

Do tratamento à

Organização da **Informação**

Reflexões sobre concepções,
perspectivas e tendências

Maria Graça Simões
Gercina Ângela de Lima
COORDENAÇÃO

IMPrensa DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

Esta obra apresenta uma reflexão teórica e histórica sobre as noções clássicas de tratamento da informação e de organização da informação, com base na esfera epistemológica da organização do conhecimento.

É Composta de oito capítulos que discorrem sobre (1) as variações terminológicas produzidas na Ciência e no Ensino Brasileiro e Português, (2) a discussão de aspectos teórico-conceituais da organização e representação da informação e a sua correlação com os aspectos práticos que envolvem os ambientes informacionais digitais; (3) o delineamento conceitual inerente à representação da informação feito a partir da discussão da catalogação como confluência de aspectos descritivos, temáticos, contextuais e de acesso à informação de modo que a interconexão entre eles esteja assegurada; (4) as diferentes perspectivas teóricas e metodológicas sobre a modelagem de domínios aplicados à elaboração de ferramentas de representação e recuperação de informação, no contexto das redes eletrônicas de informação e comunicação; (5) a identificação e exploração dos aspectos da variação terminológica na produção científica sobre Organização da Informação relativos aos termos Análise Documentária e Análise de Assunto; (6) as aplicações de representação de recurso informacional, os contextos e processos onde as inovações derivadas da ciência da computação, da inteligência artificial, da área de processamento de linguagem natural e de novas plataformas de bancos de dados têm transformado o contexto de produção de documentos, de sistemas de organização do conhecimento, de métodos de exploração semântica e de visualização.



I N V E S T I G A Ç Ã O



EDIÇÃO

Imprensa da Universidade de Coimbra
Email: imprensa@uc.pt
URL: http://www.uc.pt/imprensa_uc
Vendas online: <http://livrariadaimprensa.uc.pt>

COORDENAÇÃO EDITORIAL

Imprensa da Universidade de Coimbra

INFOGRAFIA

Margarida Albino

REVISÃO

Sónia Gabriel

EXECUÇÃO GRÁFICA

KDP – Kindle Direct Publishing

ISBN

978-989-26-1943-9

ISBN DIGITAL

978-989-26-1944-6

DOI

<https://doi.org/10.14195/978-989-26-1944-6>

Do tratamento à

Organização da **Informação**

Reflexões sobre concepções,
perspectivas e tendências

Maria Graça Simões
Gercina Ângela de Lima

COORDENAÇÃO

Benildes Coura Moreira dos Santos Maculan
Célia da Consolação Dias

ORGANIZAÇÃO

IMPrensa DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

DEDICATÓRIA

À memória da professora Maria da Graça de Melo Simões dedicamos este livro, que inspirou, encorajou e se empenhou na realização dele, mesmo passando por momentos tão desafiadores. Um exemplo a seguir na vida e na academia!

EPÍGRAFE

*Quanto mais nos elevamos, menores parecemos aos
olhos daqueles que não sabem voar (Friedrich Nietzsche).*

SUMÁRIO

Apresentação.....	11
Prefácio.....	17
CAPÍTULO 1 – Reflexões sobre os conceitos de Tratamento da Informação e de Organização da Informação	23
CAPÍTULO 2 – Tratamento Temático da Informação a partir dos trabalhos publicados nos anais do capítulo brasileiro de ISKO: pontos de partida, identidade nacional e agentes epistêmicos.....	49
CAPÍTULO 3 – Catalogação e Metadados: reflexões sobre concepções, perspectivas e tendências	83
CAPÍTULO 4 – Perspectivas de modelagem conceitual na elaboração de sistemas de organização do conhecimento: subsídios para o ensino	119
CAPÍTULO 5 – A variação terminológica em Organização da Informação: Análise de Assunto e Análise Documentária ..	145
CAPÍTULO 6 – Subsídios teórico-metodológicos para a construção de vocabulários controlados	193

