a presença da justiça social na Organização do Conhecimento a partir de um levantamento bibliográfico em bases de dados internacionais. Para tanto, partimos em um primeiro momento da compreensão de uma Organização do Conhecimento crítica que encontra fulcro na obra de dois marcos teóricos, Berman e Olson. Para os autores, a lógica da reprodução dos preconceitos que mantém o sistema de privilégios da nossa sociedade nos SOCs hegemônicos está vinculada a uma “provável psicologia do leitor” e à criação de um “público singular”, abstrato e não condizente com toda a comunidade de usuários reais e potenciais.

Nos artigos recuperados e analisados aqui, as diversas críticas e alternativas propostas a esses problemas apontam, cada uma a sua maneira, para a ampliação da participação dos grupos socialmente marginalizados na construção, crítica, atualização e adaptação dos SOCs. Porém, não abordando necessariamente o conceito de justiça social, mas incorporando-o de forma implícita nas discussões.

Ao discutirmos a compreensão da justiça social como distribuição e como reconhecimento, podemos retomar o conceito de paridade de participação de Fraser (2002), segundo o qual todos os membros de uma sociedade devem interagir entre si como pares. Para a autora as teorias distributivas não dão conta do conjunto de injustiças e opressões em nossa sociedade e, conseqüentemente, não atendem as demandas dos crescentes e diversificados movimentos sociais.

Por outro lado, a sua mera substituição pelo “reconhecimento” pode causar uma incompreensão da relação entre a desigualdade econômica e a ausência de justiça social. Ao propor a justiça social como paridade de participação, Fraser (2002) defende uma distribuição de recursos materiais que garanta independência e voz para todos, e que “padrões institucionalizados de valor cultural expressem igual respeito por todos os participantes e garantam iguais oportunidades para alcançar a consideração social” (FRASER, 2002, p. 13).

Acreditamos que a contribuição de Fraser para um conceito de justiça social, amparado na distribuição e no reconhecimento, possibilite-nos um aprofundamento no debate da construção de uma Organização do Conhecimento crítica, orientada para a justiça social, garantindo uma crítica teórica e a construção de SOCs destinados a uma ampliação da paridade de participação nos processos de organização do conhecimento em arquivos e bibliotecas. Por fim, ressaltamos que todos os trabalhos – a partir de diferentes perspectivas teóricas e metodológicas – corroboram o sentido de buscar maior participação de grupos marginalizados como forma de ampliar a justiça social nas bibliotecas e arquivos, diminuindo vieses, preconceitos e desigualdade informacional.

## 6 REFERÊNCIAS

ADLER, M. Transcending Library Catalogs: a Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in Library Thing for Transgender Books. **Journal of Web Librarianship**, v. 3, p. 309–331, 2009.

ADLER, M. The Case for Taxonomic Reparations. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 43, n. 8, p. 630-640, 2016. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko\\_43\\_2016\\_8\\_f.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko_43_2016_8_f.pdf). Acesso em: 10 fev. 2022.

ADLER, M; HARPER, L.M. Race and Ethnicity in Classification Systems: teaching knowledge organization from a social justice perspective. **Library Trends**, Baltimore, Johns Hopkins University Press, v. 67, n. 1, p. 52-73, 2018.

BAK, G; ALLARD, D; FERRIS, S. Knowledge Organization as Knowledge Creation: surfacing community participation in archival arrangement and description. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 46, n. 7, p. 502-521, 2019. Disponível em: [www.ergon-verlag.de/isko\\_ko/downloads/ko\\_46\\_2019\\_7\\_b.pdf](http://www.ergon-verlag.de/isko_ko/downloads/ko_46_2019_7_b.pdf). Acesso em: 10 fev. 2022.

BARDIN, L. **Análise de Conteúdo**. Lisboa, Portugal: Edições 70, LDA, 2009.

BERMAN, S. **Prejudices and antipathies**. Metuchen: Scarecrow Press, 1971. Disponível em: <https://www.sanfordberman.org/prejant/prejant.pdf>. Acesso em: 10 fev. 2022.

BERMAN, S; GROSS, T. Expand, Humanize, Simplify: an interview with sandy berman. **Cataloging & Classification Quarterly**, v. 55, n. 6, p. 347-360, 29 jun. 2017.

COPE, J. Neoliberalism and Library & Information Science: Using Karl Polanyi's Fictitious Commodity as an Alternative to Neoliberal Conceptions of Information. **Progressive Librarian**, v. 43, p. 67-80, 2014. Disponível em: <http://www.progressivelibrariansguild.org/PL/PL43/067.pdf>. Acesso em: 10 fev. 2022.

DICK, A. L.; BURGER, M. Transforming Subject Access: Some Critical Issues for South African Information Professionals. In: BERMAN, S.; DANKY, J. (Ed.). **Alternative Library Literature**, 1996–1997. Jefferson: McFarland & Company, 1998.

DOYLE, A. M.; LAWSON, K.; DUPONT, S. Indigenization of Knowledge Organization at the Xwi7xwa Library. **International Journal of Library And Information Studies**, Vancouver: University Of British Columbia Library, v. 2, n. 13, p. 107-134, dez. 2015. Disponível em: <https://open.library.ubc.ca/soa/cIRcle/collections/ubclibraryandarchives/494/items/1.0103204>. Acesso em: 10 fev. 2022.

DRUCKER, D. J. How Subjects Matter: the kinsey institute's sexual nomenclature. **Information & Culture: A Journal of History**, Austin: University of Texas Press, v. 52, n. 2, p. 207-228, 2017. Disponível em: <https://muse.jhu.edu/article/658180>. Acesso em: 10 fev. 2022.

DUDLEY, M Q. A Library Matter of Genocide: the library of congress and the historiography of the native american holocaust. **International Indigenous Policy Journal**, v. 8, n. 2, 2017. Disponível em: <https://ojs.lib.uwo.ca/index.php/iipj/article/view/7519/6163>. Acesso em: 10 fev. 2022.

FOX, M. J. "Priorities of Arrangement" or a "Hierarchy of Oppressions?": perspectives on intersectionality in knowledge organization. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 43, n. 5, p. 373-383, 2016. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko\\_43\\_2016\\_5\\_h.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko_43_2016_5_h.pdf). Acesso em: 10 fev. 2022.

FRASER, N. A justiça social na globalização: redistribuição, reconhecimento e participação. **Revista Crítica de Ciências Sociais**, n. 63, p. 7-20, 2002. Disponível em: <https://www.ces.uc.pt/publicacoes/rccs/artigos/63/RCCS63-Nancy%20Fraser-007-020.pdf>. Acesso em: 10 fev. 2022.

FURNER, J. Dewey Deracialized: A Critical Race-Theoretic Perspective. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 34, n. 3, p. 144-168, 2007.

GUTIÉRREZ, A.G. Declassifying Knowledge Organization. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 41, n. 5, p. 393-409, 2014. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko\\_41\\_2014\\_5\\_g.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko_41_2014_5_g.pdf). Acesso em: 10 fev. 2022.

GILLILAND, A. J. Contemplating Co-creator Rights in Archival Description. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 39, n. 5, p. 340-346, 2012. Disponível em: [www.ergon-verlag.de/isko\\_ko/downloads/ko\\_39\\_2012\\_5\\_e.pdf](http://www.ergon-verlag.de/isko_ko/downloads/ko_39_2012_5_e.pdf). Acesso em: 10 fev. 2022.

HJØRLAND, B.; ALBRECHTSEN, A. Toward a new horizon in Information Science: Domain-Analysis. **Journal of the American Society for Information Science**, v. 46, n. 6, p. 400-25, 1995.

HJØRLAND, B. Epistemology and the socio-cognitive perspective in information science. **Journal of the American Society for Information Science and Technology**, vol. 53, n. 4, p. 257-270, 2002. Disponível em: <https://doi.org/10.1002/asi.10042>.

KAIN, P. **How to do a close reading**. S.l.: Harvard, 1998. Disponível em: <http://www.fas.harvard.edu/~wricntr/documents/CloseReading.html>. Acesso em: 22 ago. 2012.

LEAZER, G. H.; MONTOYA, R. The Politics of Knowledge Organization: introduction to the special issue. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 47, n. 5, p. 367-371, 2020. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko\\_47\\_2020\\_5\\_a.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko_47_2020_5_a.pdf). Acesso em: 10 fev. 2022.

LORDE, A. **Irmã outsider: ensaios e conferências**. Belo Horizonte: Autêntica, 2019.

MAI, J. E. Ethics, Values and Morality in Contemporary Library Classifications. **Knowledge Organization**, v. 40, n. 4, p. 242-253, 2013.

MARSHALL, J. K. **On Equal Terms: A Thesaurus for Nonsexist Indexing and Cataloging**, New York: Neal-Schuman, 1977.

MARTINEZ-ÁVILA, D.; SEMIDÃO, R.; FERREIRA, M. Methodological aspects of critical theories in knowledge organization. **Knowledge Organization**,

Würzburg: Ergon Verlag, v. 43, n. 2, p. 118-125, 2016. Disponível em: [www.ergon-verlag.de/isko\\_ko/downloads/ko\\_43\\_2016\\_2\\_f.pdf](http://www.ergon-verlag.de/isko_ko/downloads/ko_43_2016_2_f.pdf). Acesso em: 10 fev. 2022.

McCLENNEN, S. A. **How to do a close reading**. S.l.: Personal PSU, 2001. Disponível em: <http://www.personal.psu.edu/users/s/a/sam50/closerread.htm>. Acesso em: 30 jul. 2012.

OLSON, H. A. **The power to name**: locating the limits or subject representation in libraries. Dordrecht: Kluwer Academic Publisher, 2002.

SADLER, B.; BOURG, C. Feminism and the Future of Library Discovery. **Code4 Journal**, 28, p.1-5, 2015. Disponível em: <https://journal.code4lib.org/articles/10425>. Acesso em: 10 fev. 2022.

SMIT, J. W. A informação na Ciência da Informação. **InCID: Revista de Ciência da Informação e Documentação**, vol. 3, n. 2, p. 84-101, 2012.

SNOW, K. Shifting sands and the prophet's dream: Exploring the future of information organization education. **Journal of Library and Information Science Education**, v. 60, n. 2, p. 139-151, 2019.

SZOSTAK, R. Classifying for Social Diversity. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 41, n. 2, p. 160-170, 2014. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko\\_41\\_2014\\_2\\_g.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko_41_2014_2_g.pdf). Acesso em: 10 fev. 2022.

THARANI, K. Just KOS! Enriching Digital Collections with Hypertexts to Enhance Accessibility of Non-Western Knowledge Materials in Libraries. **Knowledge Organization**, Würzburg: Ergon Verlag, v. 47, n. 3, p. 220-230, 2020. Disponível em: [www.ergon-verlag.de/isko\\_ko/downloads/ko\\_47\\_2020\\_3\\_b.pdf](http://www.ergon-verlag.de/isko_ko/downloads/ko_47_2020_3_b.pdf). Acesso em: 10 fev. 2022.

THORPE, Kristen. Transformative praxis – building spaces for Indigenous self-determination in libraries and archives – In the Library with the Lead Pipe. **In the Library with the lead pipe**: an open access, open peer reviewed journal, S.l., 23 Jan. 2019. Disponível em: <http://www.inthelibrarywiththeleadpipe.org/2019/transformative-praxis/>. Acesso em: 10 fev. 2022.

WATSON, B. M. There was Sex but no Sexuality: critical cataloging and the classification of asexuality in lcs. **Cataloging & Classification Quarterly**, v. 58, n. 6, p. 547-565, ago. 2020. Disponível em: <https://scholarworks.iu.edu/dspace/handle/2022/25766>. Acesso em: 10 fev. 2022.

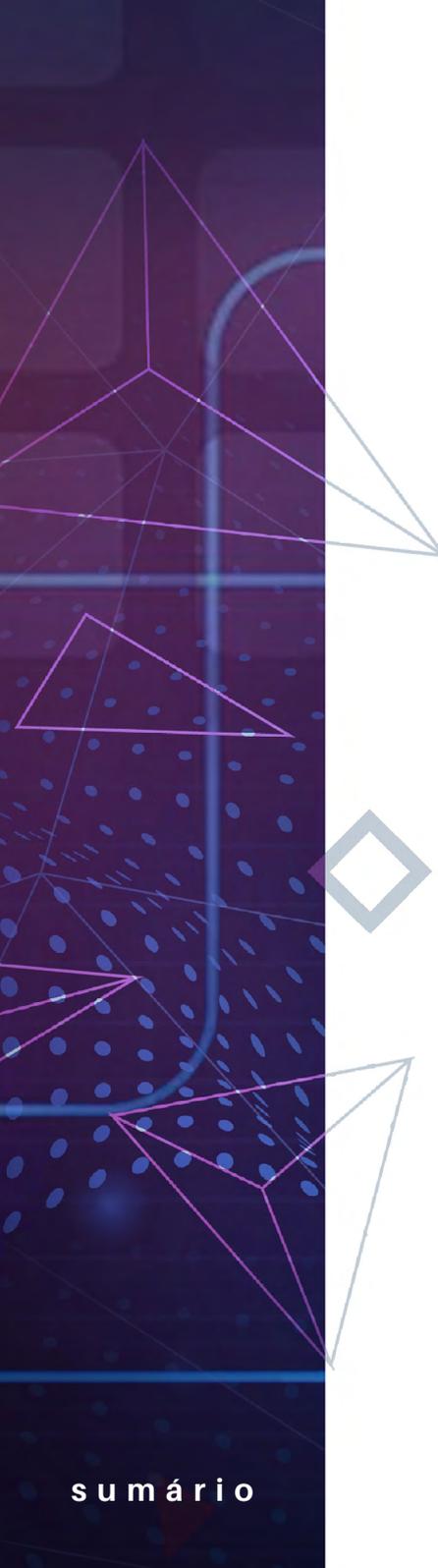
# 15

Leolíbia Luana Linden

**Organização  
do conhecimento e Arquivologia:  
diálogos e prospecções**

**Knowledge organization  
and Archival Science:  
dialogues and prospects**

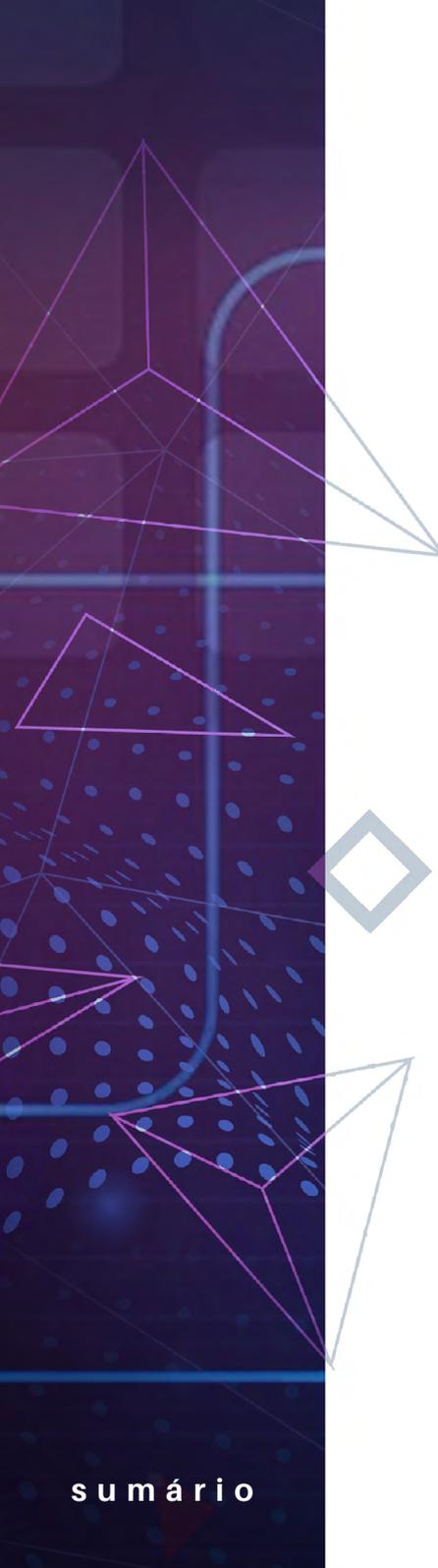
DOI: 10.31560/pimentacultural/2022.95613.15



**Resumo:**

Cada vez mais são desenvolvidas pesquisas que alinham estudos teóricos e metodológicos da Organização do Conhecimento à Arquivologia, nas quais, é possível identificar os percursos e os traços da organização e representação no âmbito dos arquivos. Dessa forma, o presente estudo busca contextualizar a Organização do Conhecimento e a Arquivologia como áreas autônomas que possuem diversos aspectos de convergência, analisando os diálogos estabelecidos entre ambas as áreas por meio da análise das funções arquivísticas de classificação e descrição à luz da organização e representação do conhecimento e da informação. Do mesmo modo, também se busca identificar as possíveis prospecções para o desenvolvimento de estudos que ampliem as possibilidades de análise nesse campo de intersecção, para que seja possível colaborar no avanço das discussões acerca da temática. Justifica-se pelas possibilidades de avanço teórico e metodológico derivados da aproximação e interposição de ambas as áreas. Por fim, ressalta-se a necessidade de melhor explorar as relações interdisciplinares entre Organização do Conhecimento e Arquivologia colaborando com seu mútuo desenvolvimento.

**Palavras-Chave:** Arquivologia; organização do conhecimento; representação em arquivos.



**Abstract:**

*There are more and more academic research that align theoretical and methodological studies on the Knowledge Organization to Archival Science, in which it is possible to identify the paths and traces of organization and representation in the context of archives. Thus, this study seeks to contextualize Knowledge Organization and Archival Science as autonomous areas that have different aspects of convergence, analyzing the updated dialogues between both areas through the analysis of archival functions of classification and description in light of the organization, representation and knowledge and of information. Likewise, it also seeks to identify possible prospects for the development of studies that expand the possibilities of analysis in this field of intersection, so that it is possible to collaborate in advancing the implications on the subject. The possibilities for theoretical and methodological advances are justified by the approach and interposition of both areas. Finally, it emphasizes the need to better explore the interdisciplinary relationships between Knowledge Organization and Archives, collaborating with their mutual development.*

**Keywords:** Archival science; knowledge organization; archives representation.

## 1 INTRODUÇÃO

A Arquivologia, enquanto disciplina científica inserida nas Ciências Sociais Aplicadas, trabalha e desenvolve-se no Campo da Informação (MARQUES, 2011), ou seja, um espaço de interlocuções entre disciplinas que têm a informação e/ou desdobramentos dela, na suas mais diversas interpretações e usos, como objeto de estudo. Dentre as disciplinas e as áreas que estabelecem esse diálogo com a Arquivologia, destacamos a Organização do Conhecimento. Por sua vez, a Organização do Conhecimento (OC) é um campo de pesquisa, ensino e prática, relacionada à Ciência da Informação, que se preocupa com a natureza e a qualidade dos processos de organização do conhecimento, bem como os Sistemas de Organização do Conhecimento (HJØRLAND, 2008).

As funções arquivísticas de classificação e de descrição, intervenções arquivísticas, são pensadas, elaboradas e contribuem para o desenvolvimento de metodologias que possibilitam facilitar a recuperação da informação. Dessa forma, é possível aproximar a Arquivologia da Organização do Conhecimento, apropriando-se das metodologias de tratamento e recuperação da informação e do conhecimento propostas por esta (BARROS, 2016). Assim, chegamos a um cenário em que a classificação e descrição arquivística são consideradas processos de organização e representação do conhecimento e da informação desenvolvidos nos arquivos (BARROS, 2010; SILVA, 2012; VITAL; MEDEIROS; BRASCHER, 2017; MARTINS, 2019; BARROS; SOUSA, 2019).

A proposta desse estudo é compreender os diálogos existentes entre OC e Arquivologia no que tange a concepção da classificação e da descrição como intervenções que correspondem à organização e representação do conhecimento e da informação no âmbito dos arquivos. Procura-se ressaltar as diferentes possibilidades de pesquisa a serem desenvolvidas que valorizem ainda mais a intersecção entre as áreas.