CAPÍTULO 7 – Aplicações tecnológicas em Representação e Organização da Informação.....	243
CAPÍTULO 8 – Da organização à descoberta do conhecimento: novos caminhos para a recuperação da informação ...	267

APRESENTAÇÃO

A motivação para a elaboração do livro **“Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências”** surgiu a partir do encontro das Professoras Gercina Lima, Benildes Maculan e Célia Dias com a Professora Graça Simões, no III Congresso ISKO Espanha-Portugal / XIII Congresso ISKO Espanha, que teve lugar na Universidade de Coimbra, em novembro de 2017. Ali, na Sala dos Professores da Faculdade de Letras, no meio de todo o burburinho que animou, durante vários dias, os dois congressos irmãos da ISKO e EDICIC, nascia mais um projeto em parceria. Desta vez, a proposta dada à luz foi a produção de um livro idealizado para colmatar uma lacuna observável na literatura científica de ambos os países. E foi com muito entusiasmo que assim surgiu o primeiro esboço do que agora, quase três anos depois, vem a ser a obra final.

Além do interesse comum em consolidar uma parceria entre a UFMG e a UC, verificou-se, sobretudo, a necessidade de reunir reflexões em torno do Tratamento da Informação e da Organização da Informação, áreas tradicionalmente suscetíveis a alguma flutuação terminológica e fluidez conceitual. Para tal, propôs-se estudar a univocidade e o estabelecimento de uma relação entre o construto conceitual, a prática e as tendências de pesquisa expressas na literatura internacional especializada. Aos autores que participaram neste desafio coletivo sugeriu-se que explorassem convergências e/ou

divergências de sentidos na ocorrência daqueles termos em situação discursiva, bem como o resgate teórico-metodológico das temáticas e das suas perspectivas e tendências de estudo e de aplicação. Assim, direcionados por estas questões, apresentaram os seus estudos que reúnem em uma só obra as vertentes teóricas, históricas e culturais da área de Biblioteconomia e Ciência da Informação (BCI).

Este projeto começou a tomar corpo logo após o final do III Congresso ISKO Espanha-Portugal, tendo como coordenadoras as Professoras Graça Simões e Gercina Lima, e como organizadoras as Professoras Benildes Maculan e Célia Dias. A Professora Graça Simões, com as suas raras capacidades de coesão e de agregação de pessoas em torno de objetivos comuns, encarregou-se de contactar os colegas portugueses e solicitar um resumo das ideias principais a desenvolver nos vários capítulos. Procedimento homólogo realizou a Professora Gercina Lima, no Brasil. Os resultados são os que se verificam na obra presentemente concluída.

Num tom mais pessoal.

O tradicional e o digital, a organização e a recuperação da informação, as divergências e as convergências, os processos e os produtos, os princípios e os fins... foram estas, certamente, as vertentes de análise que funcionaram como traves-mestras no conjunto da obra que nos legou a Graça nos seus últimos e excepcionalmente produtivos dez anos de vida, dedicados ao trabalho e à investigação, aspectos que com a sua vida pessoal se misturavam e se completavam.

O livro que apresentamos foi muito acarinhado pela Graça, possivelmente porque representava todo um alinhamento e esforço comum, convergente e direccionado num sentido muito claro: con-

tribuir para dotar a área disciplinar, ou se preferirmos, o domínio científico, de um corpo de conhecimentos coerente e consistente, teoricamente forte e, como complementar, certamente, a Graça, “inovador”.

Este projeto em que a Graça se envolveu, e que agora se dá à estampa, perto de um ano depois do seu desaparecimento, é uma derradeira demonstração do seu espírito e do sentido de missão que abraçou, e que mesmo no último ano de vida não deixou que esmorecessem.

Na impossibilidade de a lembrarmos na homenagem que pré- víamos prestar neste primeiro ano sem a sua presença, e que guardaremos para quando nos pudermos reunir e, dessa forma, revisitar integralmente as diferentes facetas que a lembrança da vida e do trabalho da Graça convoca, aproveitamos este espaço que seria da Graça – e que a Gercina generosamente nos cedeu – para destacar duas dessas facetas que este livro também e tão bem ilustra.