## 2 ORGANIZAÇÃO DO CONHECIMENTO E ARQUIVOLOGIA

### 2.1 DELINEAMENTOS DA ORGANIZAÇÃO DO CONHECIMENTO

Pondera-se que o termo OC foi utilizado inicialmente por Henry Evelyn Bliss, que também concebeu a OC como autônoma no campo científico por meio de suas obras *The Organization of Knowledge and the System of Sciences* (1929) e *Organization of Knowledge in Libraries and the Subject Approach to Books* (1933) (DAHLBERG, 1995, p. 10). Diferente da perspectiva de Bliss que restringia a aplicação do conceito ao ambiente biblioteconômico, Dahlberg (1993, p. 211) afirma que a OC é a ciência que sistematiza conceitos de acordo com suas características. Dessa forma, a autora defende que a teoria do conceito seria o pressuposto mais importante da fundamentação teórica da OC, uma vez que esta deve ser sistematizada segundo unidades do conhecimento (conceitos) e seus elementos de conhecimento (características). Essa ampliação das possibilidades de aplicação da classificação e teoria do conceito a outras áreas pode ter posicionado a OC como um possível novo campo de estudo (SALES, 2015).

Complementarmente, Bräscher e Café (2008) definem a OC como:

Delineamos a **organização do conhecimento** como o processo de modelagem do conhecimento que visa a construção de representações do conhecimento. Esse processo tem por base a análise do conceito e de suas características para o estabelecimento da posição que cada conceito ocupa num determinado domínio, bem como das suas relações com os demais conceitos que compõem esse sistema nocional (BRÄSCHER; CAFÉ, 2008, p. 8, grifo nosso).

Conjugando-se Dahlberg (1993), Sales (2015) e Bräscher e Café (2008), entendemos que a OC é um campo científico que tem seu cerne na modelagem do conhecimento e cujas bases estão na teoria do conceito. Ocupa-se em observar os domínios nos quais os conceitos são utilizados e entrega, para a sociedade, modelos apresentados em representações do conhecimento.

Hjørland (2008) afirma que a OC é um campo de pesquisa, ensino e prática que pode se ocupar de atividades amplas e restritas. Amplas no que diz a respeito à “[...] divisão social do trabalho, a organização das universidades e outras instituições de pesquisa e ensino superior, a estrutura das disciplinas e as profissões, a organização social dos meios de comunicação, a produção e disseminação do conhecimento”. Enquanto, no sentido restrito, vincula-se às atividades específicas que se configuram como técnicas, “[...] como descrição de documentos, indexação e classificação” (HJØRLAND, 2008, p. 86).

Buscando explorar as ações e produtos, ou, ainda, os modelos (representações) que são consequências subjacentes a esse processo, Barité (2001) conceitua amplamente o objeto de estudo e a atuação da OC:

O objeto de estudo da Organização do conhecimento é – a nosso juízo – o conhecimento socializado, e como disciplina dá conta do desenvolvimento de técnicas para a construção, gestão, uso e avaliação de classificações científicas, taxonomias, nomenclaturas e linguagens documentais. De outra parte, traz metodologias de uso e recuperação por linguagem natural (BARITÉ, 2001, p. 41).

Nessa perspectiva, Dahlberg (1993), Barité (2001), Hjørland (2008) e Bräscher e Café (2008) nos conduzem à reflexão de que a OC, para além de uma disciplina científica, protagoniza o desenvolvimento de metodologias de gestão e organização do conhecimento socializado. Essas metodologias, inerentes à gestão e organização no contexto da OC, podem ser utilizadas no fomento de diálogos interdisciplinares

entre a OC e a Arquivologia. Diálogos estes que transcendem os aspectos teóricos e pousam na perspectiva pragmática.

Em paralelo da OC, segundo Braschër e Café (2008, p. 8), a Organização da Informação (OI) é um processo que envolve a descrição física e de conteúdo dos documentos e tem como produto a representação da informação, que compõe os elementos descritivos de um objeto. Lima e Álvares (2012, p. 35) afirmam que o principal objetivo da OI é possibilitar a recuperação e o acesso à informação por meio da estruturação dos elementos de organização do conhecimento.

Guimarães (2009, p. 106, grifo nosso) pondera que:

[...] a **organização da informação** deve ser entendida como um conjunto de procedimentos que incidem sobre um conhecimento socializado (que, por sua vez, é um produto social e tem uma utilidade social e individual), os quais variam em virtude dos contextos em que são produzidos ou os fins a que se destinam, pois é a partir destes que se desenvolvem os parâmetros de organização.

Sob a égide da OI, a descrição de conteúdo dos documentos (BRASCHËR; CAFÉ, 2008), visando à recuperação e ao acesso (LIMA; ALVARES, 2012), apoiada nos elementos entendidos nessa pesquisa como representações e tendo como objetivo social contribuições (GUIMARÃES, 2009) para a sociedade, como um todo, pode ser identificada como um ponto de intersecção com a Arquivologia.

Svenonius (2000) corrobora com a reflexão anterior, quando indica que, para ser passível de organização, a informação precisa ser descrita. Uma vez descrita, passa-se a ter um enunciado, que, por sua vez, representará as propriedades do objeto informacional ou de suas relações com demais objetos que o identificam. O objetivo de descrever um documento é comunicá-lo, por meio de uma linguagem específica, desenvolvida e aplicada de acordo com um conjunto de preceitos que contemplam determinados atributos.

Nesse sentido, a organização da informação une os objetos que apresentam características semelhantes, em relação a seu conteúdo e formato, aos elementos que os possam descrever (SVENONIUS, 2000). Para o entendimento dessa pesquisa, o termo representação da informação pode ser entendido como um processo e, também, como produto(s) dele derivado(s). Nesse sentido, a representação busca estabelecer uma correspondência entre o objeto a ser representado e o padrão de representação utilizado (YAKEL, 2003).

Em síntese, a OI compreende a organização de conjuntos de objetos informacionais estruturados de maneira sistematizada. Já a OC propõe-se na constituição de modelos de mundo a partir de abstrações da realidade (DAHLBERG, 1993; SVENONIUS, 2000; BRÄSCHER; CAFÉ, 2008; LIMA & ALVARES, 2012), demonstrados no Quadro 1:

**Quadro 1 – Síntese dos elementos de OI e OC**

	<b>ORGANIZAÇÃO DA INFORMAÇÃO</b>	<b>ORGANIZAÇÃO DO CONHECIMENTO</b>
<b>CONTEXTO</b>	Objetos físicos	Cognição
<b>PROCESSO</b>	Descrição física e de conteúdo dos objetos informacionais.	Construção de modelos de mundo que se constituem em abstrações da realidade a partir de unidades do pensamento (conceitos).
<b>MATERIALI-ZAÇÃO</b>	Representação da informação, conjunto de elementos descritivos que representam os atributos de um objeto informacional específico.	Representação do conhecimento que reflete, assim, um modelo de abstração do mundo real, construído para determinada finalidade.

Fonte: Elaborado pela autora, a partir de Bräscher e Café (2008).

Como resultado dos processos de OC e OI, identificam-se dois tipos de representação: a Representação do Conhecimento (RC) e a Representação da Informação (RI), estes, por sua vez, processados em Sistemas de Organização do Conhecimento (SOCs). Os SOCs são planejados e desenvolvidos para auxiliar nos procedimentos de gestão

e recuperação do conhecimento registrado, como, por exemplo, tesouros, ontologias e demais instrumentos de controle de vocabulário. Eles cumprem importante papel de padronização da terminologia adotada para organização e recuperação de informações, ao delimitar o uso de termos e definir conceitos (teoria do conceito) e suas relações em alguma área do conhecimento, de forma compartilhada e consensual (DAHLBERG, 1993; BRÄSCHER; CARLAN, 2010).

Barité (2001), em uma visão mais ampla, descreve que, enquanto disciplina, a organização do conhecimento desenvolve técnicas para a construção de representações abstratas da realidade por meio de seus instrumentos (SOCs), transpondo esse aporte teórico e metodológico a aplicações em unidades de informação. Segundo Hodge (2000, p. 1, tradução nossa, grifo nosso), os SOCs:

[...] englobam todos os tipos de instrumentos usados para organizar a informação e promover o gerenciamento do conhecimento. Incluem os **esquemas de classificação** que organizam materiais em nível geral (como livros em estantes), **cabeçalhos de assunto** que provêm acesso mais detalhado e listas de autoridade que controlam versões variantes de chaves de acesso à informação (nomes geográficos e nomes de pessoas). Incluem, ainda, esquemas menos tradicionais, tais como redes semânticas e ontologias.

No entendimento do autor, percebe-se a ampliação das aplicabilidades de SOC no que tange à organização da informação aplicada à Arquivologia, como, na classificação, materializada em seus esquemas e, na descrição, quando da definição de elementos de descritivos, por exemplo, pelos cabeçalhos de assunto. Corroborando com esse entendimento, Hodge (2000, p. 1, tradução nossa) assinala que “[...] sistemas de organização do conhecimento são mecanismos para organizar informações e, estão no ‘coração’ de todas as bibliotecas, museus e arquivos.”

Desse modo, os SOCs passam a se tornar um meio de interface comunicativa entre produtores e utilizadores da informação, por representarem, de forma materializada, a organização do conhecimento de determinada realidade (AGUIAR; KOBASHI, 2013, p. 8). Por meio dos SOCs, é possível ampliar a compreensão das estruturas abstratas propostas com a finalidade de melhorar os processos de comunicação, pesquisa, recuperação da informação e estruturas de modelos conceituais.

## 2.2 ARQUIVOLOGIA

Uma vez que tudo é social e culturalmente construído no mundo pós-moderno, desconstruir e reformular parecem ser a melhor forma de refletir a diversidade na produção e na organização do conhecimento arquivístico contemporâneo (TOGNOLI, 2012). Por outro lado, é preciso avançar na teoria e na prática para que a arquivística possa ter uma cientificidade indiscutível, isto é, a liberação do senso comum (LOPES, 2009).

O ano de 1789 marcou o início da Idade Contemporânea com a Revolução Francesa que inaugura um período de garantia de direitos dos cidadãos. Nesse contexto, surge o primeiro Arquivo Nacional do mundo, durante a Assembleia Nacional Francesa em 1789. Esse fato repercutiu em importantes realizações para o campo arquivístico por meio da criação de uma administração nacional dos arquivos, da publicidade de acesso aos arquivos e da responsabilidade do Estado em custodiar esses documentos (SCHELLENBERG, 2006, p. 26).

Em 1922, o inglês Hilary Jenkinson registra, em *A manual of archive administration*, as teorias e práticas de arquivo de acordo com suas experiências. Jenkinson enfatizou o trabalho do arquivista como guardião de evidências, com a finalidade de tornar o trabalho de outras pessoas possível. Para Cook (1997), o acúmulo de documentos gerados

pela Primeira Guerra Mundial (1914-1918), além de documentos produzidos durante a Idade Média, possibilitou uma visão diferente daquela partilhada pelos arquivistas holandeses em relação ao arquivo.