Em primeiro lugar, a ideia de diálogo, de colóquio. Desde há muito, a Graça esteve envolvida no desenvolvimento da cooperação académica internacional. A sua passagem pela coordenação da mobilidade docente e discente da Secção de Informação da Faculdade de Letras permitiu continuar de forma muito viva uma prática que a Secção assumiu como prioritária: a cooperação e o diálogo com escolas de outros países, em particular da Espanha e do Brasil. Esta atividade tem-se revelado crucial para o desenvolvimento da área da Ciência da Informação, e foi preocupação da Graça que se manifestasse em diversas missões de intercâmbio que a fizeram até pisar solo africano. Mas, mesmo depois de ter terminado o seu mandato, muitas vezes por sua iniciativa individual continuou a contribuir para que as sementes de um trabalho de anos se não perdessem e, pelo contrário, se desenvolvessem e frutificassem.

Este livro é um testemunho e resultado dessa preocupação e dessa iniciativa, que a Graça também foi desenvolvendo no seu trabalho no braço ibérico da ISKO, e cujo valor pode também aqui ser publicamente apreciado. Ainda que já sem a presença da Graça, comungamos desta ideia de comunidade em colóquio, geradora de discussão e diálogo entre as diferentes faces da área de saber.

Em segundo lugar, a de que estas pontes não se lançam nem sobrevivem sem afetos. Como a conhecemos, nada do que era acadêmico, científico, ou profissional na Graça, poderia deixar de envolver uma forte componente de afeto ou cumplicidade pessoal, que nunca quis que se confundisse ou sequer se contaminasse pelo amiguismo. A Graça não pensava a Ciência sem os seus agentes, e via entre eles a necessidade do estabelecimento de um *pathos* de fortes e, por vezes, exigentes dinâmicas, que alimentava tanto quanto se alimentava do que era e é sobretudo *logos*. Este reconhecimento, que não é apenas nosso, vemo-lo igualmente na voz dos seus estudantes, dos seus colegas, das suas colaborações, daqueles e daquelas que com ela privaram de alguma forma nos meios profissionais em que se moveu e que a moveram.

Difícilmente um livro poderia ser mais clara expressão de símbolo e legado do generoso espírito que combina aquelas facetas do que este que Gercina e Graça conceberam e dirigiram, e que a Gercina por ambas, mas também por nós todos, agora conclui e apresenta. E que nos resta agradecer e ler, com saudades do passado, mas com iguais vontades do futuro: trata-se, como em Ciência deve acontecer em simultâneo, de um ponto de chegada e de um novo ponto de partida que assenta na memória dos caminhos percorridos e que desenha, com esses mesmos passos, aqueles onde havemos de pisar.

António Tavares Lopes
Universidade de Coimbra (UC)
Faculdade de Letras/CEIS20
atlopes@fl.uc.pt
ORCID: 0000-0002-8876-5387

Maria Cristina Vieira de Freitas
Universidade de Coimbra (UC)
Faculdade de Letras/CEIS20
cristina.freitas@fl.uc.pt
ORCID: 0000-0002-8849-8792

Maria Manuel Borges
Universidade de Coimbra (UC)
Faculdade de Letras/CEIS20
mmb@fl.uc.pt
ORCID: 0000-0002-7755-6168

PREFÁCIO

A coletânea que agora se oferece ao público resulta de uma iniciativa das professoras Graça Simões, Gercina Lima, Benildes Maculan e Célia Dias, dentro de um convênio com a Universidade de Coimbra. Como de costume, o volume aborda um arco abrangente que inclui reflexões e ações do ato de representar a unidade de informação até sua integração em grandes coleções/bases de dados, questões de interoperabilidade, e organização/estruturação dos próprios documentos com vista à descoberta do conhecimento via ontologias ou via ferramentas tecnológicas providas pela computação. Não deixam de ser abordados, nesse arco, os instrumentos de natureza semântica de mediação como cabeçalhos de assunto, tesouros e taxonomias. Também a Tecnologia da Informação (TI) tem sido chamada a resolver questões de tratamento, representação e recuperação que a Biblioteconomia enfrentava com as limitações tecnológicas da época, hoje integrando a Ciência da Informação (BCI). O painel mostra as primeiras contribuições teóricas, cujos princípios já vinham sendo estabelecidos desde o início do século XX, mas voltados para a organização de grandes acervos nas bibliotecas, resultando na criação de esquemas gerais de classificação para livros. Acervos especializados configuram outro ambiente de trabalho, com natural afastamento das questões bibliotecárias com seus acervos gerais. Mas contribuem, de alguma forma, para a representação de domínios, incorporando contribuições dos ontologistas.

Como afirma o Professor Shera, Ciência da Informação é a Biblioteconomia uma oitava acima.

O amadurecimento desse novo espaço de conhecimento se consolida como área científica.

Como seria de esperar, diferentes áreas têm contribuído para o desenvolvimento da CI como uma área científica. Em um processo simbiótico com a Ciência da Informação, a TI vem contribuindo, de um lado, para a resolução dos problemas identificados na BCI e, de outro, durante o processo, a TI identifica novas questões que, por sua vez, requerem adoção de princípios e técnicas da CI produzindo-se um círculo virtuoso. Visto por um outro ângulo, funcionam como uma moeda: cara e coroa. Não faltam, assim, temas de estudo para a comunidade acadêmica. De modo bastante sucinto, pode-se afirmar que as pesquisas têm se concentrado então no desenvolvimento de princípios e padrões para os instrumentos de descrição das unidades de informação – objetos de natureza diversa – através de metadados e, por consequência, na produção de taxonomias para tais metadados; e ainda na organização do conhecimento com seus modelos de representação.

É neste contexto que se apresentam os estudos incluídos neste volume.

A presente Coletânea reúne artigos de revisão de literatura nos mais diversos aspectos da BCI, oferecendo rico material de referência para outros estudos e, por vezes, trazendo inquietações que podem levar a novos estudos.

O capítulo 1 da Coletânea evidencia, desde logo, que o norte dos estudos se estabelece na separação das atividades de organização de informação e de organização do conhecimento. Os autores adotam a proposta de Brascher e Café, caracterizando a organização da informação como ‘ocorrências individuais de objetos informacionais’, depreendendo-se do próprio texto do artigo, que o conceito se estabelece no contexto das bibliotecas e catálogos, com o registro

descritivo de cada item de informação, inclusive nos repertórios bibliográficos. Nesse momento, os autores estabelecem limite com o processo de organização do conhecimento, que estaria restrito à construção de modelos de mundo.

No capítulo 2, os autores abordam o tema ‘tratamento temático da informação’ nos anais da ISKO brasileira, analisando diversas definições de reconhecidos pesquisadores internacionais e sugerindo, a partir destes elementos, maior influência francesa no desenvolvimento da pesquisa brasileira.

No capítulo 3, os autores buscam conceptualizar os conceitos metadado e catalogação e suas relações, sugerindo que são fenômenos complementares, interdependentes. O uso de metadados na catalogação consolida a representação simbólica de um objeto remetendo-o a ele de modo simbólico. Tais desenvolvimentos na catalogação trazem, para esta atividade, o entendimento de uma área que, além de seus aspectos técnicos, também requer procedimentos intelectuais.

O capítulo 4 visa a incorporação, na formação do profissional, de conteúdos de ensino para Organização do Conhecimento. As autoras apresentam um painel histórico e conceitual para, ao final, apresentar conteúdos que consideram aplicáveis a este novo campo de conhecimento, se assim se pode afirmar. Na base da organização do conhecimento estaria a classificação cujos princípios remontam a Aristóteles com suas categorias, mas há necessidade de incluir mecanismos de representação baseados em lógica para derivar conhecimento, como buscam as ontologias.