Ainda de acordo com Cook (1997), Jenkinson (1922) preocupou-se em estudar o valor dos documentos de ordem administrativa e, posteriormente, documentos de valor de prova, enfatizando as características de organicidade, imparcialidade, unicidade e autenticidade. Dessa forma, Jenkinson despertou o interesse sobre os problemas de avaliação documental, que serviram como insumo necessário às teorias norte-americanas relacionadas ao ciclo de vida dos documentos.

Em 1934, é criado o National Archives nos Estados Unidos, por consequência de situações de sinistros na documentação e influência da Associação Histórica Americana (SCHELLENBERG, 2006).

[...] os Arquivos Nacionais assumiram a responsabilidade por 10 milhões de metros cúbicos de documentos que haviam sido acumulados durante um período de 150 anos. Além disso, programas iniciados durante a Grande Depressão resultaram em uma expansão dos serviços governamentais e em um aumento no volume de documentos [...] (STAPLETON, 1983, p. 76, tradução nossa).

A partir das revoluções tecnológicas e sociais na década de 1980, começa a ser discutida a inclusão da arquivística na área da ciência da informação (RIBEIRO, 2011). Em 1982, Carol Couture e Jean-Yves Rosseau publicaram *Les archives au XX siècle*, em que é proposta uma arquivística preocupada em integrar tanto as preocupações do arquivo permanente quanto as atribuídas ao arquivo administrativo:

[...] garantir a unidade e a continuidade das intervenções do arquivista nos documentos de um organismo e permitir assim uma perspectiva do princípio das três idades e das noções de valor primário e secundário; permitir a articulação e a estruturação das atividades arquivísticas numa política de organização de arquivos; integrar o valor primário e o valor secundário numa definição alargada de arquivo (ROUSSEAU; COUTURE, 1998, p. 70).

Essas intervenções que necessitam de continuidade são especificadas pelos próprios autores na sistematização das funções arquivísticas, que seriam as atividades norteadoras no contexto do arquivo: produção/criação, avaliação, classificação, descrição, difusão, preservação e aquisição. As funções arquivísticas podem ser cumpridas de maneira dinâmica, gradual e até mesmo simultânea, o que abre espaço para sua interlocução e integração com a Organização do Conhecimento.

### 3 INTERSECÇÕES ENTRE ORGANIZAÇÃO DO CONHECIMENTO E ARQUIVOLOGIA

Pesquisas na área de Arquivologia indicam que os estudos teóricos e metodológicos da área de Organização e Representação do Conhecimento e da Informação são essenciais para o aprofundamento da Arquivologia e vice e versa. Essas pesquisas aproximam-se cada vez mais do entendimento das funções de classificação e descrição arquivística como processos de representação do conhecimento e da informação (BARROS, 2010; 2016; TOGNOLI, 2012; SILVA, 2012; VITAL; MEDEIROS; BRASCHER, 2017; MARTINS, 2019; BARROS, SOUSA, 2019; LEHMKUHL *et al.*, 2019).

Lehmkuhl *et al.* (2019) confrontam as características dos agrupamentos de SOC, sistematizados por Hodge (2000), com as funções arquivísticas, concluindo que a OC e a Arquivologia possuem bases metodológicas semelhantes quanto à construção de representações de domínios de conhecimento e vocabulários, que resultam em distintos instrumentos, tendo em vista as especificidades de cada contexto. Assim, os arquivos trabalham com domínios específicos do conhecimento, apresentando metodologias que também contribuem com a OC no sentido de abordar as diversidades sociais e políticas inerentes às atividades humanas.

Barros (2014), nessa linha, parte do pressuposto de que a representação vinculada à Arquivologia está baseada nas funções arquivísticas de classificação e descrição, indicando que o termo “representação” amplia as possibilidades de compreensão desses processos. O autor expande a reflexão indicando que

A compreensão da **descrição** e **classificação** enquanto atividades conjuntas possibilita a existência de relações interdisciplinares entre a Arquivística e a Organização do Conhecimento, possibilitando as primeiras abordagens teórico metodológicas constituindo um novo momento para a organização de arquivos e neste universo apropriar-se de metodologias de tratamento [...] (BARROS, 2016, p. 39-40, grifo nosso).

Nesse entendimento, abre-se as possibilidades acerca da teoria e da prática arquivística baseadas nos preceitos advindos da área da Organização do Conhecimento, dado o entendimento de que as funções arquivísticas de classificação e descrição são as atividades a serem planejadas e implementadas de maneira conjunta. A partir desse entendimento, as lentes teóricas dessa análise se calcam respaldadas pelos autores que se seguem.

Esteban Navarro (1993) é pontual quando indica que a classificação e a descrição arquivísticas são operações nas quais ocorrem a representação do conhecimento e da informação nos arquivos. Tognoli (2012) assinala que a classificação e a descrição arquivística são a materialização da representação nos arquivos, sobretudo na definição da proveniência dos documentos:

[...] 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 de **classificação/arranjo e descrição** do conhecimento arquivístico (TOGNOLI, 2012, p. 85, grifo nosso).

Tognoli e Barros (2015) buscam evidenciar elementos históricos e conceituais de classificação e descrição, relacionando-os diretamente com a organização e representação de arquivos:

A **Classificação** e a **Descrição** são atividades nucleares no processo de organização e representação de arquivos. Esse conhecimento pode ser entendido como todo conhecimento gerado a partir das atividades realizadas por pessoas físicas ou jurídicas e que resultam em informações orgânicas (TOGNOLI; BARROS, 2015, p. 94, grifo nosso).

No protagonismo dos instrumentos arquivísticos, as funções arquivísticas da classificação e da descrição, técnicas arquivísticas, são pensadas, elaboradas e contribuem no desenvolvimento de metodologias que possibilitam facilitar a recuperação da informação. Dessa forma, enquanto atividades conjuntas, aproximam a Arquivologia da Organização do Conhecimento, apropriando-se das metodologias de tratamento e recuperação da informação e do conhecimento (BARROS, 2016).

Freitas (2017) trata da “Classificação Arquivística como Sistema de Organização do Conhecimento” em uma das seções do texto, apresentando um vasto aprofundamento teórico sobre o tema. Conclui seu texto afirmando:

Sendo assim, a **classificação arquivística**, como um sistema de organização do conhecimento, objetiva assegurar a proteção e salvaguarda dos documentos, que servem as estratégias dos negócios organizacionais, pois o processo de construção das classificações arquivísticas resulta em uma representação do universo contextual de uma instituição (FREITAS, 2017, p. 129).

Segundo Yakel (2003), sobre o processo de representação da informação aplicado aos arquivos:

Representação refere-se a ambos os processos **arranjo** (respeitando ou desrespeitando a ordem) e **descrição**, tais como a criação de ferramentas de acesso (guias, inventários, instrumentos de pesquisa, bibliografia registros) ou sistemas (catálogos de

cartões, bancos de dados bibliográficos, bancos de dados EAD) resultantes dessas atividades (YAKEL, 2003, p. 2, grifo nosso).

Endossando esta discussão, Aguiar e Kobashi (2013) acentuam que é no exercício das atividades arquivísticas de classificação e de descrição que é realizada a representação da informação em documentos de arquivo. As autoras ainda asseguram que, para a organização e representação da informação arquivística, tem-se como ponto de partida os princípios de proveniência e da organicidade.

Destacam-se os conceitos da “proveniência e da organicidade” enquanto princípios teórico-conceituais para orientar os fundamentos metodológicos da Organização e Representação da Informação Arquivística (ORIA). O princípio da proveniência vincula e fixa na informação arquivística a estrutura e o contexto ao seu produtor, já a organicidade reflete a estrutura, funções e atividades da entidade produtora ou acumuladora (AGUIAR; KOBASHI, 2013, p. 9).

Nessa mesma linha, Silva e Moreira (2015) indicam que, dentre as intervenções arquivísticas, é possível identificar a classificação e a descrição, expressadas por meio da indexação, como elementos que se conectam diretamente à OC em sua forma e materialização mais específica, por definir de qual maneira o conhecimento será representado em determinado contexto, por meio de seus processos e instrumentos, quando por fim socializado. Assim,

A organização do conhecimento se preocupa, em seu sentido amplo, em identificar como o conhecimento é socialmente organizado para então desenvolver as práticas de organização do conhecimento, tais como **indexação** e **classificação**, que são o sentido mais estreito de Organização do Conhecimento (SILVA; MOREIRA, 2015, p. 2, grifo nosso).

Silva *et al.* (2017, p. 51) nos asseguram que, “[...] por representarem pontos de intersecção entre as demais ações relacionadas aos documentos de arquivo, revelando um universo axiológico rico e diversificado”, as funções arquivísticas de classificação e descrição

são elementarmente indispensáveis e integradas na organização e representação do conhecimento arquivístico. Barros (2016, p. 35) complementa explicando que a teoria arquivística começa a ser entendida “como um processo único e contínuo de análise e síntese da informação contida nos arquivos, ou seja, o processo compreendido aqui como a representação orgânico-contextual”.

Vital e Bräscher (2016) buscam identificar as funções de classificação e descrição arquivísticas no contexto do processo de organização e representação da informação no âmbito dos arquivos por meio de suas características. As autoras relacionam a classificação, a ordenação física e intelectual dos documentos, enquanto a descrição se ocupa da organização e representação da informação. A classificação, neste íterim, em nossa visão, tem resultado na manutenção das relações entre os documentos, e a descrição resulta na recuperação das informações contidas nos documentos de arquivo.

O desenvolvimento de pesquisas e estudos acerca da temática da descrição arquivística e seus pontos de intersecção interdisciplinar com a área da representação da informação trouxe à área novas perspectivas teóricas e metodológicas. Para tanto, na visão de Michael Cook (apud HAGEN, 1998, p. 2), “a descrição tem como base a teoria da representação”.

Nesse entendimento, García Marco (1995, p. 110, tradução nossa) indica que “[...] a descrição documental e arquivística com designação das tarefas de representação documental para facilitar o acesso aos fundos de um arquivo tem sido denominada descrição”. Essa concepção foi ampliada pelos aspectos convergentes nas áreas de Ciência da Informação e Organização do Conhecimento com a Arquivística e, claro, pelas novas teorias inauguradas pelas abordagens arquivísticas contemporâneas.

No âmbito da produção científica nacional, Lehmkuhl *et al.* (2019) indicam o crescimento exponencial dos trabalhos publicados

no âmbito da Organização do Conhecimento, no Brasil, por meio de uma análise nos anais da ISKO Brasil (2012-2015). Estes evidenciam pontos de convergência e intersecção entre ambas as áreas que trilham os caminhos da interdisciplinaridade, em um período de ascensão da produção científica, no qual as temáticas de maior interesse têm sido: a) classificação arquivística; b) descrição arquivística; e c) linguagens documentárias aplicadas aos arquivos.

Barros (2016) descreve, ao longo de seu estudo, o percurso histórico da classificação e descrição arquivísticas, visando demarcar o espaço de atuação do processo de representação em arquivos. Indica, ainda, que a classificação serviria como fundamento para o processo de representação, enquanto a descrição faria uso das categorias de classificação para representar os documentos.

Com isso, Vital, Medeiros e Brascher (2017) buscam alinhar o processo de organização e representação do conhecimento à classificação arquivística, assim como o processo de organização e representação da informação à descrição arquivística, demonstrado abaixo:

**Quadro 2 – Síntese de organização e representação do conhecimento e da informação na Arquivologia**

ORGANIZAÇÃO	REPRESENTAÇÃO	ARQUIVOLOGIA
ORGANIZAÇÃO DO CONHECIMENTO	REPRESENTAÇÃO DO CONHECIMENTO	CLASSIFICAÇÃO ARQUIVÍSTICA
ORGANIZAÇÃO DA INFORMAÇÃO	REPRESENTAÇÃO DA INFORMAÇÃO	DESCRIÇÃO ARQUIVÍSTICA

Fonte: Vital *et al.*, 2017.

Dessa maneira, identificamos que a função de classificação arquivística está contida na organização do conhecimento e da informação, visando construir classes por meio da análise das funções, atividades e estrutura organizacional, ou seja, conhecimento contextual. Do mesmo modo, a descrição arquivística está contida na organização e representação da informação, fundamentando-se na

análise contextual e, posteriormente, o processo de representação das informações dos conjuntos documentais considerado relevante para seu entendimento e recuperação.

Faz-se necessário ressaltar a natureza interseccional do processo de representação em arquivos, em que as definições de conceitos da estrutura de classificação podem ser consideradas descritores e a descrição parte do pressuposto da classificação para cumprir seu objetivo (VITAL; MEDEIROS; BRASCHER, 2017). Ou seja, essas fronteiras não são rígidas, são porosas e demonstram a possibilidade de versatilidade desses processos, associando de modo transversal os conceitos de conhecimento e informação.

Observa-se que a classificação e a descrição arquivísticas são, de fato, as intervenções arquivísticas que são mais vinculadas aos processos de representação em arquivos no âmbito da OC. Diante dessas circunstâncias, pontuam-se algumas diferentes possibilidades que podem endossar as articulações metodológicas entre as áreas e prospectar outras oportunidades e perspectivas de análises e estudos.

Para além das funções de classificação e descrição arquivísticas, cabe-nos ressaltar que as demais intervenções delimitadas por Rousseau e Couture (1998) também são capazes de estabelecer suas interseções com a OC. As funções de produção e avaliação também reúnem características que as aproximam do debate da organização e representação do conhecimento e da informação, embora esses estudos ainda não tenham sido realizados.

A exploração dos diferentes SOCs e suas respectivas características também são importantes para que seja possível traçar um paralelo com as características de intervenções arquivísticas e seus respectivos instrumentos de gestão e descrição. Além disso, é possível desenvolver estudos que possam caracterizar e identificar o uso de SOCs no âmbito da prática da arquivologia. Outra oportunidade de

pesquisa e análise é a perspectiva de definição de diretrizes e procedimentos que possam auxiliar no processo de padronização estratégica e operacional para a organização e representação do conhecimento e da informação em arquivos, a fim de tornar factível a aplicação dessas metodologias no cenário arquivístico.

Quanto à organização do conhecimento, é possível desenvolver estudos que identifiquem quais são os SOCs utilizados no âmbito de cenários institucionais que possam revelar cenários contextuais que contribuam nas atividades de intervenções arquivística. Isso é possível, uma vez que toda intervenção arquivística, estratégica ou técnica, precede do reconhecimento e caracterização do contexto organizacional responsável pelo acúmulo dos documentos.

Infere-se, ainda, a relevância de ampliar as discussões quanto a análise de domínio no âmbito da Arquivologia como possibilidades de investigação. Isso se dá no que tange a alinhamentos metodológicos da caracterização do conhecimento de uma comunidade ou conjunto de atores discursivos.

Com o avanço das tecnologias e sua aplicabilidade no cenário dos arquivos, ampliam-se as possibilidades de aplicação e configuração da representação em arquivos quando se trata do uso de sistemas informatizados e plataformas de acesso aos documentos. No cenário tecnológico, intensificam-se os usos de SOCs no que tange aos mecanismos de produção, tramitação e controle dos documentos por conta da complexidade da estrutura digital.

## 4 CONSIDERAÇÕES FINAIS

A organização e representação do conhecimento e da informação aplicada a documentos de arquivo apresenta um potencial significativo de desenvolvimento de pesquisas, quanto a esses tipos de

documentos. A ampliação do diálogo entre as áreas de Arquivologia e Organização do Conhecimento se faz necessária e construtiva para que se tenham avanços no campo teórico e prático dessa intersecção.

Verificou-se que as funções de classificação e descrição arquivísticas são condicionantes na relação da Arquivologia com a organização e representação do conhecimento e da informação e integram, entre si, um primeiro nível (explícito) de relacionamento entre as áreas. Ademais, em um segundo nível (não-explícito), é possível desenvolver estudos em que as aplicações de OC sejam capazes de inferir demais relações multifacetadas da Arquivologia com a OC.

Identifica-se que há um reconhecimento crescente acerca da relevância de promover interfaces da Arquivologia em vista do desenvolvimento teórico-metodológico da OC, tendo em vista que o desenvolvimento de estudos em OC podem contribuir com a própria Arquivologia. Não se buscou uma limitação nos pontos de convergência entre OC e Arquivologia, mas sim o entendimento de que a ampliação de nossas perspectivas e a flexibilidade dos conceitos já consagrados podem nos levar a descobertas e redescobertas sobre a compreensão dessas concepções em um patamar de aplicação diferente.

## REFERÊNCIAS

AGUIAR, Francisco Lopes de; KOBASHI, Nair Yumiko. Organização e Representação do Conhecimento: Perspectivas de Interlocação Interdisciplinar entre Ciência da Informação e Arquivologia. *In: XIV Encontro Nacional de Pesquisa em Ciência da Informação, 2013, Florianópolis. Anais [...]*. Disponível em: <http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/155/147>. Acesso em: 30 jan. 2021.

BARITÉ, Mario. Organización del Conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. *In: CARRARA, K. (Org.). Educación, universidad e pesquisa. Marília, 2001. p.35-50.*

BARROS, Thiago Henrique Bragato. **A construção discursiva em arquivística: uma análise do percurso histórico e conceitual da disciplina por meio dos conceitos de classificação e descrição**. 2010. 132 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. São Paulo: Universidade Estadual Paulista, 2010. Disponível em: <http://hdl.handle.net/11449/93667>. Acesso em: 11 out. 2021.

BARROS, Thiago Henrique Bragato. **A representação da informação arquivística: uma análise do discurso teórico e institucional a partir dos contextos Espanhol, Canadense e Brasileiro**. 2014. 222 f. Dissertação (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências de Marília, Universidade Estadual Paulista, 2014. Disponível em: <http://hdl.handle.net/11449/110391>. Acesso em: 11 out. 2021.

BARROS, Thiago Henrique Bragato. A Indexação e a Arquivística: aproximações iniciais no universo teórico da organização e representação do conhecimento. *Encontros Bibli*, Florianópolis, v.21, n.46, p.33-44, 2016. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2016v-21n46p33/31600>. Acesso em: 30 jun. 2021

BARROS, Thiago Henrique Bragato.; SOUSA, Renato Tarcísio Barbosa. Organização do conhecimento e arquivologia: abordagens metodológicas. *Informação & Informação*, Londrina, v. 24, n. 2, p. 76-92, 2019. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/38290>. Acesso em: 10 nov. 2021.

BRÄSCHER, Marisa.; CAFÉ, Lúgia. Organização da informação ou organização do conhecimento?. *In: Encontro Nacional de Pesquisa em Ciência da Informação*, 9., 2008, São Paulo. **Anais [...]**. São Paulo: Universidade de São Paulo, 2008.

BRÄSCHER, Marisa; CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. *In: ROBREDO, J.; BRÄSCHER, M. (Orgs.). Passeios no Bosque da Informação: Estudos sobre Representação e Organização da Informação e do Conhecimento*. Brasília DF: IBICT, 2010. p. 147-176.

COOK, Terry. **What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift**. Ottawa: Archivaria, 1997.

DAHLBERG, Ingetraut. Current trends in knowledge organization. *In: GARCIA MARCO, F. J. (Org.). Organización del conocimiento em sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza, 1995. p. 7-25.

DAHLBERG, Ingetraut. Knowledge organization: its scope and possibilities. **Knowledge Organization**, v. 20, n. 4, p. 211-222, 1993.

ESTEBAN NAVARRO, Miguel Ángel. La representación y la organización del conocimiento en los archivos. *In: Actas del I Encuentro de Isko-España*, Zaragoza, 1993. **Anais** [...]. Zaragoza: Universidad Librería General. 1993.

FREITAS, Lidiane Marques. A organização do conhecimento sensível sob a ótica da Arquivologia: um estudo de caso na Embrapa Soja. 2017. 150 f. Dissertação (Mestrado em Ciência da Informação) – Curso de Ciência da Informação, Programa de Pós-graduação em Ciência da Informação, Universidade Estadual de Londrina. Londrina: Universidade Estadual de Londrina, 2017. Disponível em: <http://www.bibliotecadigital.uel.br/document/?code=-vtls000215310>. Acesso em: 10 abr. 2020.

GARCÍA MARCO, Francisco Javier. Paradigmas científicos en recuperación de información. *In: GARCÍA MARCO, F. J. (Ed.). Organización del conocimiento en sistemas de información y documentación*. Zaragoza: ISKO, 1995. p.99-112.