No capítulo 5, os autores estudam variantes terminológicas como Análise de Assunto e Análise Documentária e seu conteúdo, nos trabalhos da ISKO de Espanha, Portugal e Brasil. E para tanto o fazem a partir das definições, objetos, princípios teóricos e metodológicos encontrados nas comunicações do período estudado. A pesquisa evidencia a impropriedade de considerar que ambas

expressões possam se referir a uma mesma forma de pensar e atuar.

Enquanto os estudos anteriores se debruçam sobre processos, os autores fazem, no capítulo 6, extensa descrição dos instrumentos de apoio aos diversos processos envolvidos na organização da informação. Com base nas normas internacionais são apresentadas as bases para organizar esquemas de classificação, taxonomias, listas de cabeçalhos de assunto, ontologias e terminologias. Neste capítulo cada instrumento está descrito em separado, mas é possível observar aspectos comuns em algumas normas. Estas ressaltam, ainda, relevância dos sistemas facetados aplicáveis a objetos, e-comércio, modelagem de objetos computacionais, dada a facilidade de organizar informações a partir de diferentes facetas.

O capítulo 7 aborda as questões tecnológicas que perpassam as atividades e instrumentos desenvolvidos para a descrição e representação da informação, descrevendo os principais artefatos computacionais relacionados. Os autores citam, de início, suas apreciações a partir da tecnologia de base, conhecida como Web Semântica, que promoveu subsídios para a estrutura semântica e sintática dos dados. Em relação ao impacto destas tecnologias no universo bibliotecário, citam-se a restrição de padrões para o universo bibliográfico e a necessidade de ‘repensar e reestruturar as práticas e instrumentos de representação’ que precisam lidar com novas tipologias de documento, de suporte, de navegação e compartilhamento com outros serviços voltados para o ambiente cultural como arquivos e museus.

A Coletânea termina com o capítulo 8 com o tema da descoberta do conhecimento, complementando, de certo modo, o texto anterior, visto que exige repensar as práticas adotadas até então para a recuperação da informação. Os autores fazem uma ampla busca exploratória na literatura, relatando os ‘principais conceitos subjacentes aos mecanismos dos sistemas de descoberta, os quais

se enquadram nos propósitos da Recuperação da Informação'. O tema é considerado por uns como pertencendo à BCI, outros contestam porque isso envolveria perspectivas disciplinares diferentes. Organização do conhecimento não pertenceria ao domínio exclusivo da BCI porque o conhecimento é classificado e organizado em várias áreas como Biologia, Sociologia, Filosofia, 'obedecendo a princípios teóricos consensuais ou não.' No entanto, o percurso histórico feito pelos autores revela 'relativa maturidade como campo e como tema de pesquisa, e torna-se uma área de toque da Ciência da Informação, contribuindo, inclusivamente, para o seu desenvolvimento científico'.

A presente Coletânea nos oferece então um panorama amplo das questões que envolvem, em especial, organização da informação, não descurando aspectos ligados a organização do conhecimento e a seus meios de apoio àquelas atividades. O ambiente em que tais questões ocorrem é o ambiente das bibliotecas e, como consequência, contempla documentos de natureza científica e cultural.

As implicações da TI na BCI não poderiam deixar de integrar a Coletânea, seja em relação à construção de linguagens de indexação/recuperação, seja em relação ao ensino. Os conteúdos sugeridos contemplam conteúdos familiares aos profissionais de informação acrescidos de conhecimentos já abordados na TI, mas imprescindíveis para uma satisfatória coparticipação proveitosa.