GUIMARÃES, José Augusto Chaves Guimarães. Abordagens teóricas de tratamento temático da informação: catalogação de assunto, indexação e análise documental. **Ibersid**: revista de sistemas de información y documentación, v. 3, p. 105-117, 2009.

HAGEN, Acácia Maria Maduro. Algumas considerações a partir do processo de padronização da descrição arquivística. **Ciência da Informação**, Brasília, v. 27, n. 3, 1998.

HJØRLAND, Birger. What is knowledge organization (KO)? **Royal School of Library and Information Science**, 6 Birketinget, Denmark, 2008. Disponível em: [https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/What\\_is\\_Knowledge\\_Organization.pdf](https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/What_is_Knowledge_Organization.pdf). Acesso em: 5 mar. 2020.

HODGE, Gail. **Systems of Knowledge Organization for Digital Libraries: beyond traditional authority files**. Washington, DC: The Council on Library and Information Resources, 2000. Disponível em: <http://www.clir.org/wp-content/uploads/sites/9/pub91.pdf>. Acesso em: 30 jan. 2019.

JENKINSON, Hilary. **A manual of Archive Administration**. Oxford: Oxford University Press, 1922.

LEHMKUHL, Camila Schwinden.; LINDEN, Leolíbia Luana. BARROS, Camila Monteiro; SILVA, Eva Cristina Leal da.; VITAL, Luciane Paula. Sistemas de organização do conhecimento e arquivologia: diálogos possíveis. **ISKO Brasil**, Londrina, v. 6, p. 58-66, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/123195>. Acesso em: 10 nov. 2021

LIMA, José Leonardo de Oliveira; ALVARES, Lilian. Organização e representação da informação e do conhecimento. *In: ALVARES, Lilian. (Org.)*.

**Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações.** São Paulo: B4 Editores, 2012. p. 21-47.

LOPES, Luis Carlos. A nova arquivística na modernização administrativa. 2. ed. Brasília: Projeto Editorial, 2009. 416p

MARTINS, Wanessa Rodrigues. **Representação arquivística:** o papel do CIA e do CONARQ para o desenvolvimento de políticas de descrição. 2019. 80 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-graduação em Ciência da Informação, Instituto de Ciências Sociais Aplicadas, Universidade Federal do Pará. Belém: Universidade Federal do Pará, 2019. Disponível em: <http://repositorio.ufpa.br/jspui/handle/2011/11427>. Acesso em: 10 out. 2021.

MARQUES, Angélica Alves da Cunha. Interlocuções entre a arquivologia nacional e a internacional no delineamento da disciplina no Brasil. 2011. 399 f. Tese (Doutorado em Ciência da Informação)-Universidade de Brasília, Brasília, 2011. Disponível em: <https://repositorio.unb.br/handle/10482/8730> Acesso em: 10 out. 2021.

RIBEIRO, Fernanda. A Arquivística como disciplina aplicada no campo da Ciência da Informação. *Perspectivas em Gestão & Conhecimento*, João Pessoa, v. 1, n. 1, p. 59-73, jan./jun. 2011. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/pgc/article/view/9887>. Acesso em: 28 nov. 2021.

ROUSSEAU, Jean.-Yves; COUTURE, Carol. **Os Fundamentos da Disciplina Arquivística.** Lisboa: Publicações D. Quixote, 1998. 356p.

SALES, Rodrigo de. O diálogo entre a organização do conhecimento e a ciência da informação na comunidade científica da ISKO-Brasil. *In: XIV Encontro Nacional de Pesquisa em Ciência da Informação*, 2015, João Pessoa. **Anais** [...]. João Pessoa: Universidade Federal da Paraíba, 2015.

SILVA, Andrieli Pachuda; TOGNOLI, Natalia Bolfarini.; GUIMARÃES, José Augusto Chaves Guimarães. Os valores éticos na organização e representação do conhecimento arquivístico. *Brazilian Journal of Information Science*, Marília, v. 11, n. 1, 2017. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/6678>. DOI:10.5016/brajis.v11i1.6678

SHELLENBERG, Theodore Roosevelt. R. **Arquivos modernos:** princípios e técnicas. Tradução de Nilza Teixeira Soares. Rio de Janeiro: Editora FGV, 2006.

SILVA, Eliezer Pires. A noção de informação arquivística. *In: RODRIGUES, G. M.; COSTA, M. G. (Orgs.). Arquivologia: configurações da pesquisa no Brasil: epistemologia, formação, preservação, uso e acesso.* Brasília: Editora Universidade de Brasília, 2012.

SILVA, M. A.; MOREIRA, Walter. Relações conceituais em ontologias e linguagens documentárias: análise de periódicos Qualis A1 de ciência da informação. *In: Seminário de Arquivologia e Biblioteconomia, 2015, São Paulo. Anais [...]*. São Paulo: Universidade Estadual de São Paulo, 2015.

STAPLETON, Richard. Jenkinson and Schellenberg: a comparison. *Archivaria*, Ottawa, n.17, p. 75-85, 1983.

SVENONIUS, Elaine. **The intellectual foundations of information organization**. Cambridge: The MIT Press, 2000. 255p.

TOGNOLI, Natalia Bolfarini. **A contribuição epistemológica canadense para a construção da Arquivística Contemporânea**. 2010. 120 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. Marília: Universidade Estadual Paulista, 2010.

TOGNOLI, Natalia Bolfarini. A representação na Arquivística contemporânea. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 5, n. 2, p. 79-92, 2012. Disponível em: <http://hdl.handle.net/11449/115396>. Acesso em: 30 jan. 2019.

TOGNOLI, Natalia Bolfarini; BARROS, Thiago Henrique Bragato. Os processos de representação do conhecimento arquivístico: elementos históricos e conceituais da classificação e descrição. **ISKO Brasil**, v. 3, p. 94-99, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/135201>. Acesso em: 10 nov. 2021.

VITAL, Luciane Paula, MEDEIROS, Graziela Martins de; BRASCHER, Marisa. Classificação e descrição arquivística como atividades de organização e representação da informação e do conhecimento. **Brazilian Journal of Information Science: Research Trends**, São Paulo, v. 11, n. 4, dez. 2017. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/7507>. Acesso em: 05 fev. 2019.

VITAL Luciane Paula; BRASCHER, Marisa. Descrição Arquivística: uma discussão conceitual. **Informação & Informação**, Londrina, v. 21, n. 1, p. 213-229, jun. 2016. ISSN 1981-8920. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/22497/18994>. Acesso em: 05 fev. 2019.

YAKEL, Elizabeth. Archival representation. **Archival Science**, Switzerland, v. 3, n. 1, p. 1-25. 2003.

## Sobre os organizadores

### **Thiago Henrique Bragato Barros**

Professor Adjunto no Departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS), professor permanente nos Programas de Pós-Graduação em Ciência da Informação da UFRGS e UFPA. Pós-doutorado em Ciência da Informação pela Universidade de Brasília (2018-2019). Pesquisador em grupos de pesquisa da UNB, UFPA, UFSCAR e UFRGS. Presidente eleito da International Society for Knowledge Organization (2020-2024). Deputy Editor da revista Knowledge Organization. Coordenador do Programa de Pós-Graduação em Ciência da Informação – PPGCIN UFRGS (2020-2022). Membro da Comissão Editorial da Em questão. Tem experiência na área de Arquivologia e Ciência da Informação, atuando principalmente nos seguintes temas: Abordagens linguísticas em Ciência da informação e Representação e Organização do Conhecimento.

### **Rita do Carmo Ferreira Laipet**

Professora Adjunta do Departamento de Ciências da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul (UFRGS). Doutora em Linguística Aplicada pela Universidade do Vale do Rio dos Sinos (UNISINOS). Mestre em Comunicação e Informação pelo Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul (UFRGS) e Bacharel em Biblioteconomia pela UFRGS. Chefe do Departamento de Ciências da Informação (DCI) da Faculdade de Biblioteconomia e Comunicação da UFRGS. Pesquisadora do Grupo de Pesquisa de Comunicação Científica da UFRGS com atuação na linha de Estudos em Organização do Conhecimento. Membro associado da ISKO-BRASIL (Sociedade Brasileira de Organização do Conhecimento). Líder do grupo de pesquisa Organização e Representação do conhecimento abordagens linguísticas em arquivos e bibliotecas - ORCALAB. Tem experiência na área de Ciência da Informação, com ênfase em Biblioteconomia nos seguintes temas: Representação do Conhecimento, Sistemas de Organização do Conhecimento (SKOS), Terminologia, Web Semântica, Ontologias, Folksonomia, Telecentros Comunitários, Inclusão Digital, Acesso à Informação. Os temas de pesquisas atuais envolvem o estudo e aplicação da Terminologia na Ciência da Informação, Teoria do Conceito, Teoria da Classificação Facetada, Relações Semânticas entre termos, Sistemas de Organização do Conhecimento (SKOS), Elaboração de Tesouros, Recuperação da Informação, Análise de Logs.

## Sobre os autores

### **Andréa Fontoura da Silva**

Mestranda em Ciência da Informação no Programa de Pós-Graduação em Ciência da Informação na UFRGS. Possui graduação em Arquivologia (2012) e Biblioteconomia (2000), também pela UFRGS. Participa do Grupo de Pesquisa Organização e Representação do Conhecimento - ORCALAB. É tutora desde 2020 do Curso EAD em Biblioteconomia da UFRGS pela Universidade Aberta do Brasil UAB, bibliotecária chefe na Província Franciscana do Rio Grande do Sul no Brasil, da Escola Superior de Teologia e Espiritualidade Franciscana - ESTEF e da Escola Técnica Cristo Redentor - ETCR. Tem experiência na área de gestão de bibliotecas, coordenando equipes e eventos. Coordena projetos de incentivo à leitura. Formanda em Letras Português/Inglês - UNICESUMAR.

### **Adriana Santos**

Mestranda em Ciência da Informação PPGCI (UFPA). Graduada em Letras/Português (UFPA), Especialista em Educação, Cultura e Organização Social (UFPA). Graduanda em Arquivologia (UFPA). Estagiária do setor de documentação do Arquivo Público do Estado do Pará. Membro do Grupo de Estudos: ORCALAB/UFRGS. Voluntária no Projeto de Extensão: Difundindo Arquivos, revelando memórias e narrando histórias (UFPA).

### **Brisa Pozzi de Sousa**

Concluiu, em 2019, o doutorado em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG). O mestrado em Ciência da Informação, pela Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP câmpus de Marília) em 2012 e, o bacharelado em Biblioteconomia, pela Universidade Federal do Espírito Santo (UFES), em 2004. Desde 2013 atua como docente nos cursos de bacharelado e licenciatura em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Pesquisa na Biblioteconomia e Ciência da Informação o contexto temático que envolve a Organização da Informação e a Organização do Conhecimento, destacando-se os fundamentos teóricos e práticos sobre a noção de assunto, a indexação, a classificação e a catalogação.

### **Bruna Carballo Dominguez de Almeida**

Mestranda no Programa de Pós-Graduação em Ciência da Informação pela Universidade Federal do Rio Grande do Sul. Especialista em Gestão em Arquivos pela Universidade Federal de Santa Maria (2017). Possui graduação em Arquivologia pela Universidade Federal do Rio Grande (2014). É Técnica Administrativa em Educação (TAE) na Universidade Federal do Rio Grande. Pesquisadora nos Grupos de Pesquisa: Organização e Representação do conhecimento abordagens linguísticas em arquivos e bibliotecas - ORCALAB e Arquivologia e Memória: documentos e identidade.

### **Camila Monteiro de Barros**

Doutora em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC), com período sanduíche na Université de Montréal (UdeM, Canadá), mestre em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC), graduada em Biblioteconomia pela Universidade do Estado de Santa Catarina (UDESC). Professora do Departamento de Ciência da Informação (CIN/UFSC), subcoordenadora do Programa de Pós-Graduação em Ciência da Informação (PGCIN/UFSC). Líder do Grupo de Pesquisa Representação e Organização do Conhecimento (ROC/UFSC). Membro do grupo de pesquisa Fundamentos Teóricos da Informação (FTI) da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Atua nas áreas de Organização e Representação do Conhecimento e da Informação, com ênfase teórico-metodológica nos sistemas de organização do conhecimento. Tem interesse nas questões relativas à informação musical.

### **Carine Melo Cogo Bastos**

Graduada em Arquivologia pela Universidade Federal de Santa Maria. Pós-graduada em Gestão Estratégica de Tecnologia da Informação pela Pontifícia Universidade Católica do RS, PUC/RS. Pós-graduada em Gestão de Arquivos pela Universidade Federal de Santa Maria, UFSM/RS. Mestre pelo Programa de Pós-graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul UFRGS/RS.

### **Daniel Martínez Ávila**

Professor assistente doutor da Universidad de León, na Espanha, e professor permanente do Programa de Pós-Graduação em Ciência da Informação (PP-GCI) da Universidade Estadual Paulista (UNESP). Possui graduação em Biblioteconomia y Documentación pela Universidad Carlos III de Madrid (2005),

licenciatura em Documentación pela Universidad Carlos III de Madrid (2007), mestrado em Investigación en Documentación pela Universidad Carlos III de Madrid (2009), e doutorado internacional pelo Programa Oficial de Doctorado en Documentación: Archivos y Bibliotecas en el Entorno Digital da Universidad Carlos III de Madrid (2012). Foi professor assistente doutor da Universidad Carlos III de Madrid até agosto de 2013 e desde setembro de 2019 até agosto de 2021, pesquisador e instrutor adjunto da University of Wisconsin-Milwaukee até 2014, professor assistente doutor da Universidade Estadual Paulista desde outubro de 2014 até agosto de 2019, e professor assistente doutor da Universidad Complutense de Madrid desde setembro de 2021 até fevereiro de 2022.

### **Donald Grant Campbell**

Phd in English from Queen's University (1990) Chair of Master of Library & Information Science MLIS Program. Associate Professor at Faculty of Information and Media Studies, Western University of Ontario.

### **Edgar Bisset Alvarez**

Doutor em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP), com período sanduíche na Universidad de Murcia/Espanha. Mestrado em Biblioteconomia e Ciência da Informação - Universidade da Havana (2012) e Graduação em Informação Científico-Técnica e Biblioteconomia - Universidade da Havana (2007). Atua como Professor nos cursos de Graduação e Pós-Graduação, do Departamento de Ciência da Informação no Centro de Ciências da Educação, Universidade Federal de Santa Catarina (UFSC), Coordenador do Programa de Pós-Graduação em Ciência da Informação da UFSC e Editor da Revista Encontros Bibli. Foi Professor Assistente na Faculdade de Comunicação Social e Especialista Principal na Direção de Informação da Universidade da Havana, Coordenador da Rede de Bibliotecas da Universidade da Havana. Pertence à Associação Cubana de Bibliotecarios (ASCUBI), também ao Grupo de Trabalho GT-8 Informação e Tecnologia da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). Participa como membro da Linha de Pesquisa 02: Informação, Comunicação Científica e Competência do PGCIN e pesquisador no grupo de pesquisa ITI-RG (Inteligência, Tecnologia e Informação - Research Group) da UFSC, e atua como pesquisador no Grupo de pesquisa das Novas Tecnologias da Informação, da Linha de Informação e Tecnologia do Programa de Pós-Graduação de Ciência da Informação da Universidade Estadual Paulista "Hyginio Mesquita Filho" UNESP/Marília-SP.

### **Fabio Assis Pinho**

Professor na Universidade Federal de Pernambuco (UFPE) atuando nos cursos de graduação e pós-graduação. Possui graduação em Biblioteconomia e Ciência da Informação (2002) pela UFSCar, Mestrado (2006) e Doutorado (2010) em Ciência da Informação pela UNESP. É bolsista de produtividade do CNPq (PQ2) e coordenador adjunto da área de Comunicação e Informação na CAPES (2018-2022). Tem interesse nos temas de pesquisa de Organização e Representação do Conhecimento e Memória da Informação Científica e Tecnológica.

### **Isabel Cristina Pereira dos Santos**

Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação (PPGCIN) da Universidade Federal do Rio Grande do Sul (UFRGS). Bacharela em Letras pela UFRGS. Servidora técnico-administrativa lotada no Programa de Pós-Graduação em Controladoria e Contabilidade da Faculdade de Ciências Econômicas da UFRGS. Tem experiência na área de Letras, com ênfase em Revisão Textual, atuando principalmente em temas relacionados a Ciências Econômicas e Ciências Contábeis.

### **Isadora Victorino Evangelista**

Possui graduação em Biblioteconomia pela Universidade Estadual Paulista – UNESP, Mestrado e Doutorado em Ciência da Informação pela mesma universidade. Fez parte de sua pesquisa de mestrado na Universidade de Coimbra, sob supervisão da Dra. Maria da Graça de Melo Simões, investigando sobre os aspectos éticos de características da Indexação como a exaustividade e especificidade. Parte de sua pesquisa de doutorado foi desenvolvida na Western Ontario University, em London/Canadá, investigando sobre a comunidade epistêmica e as revoluções científicas ocorridas no âmbito das pesquisas éticas em Organização do Conhecimento. Atualmente atua como Bibliotecária na Universidade Federal de São Carlos, campus Lagoa do Sino. Possui como interesse de pesquisa Ética em Organização do Conhecimento, Sociologia do Conhecimento e Comunidades Epistêmicas.

### **Joana Peregrina Hernandes**

Possui graduação em Arquivologia pela Universidade Federal do Rio Grande do Sul (2011) e mestranda em Ciência da Informação pela Universidade Federal do Rio grande do Sul (2020). Tem experiência na área de Gerenciamento Eletrônico de Documentos, Gestão de Arquivos e atualmente complementando estudos acadêmicos na área da Ciência da Informação: com ênfase em Arquivologia, atuando principalmente no seguinte tema: Organização do conhecimento e estudos de usuários.

### **José Augusto Chaves Guimarães**

Bacharel em Biblioteconomia (UNESP, 1981) e em Direito (Univem, 1981), Mestre em Ciência da Comunicação (USP, 1989) e Doutor em Ciência da Informação (USP, 1994). Realizou estágio pós-doutoral em Documentação na Universidad Carlos III de Madrid (2009). É Professor Titular aposentado da UNESP-Marília, onde ainda atua como Docente Permanente do Programa de Pós-graduação em Ciência da Informação. Foi fundador do capítulo Brasileiro da International Society for Knowledge Organization (ISKO-Brasil), do qual foi presidente por três gestões. Foi membro do Executive Board da ISKO e organizou a 14th ISKO International Conference (Rio de Janeiro, 2016). Foi presidente da Associação Brasileira de Ensino de Biblioteconomia - ABEED e da Asociación de Educadores e Investigadores en Biblioteconomía, Documentación y Ciencias de la Información de Iberoamérica y el Caribe - EDICIC na área de Documentação. Atualmente atua como membro titular do Comitê Assessor de Artes, Ciência da Informação e Comunicação do CNPq. Atua também como membro de comitê científico de várias revistas e eventos na área de Ciência da Informação no Brasil e no exterior assim como parecerista ad hoc de revistas evento e agências de fomento no Brasil e no exterior. Suas áreas de pesquisa são: organização do conhecimento, análise de domínio, ética informacional, e produção científica.

### **Leolíbia Luana Linden**

Professora Assistente do Curso de Arquivologia no Departamento de Ciências da Informação (DCI) da Faculdade de Biblioteconomia e Comunicação (FABICO) da Universidade Federal do Rio Grande do Sul (UFRGS). Arquivista formada pela Universidade Federal de Santa Catarina (UFSC). Mestre e Doutora em Ciência da Informação pelo Programa de Pós Graduação em Ciência da Informação (PGCIN) na UFSC. Possui experiência na área de Ciência da Informação com ênfase em Arquivologia. Tem interesse nos temas relacionados à Organização e Representação da Informação, Políticas e Programas de Gestão de Documentos.

### **Luís M. Machado**

Doutorando e Mestre (2016) em Ciência da Informação na Universidade de Coimbra, graduado (2001) em Ensino pelo Instituto Politécnico de Lisboa. Investigador colaborador no Centro de Estudos Interdisciplinares do Século XX da Universidade de Coimbra (2018), no Grupo de Pesquisa Representação do Conhecimento, Ontologias e Línguas, da Universidade Federal de Minas Gerais (2020) e no Grupo de Pesquisa Fundamentos Teóricos da Informação da Universidade Estadual Paulista (2021). Editor convidado na Multidisciplinary Digital Publishing Institute (2021).