Novos horizontes de pesquisa são sugeridos, levantando questões atuais como compartilhamento, interoperabilidade, ontologias. Adoção de padrões de registro de dados, com seus metadados e taxonomias relacionadas, como a mesma definição de conceitos e de forma legível por máquinas não basta. Não importa qual base teórica adotar, qual escola de pensamento adotar, o conceito principal é 'organização' e isto significa classificar: classificar para

modelar, classificar metadados, classificar para estruturar os dados nas diversas taxonomias, classificar para elaborar tesouros. De alguma forma a classificação subjaz, mesmo que implicitamente, os diversos textos da Coletânea, seja quando os autores discutem com rigor as diversas variantes terminológicas, ou quando apresentam os instrumentos de mediação, ou os conteúdos de ensino, ou as novas possibilidades de recuperação – mesmo quando não explicitam os princípios adotados.

No processo da passagem da oitava acima a que nos referimos inicialmente, é importante refletir sobre as contribuições da Biblioteconomia: o que permanece, o que foi alterado, o que foi substituído e o que é novo.

Assim, cada Coletânea provoca novas indagações.

Hagar Espanha Gomes

Universidade Federal Fluminense (UFF)

hagarespanhagomes@gmail.com

CAPÍTULO 1
REFLEXÕES SOBRE OS CONCEITOS
DE TRATAMENTO DA INFORMAÇÃO
E DE ORGANIZAÇÃO DA INFORMAÇÃO
REFLECTIONS ON THE CONCEPTS
OF INFORMATION PROCESSING
AND INFORMATION ORGANIZATION

Gustavo Silva Saldanha

Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
gustavosaldaña@ibict.br
ORCID: 0000-0002-7679-8552

Rodrigo de Sales

Universidade Federal de Santa Catarina (UFSC)
rodrigo.sales@ufsc.br
ORCID: 0000-0002-8695-9807

Ligia Maria Arruda Café

Universidade Federal de Santa Catarina (UFSC)
ligia.cafe@ufsc.br
ORCID: 0000-0002-3934-106X

Resumo: A construção de um domínio de investigação teórica e aplicada em torno dos conceitos de “tratamento da informação” e de “organização da informação” guarda um conjunto de questões terminológicas que colocam o discurso sobre a ordenação da relação entre palavras, seres e coisas em permanente reflexão. Esta pesquisa coloca em discussão estes conceitos a partir de uma perspectiva crítico-pragmática. Como procedimentos metodológicos, a investigação parte do ponto de vista do léxico da lusofonia brasileira em Ciência da Informação, recorrendo às evidências de produção científica via Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); do ponto de vista internacional, o estudo adota a base de dados *Library and Information Science Abstracts* (LISA). O recorte temporal para as assertivas foi integral, ou seja, a procura estava por toda a repercussão da BRAPCI (de 1972 a 2019) e da LISA (de 1969 a 2019), com foco na presença dos conceitos nas figurações pré-textuais da produção científica do campo, a saber, título, resumo e palavras-chave. O percurso de análise e discussão coloca o panorama dos levantamentos nas bases sob as lentes geopolíticas e historiográficas para identificar suas nuances. Os resultados demonstram as variações históricas dos seus conceitos em diferentes contextos de uso, identificando as tendências de desenvolvimento de cada noção.

Palavras-Chave: Tratamento da Informação. Organização da Informação. Organização do Conhecimento. Pragmática informacional. Terminologia – Ciência da Informação.

Abstract: The construction of a theoretical and applied research around the concepts of “information processing” and “information organization” holds a set of terminological questions that put the discourse on the ordering of the relationship between words, beings and things in permanent reflection. This research discusses these concepts from a critical-pragmatic perspective. The research starts from the Brazilian Lusophone lexicon in Information Science, resorting to the evidences of scientific production via the Reference Database of Periodical Articles in Brazilian Information Science (BRAPCI); from an international point of view, the study adopts the Library and Information

Science Abstracts (LISA) database. The temporal scope was integral, that is, the demand was for all the repercussions of BRAPCI (from 1972 to 2019) and from LISA (from 1969 to 2019), focusing on the presence of the concepts in the pre-textual figurations of production in the scientific field, namely, title, abstract and keywords. The course of analysis and discussion places the panorama of the bases under the geopolitical and historiographic lenses to identify their nuances. The results demonstrate the historical variations of their concepts in different contexts of use, identifying the development trends of each notion.