### **Luciana Monteiro-Krebs**

Doutora em Ciências Sociais (KU Leuven, Bélgica) e Doutora em Comunicação e Informação (UFRGS, Brasil). Luciana é mestre em Linguística Aplicada e bacharel em Biblioteconomia, com experiência internacional em pesquisa Universidade do Porto – Portugal (2010-2011) e na KU Leuven – Bélgica (2018-2022). No Meaningful Interactions Lab - Mintlab (KU Leuven), Luciana trabalha em projetos de pesquisa em Interação Humano-Computador, como o Algorithmic Mediation in Academic Social Systems (AMASS) e o Algorithmic Transparency & Accountability in Practice (ATAP), pesquisando os efeitos dos algoritmos de recomendação na mídia a partir de uma abordagem centrada no usuário. Luciana é membro dos grupos de pesquisa Organização e Representação do conhecimento - abordagens linguísticas em arquivos e bibliotecas (ORCALAB) do PPGCIN/UFRGS, e Comunicação Científica (GP ComCient) do PPGCOM/UFRGS. Seus tópicos de interesse são algoritmos de recomendação, comunicação científica, padrões de uso em mídias sociais online, folksonomias, ontologias, análise de logs, recuperação da informação e terminologias das áreas jurídica e saúde.

### **Lucas Andrade Sá Corrêa**

Doutorando em Ciência da Informação pelo PPGCI/UFF, Mestre em História, Política e Bens Culturais pelo Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC/FGV) (2014) e Bacharel em Ciências Sociais (CPDOC/FGV) (2011). Graduando em Arquivologia (UFF). É integrante do Grupo de Pesquisa Arquivos, Bibliotecas e Organização do Conhecimento (GP-ABOC).

### **Marcos Oliveira da Costa**

Graduado em Biblioteconomia - UFPA, Especialista em Neuropsicopedagogia – (UNICID), Mestrando do Programa de Pós-graduação em Ciência da Informação (UFPA). Integrante do Grupo de Pesquisa ORCALAB (UFRGS). Atualmente é Bolsista Capes e desenvolve pesquisas de cunho teórico - metodológico e tem como áreas de interesse: Estudos Linguísticos no contexto da Ciência da Informação e da Organização do Conhecimento; Epistemologia da Ciência da Informação e da Organização do Conhecimento (OC); Tecnologias da Informação e Comunicação; Problemática de Acesso; Inclusão e Exclusão educativa; Teoria Crítica da CI.

### **Mario Barité**

Es Doctor y Máster en Información Científica por la Universidad de Granada, España. Asimismo, cuenta con un Diploma de Estudios Avanzados de la Universidad de Granada, España. Es Licenciado en Bibliotecología y Escribano Público por la Universidad de la República del Uruguay. Actualmente reviste como Profesor Titular efectivo, en régimen de dedicación total, del Instituto de Información de la Facultad de Información y Comunicación, Universidad de la República, con docencia de grado y posgrado, e investigación en las áreas de organización del conocimiento, terminología y gobierno electrónico. Es responsable del grupo de investigación "Terminología y Organización del Conocimiento" (GTERM), que cuenta con miembros provenientes de la ciencia de la información, la comunicación, la lingüística y la traducción. Fue Pro-Rector de Extensión y Relaciones con el Medio de la Universidad de la República. Autor y/o coordinador de una quincena de libros ha realizado alrededor de un centenar de contribuciones en forma de capítulos de libros, ponencias a congresos y artículos en revistas especializadas de varios países. Es editor responsable de la revista académica *Informatio*.

### **Natália Bolfarini Tognoli**

Bacharel em Arquivologia, Mestre e Doutora em Ciência da Informação. É professora adjunta no departamento de Ciência da Informação e no Programa de Pós-graduação em Ciência da Informação da Universidade Federal Fluminense (UFF). É líder do Grupo de Pesquisa Arquivos, Bibliotecas e Organização do Conhecimento (GP-ABOC). Presidente do Capítulo Brasileiro da International Society for Knowledge Organization (2020-2021; 2022-).

### **Rafael Port da Rocha**

Possui graduação em Ciências de Computação pela Universidade Federal do Rio Grande do Sul (1988), mestrado em Computação pela Universidade Federal do Rio Grande do Sul (1992) e doutorado em Computação pela Universidade Federal do Rio Grande do Sul (2000). Atualmente é professor associado do Departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul. Tem experiência nas áreas de Ciência da Informação e Ciência da Computação, atuando principalmente nos seguintes temas: metadados, banco de dados, ontologias, preservação digital, gestão de dados da pesquisa. É fundador e atualmente coordena o Centro de Documentação e Acervo Digital da Pesquisa (CEDAP), órgão auxiliar da FABICO/UFRGS que busca dar suporte para pesquisas científicas, tecnológicas, artísticas e culturais realizada na Universidade, visando à gestão, preservação e uso científico e cultural de seus ativos digitais de pesquisa, natos ou decorrentes da digitalização.

### **Rita do Carmo Ferreira Laipelt**

Professora Adjunta do Departamento de Ciências da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul (UFRGS). Doutora em Linguística Aplicada pela Universidade do Vale do Rio dos Sinos (UNISINOS). Mestre em Comunicação e Informação pelo Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul (UFRGS) e Bacharel em Biblioteconomia pela UFRGS. Chefe do Departamento de Ciências da Informação (DCI) da Faculdade de Biblioteconomia e Comunicação da UFRGS. Pesquisadora do Grupo de Pesquisa de Comunicação Científica da UFRGS com atuação na linha de Estudos em Organização do Conhecimento. Membro associado da ISKO-BRASIL (Sociedade Brasileira de Organização do Conhecimento). Líder do grupo de pesquisa Organização e Representação do conhecimento abordagens linguísticas em arquivos e bibliotecas - ORCALAB. Tem experiência na área de Ciência da Informação, com ênfase em Biblioteconomia nos seguintes temas: Representação do Conhecimento, Sistemas de Organização do Conhecimento (SKOS), Terminologia, Web Semântica, Ontologias, Folksonomia, Telecentros Comunitários, Inclusão Digital, Acesso à Informação. Os temas de pesquisas atuais envolvem o estudo e aplicação da Terminologia na Ciência da Informação, Teoria do Conceito, Teoria da Classificação Facetada, Relações Semânticas entre termos, Sistemas de Organização do Conhecimento (SKOS), Elaboração de Tesouros, Recuperação da Informação, Análise de Logs.

### **Renato Tarciso Barbosa de Sousa**

Graduação em História pela Universidade de Brasília (1990), especialização em Organização de Arquivos pela Universidade de São Paulo (1992), mestrado em Biblioteconomia e Documentação pela Universidade de Brasília (1995) e doutorado em História Social pela Universidade de São Paulo (2005). Atualmente, é Professor Associado IV do Curso de Graduação em Arquivologia e do Programa de Pós-Graduação em Ciência da Informação, da Faculdade de Ciência da Informação, da Universidade de Brasília. É diretor da Faculdade de Ciência da Informação no quadriênio 2018-2022. Vice-líder do Grupo de Pesquisa Estudos de Representação e Organização da Informação e do Conhecimento. Tem experiência na área de Ciência da Informação, com ênfase em Organização de Arquivos, atuando principalmente nos seguintes temas: arquivologia, organização e recuperação de documentos de arquivo, gestão de documentos e políticas públicas de arquivo.

### **Rai Costa**

Possui graduação em Arquivologia pela Universidade Federal do Pará (2018). Especialista em Gestão Pública, pelo Centro de Estudos de Especialização e Extensão (2022), Mestrando em Ciência da Informação, PPGCI/UFGA. Integrante do Grupo de Pesquisa Orcalab (UFRGS). Atualmente é Arquivista da Universidade Federal do Pará.

### **Thiago Henrique Bragato Barros**

Professor Adjunto no Departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul (UFRGS), professor permanente nos Programas de Pós-Graduação em Ciência da Informação da UFRGS e UFPA. Pós-doutorado em Ciência da Informação pela Universidade de Brasília (2018-2019). Pesquisador em grupos de pesquisa da UNB, UFPA, UFSCAR e UFRGS. Presidente da International Society for Knowledge Organization (2022-2026). Deputy Editor da revista Knowledge Organization. Coordenador do Programa de Pós-Graduação em Ciência da Informação - PPGCIN UFRGS (2020-2022). Membro da Comissão Editorial da Emquestão. Tem experiência na área de Arquivologia e Ciência da Informação, atuando principalmente nos seguintes temas: Abordagens linguísticas em Ciência da informação e Representação e Organização do Conhecimento.

### **Talita Morgana Arruda Tavares**

Graduada em Jornalismo pela UNICAP (2009), Especialista em Marketing e Assessoria Política pela USURP (2011). Bacharel em Biblioteconomia pela UFPE (2021). Exerceu atividades de docência, particularmente de professora tutora.

### **Wilson Roberto Veronez Junior**

Doutorando e Representante Discente do Programa de Pós-Graduação em Ciência da Informação (PPGCI-UNESP 2021-). Membro do Grupo de Pesquisa: Fundamentos Teóricos da Informação (GPFTI). Membro do Conselho do Programa de Pós-graduação em Ciência da Informação (PPGCI/UNESP) e da Comissão Gestora de Recursos PROEX. Especialização em Processos Didáticos-Pedagógicos para Cursos na Modalidade à Distância na UNIVESP (2019-2021). Atuou como facilitador de ensino à distância nos cursos de Gestão Pública, Engenharia de Computação, Engenharia de Produção e Pedagogia da UNIVESP (2019-2021). Realizou consultoria arquivística na empresa CAP Arquitetura & Construção LTDA (2019). Mestre em Ciência da Informação pelo Programa de Pós-graduação em Ciência da Informação (PPGCI-UNESP

2019-2021). Graduação interrompida em Biblioteconomia pela Unesp de Marília (2017-2019). Bacharel em Arquivologia pela Unesp de Marília (2012-2017). Tem como interesse de pesquisa: Arquivologia Científica, Arquivo Histórico do Partido dos Trabalhadores, Classificação Arquivística, Diplomática Arquivística, Epistemologia Social, Gestão de Arquivos, Linguagem Documentária, Organização e Representação da Informação e do Conhecimento Arquivístico.

[www.pimentacultural.com](http://www.pimentacultural.com)

# Organização e Representação do Conhecimento em Múltiplas Abordagens