Keywords: Information processing. Organization of information. Knowledge organization. Informational pragmatics. Terminology – Information Science.

1. Introdução

A construção de um domínio de investigação teórica e aplicada em torno dos conceitos de “tratamento da informação” e de “organização da informação” guarda um conjunto de questões terminológicas que colocam o discurso sobre a ordenação da relação entre palavras, seres e coisas em permanente reflexão. No entanto, as dificuldades inerentes ao plano da terminologia não podem ser colocadas como entraves ao próprio desenvolvimento destes conceitos e em sua aplicação. Uma reflexão sobre a pragmática da vivência de tais conceitos demonstra que o imperativo semântico, mesmo determinando uma dada barreira de sentido, não provoca nem a paralisia no seu uso, nem a ausência integral de comunicação entre os pares.

A crise terminológica do domínio que abrange esses conceitos foi investigada em diferentes cenários. É o caso, por exemplo, das reflexões oriundas das pesquisas de Brascher (1986), Svenonius (2000), Alvarenga (2003), Gilchrist (2003), Lancaster (2004), Lara

(2004), Naves e Kuramoto (2006), Kobashi (2007), Hjørland (2008), Brascher e Café (2010).

O caso da imprecisão terminológica e da dificuldade de “organização dos termos” no plano conceitual do próprio domínio pode ainda hoje repercutir de modo objetivo as conclusões de Brascher (1986) sobre a questão lexical da epistemologia da Ciência da Informação. Conforme a pesquisadora, os problemas terminológicos giravam ali em torno da proliferação: a) de significados diferentes para o mesmo termo, de noções oriundas da linguagem comum; b) da produtividade de conceitos novos resultante da aplicação da variedade de modelos teóricos; c) do uso de termos adotados por convivência contingencial, sem a preocupação com os possíveis padrões do campo; d) da criação desnecessária de novos termos para representar conceitos já existentes; e) de variações semânticas de termos a partir da variação no uso. Em síntese, a investigação de Brascher (1986), apontava para um conjunto de problemas oriundos dos cenários de “empréstimo”, “ambiguidade” e “neologia”.

Esse mesmo impacto é apontado pela apresentação crítica de Lancaster em sua obra *Indexing and abstracting in theory and practice*, publicada em 1991 (Lancaster, 2004). O olhar do pesquisador no breve preâmbulo intitulado “A reinvenção da roda” está justamente na análise da ausência de perspectiva científica e autocrítica do campo, especificamente no domínio de investigação que abrange, dentre muitos outros, os conceitos de “tratamento da informação” e “organização da informação” (atravessando ainda noções como classificação, taxonomia, ontologia).

Os possíveis problemas se multiplicam no decurso dos últimos 30 anos de produção teórica e aplicada no que hoje, em geral, tratamos como “organização do conhecimento”, não sem um agregado de dilemas no âmbito da definição e do uso. As transformações sociopolíticas dos anos 1990 e do início do século XXI, a ampliação dos diálogos disciplinares junto a campos do conhecimento

como Filosofia, Computação, Sociologia, Linguística, Ciências Cognitivas, dentre outros, e a contínua luta epistemológica para a emancipação do domínio representam parte considerável das razões que conduzem-nos ao aprofundamento de barreiras iniciais para a compreensão do vocabulário científico em questão. Esses fenômenos são claramente retratados nos trabalhos de fundo epistemológico de González de Gómez (1993), Dahlberg (2006) e Hjørland (2008).

Dada a contextualização do cenário, nossa procura nesta reflexão está em nos interrogar pela construção e aplicação de conceitos centrais no desenvolvimento do domínio de práticas de organização e representação de palavras, seres e coisas, adentrando o labirinto de sua terminologia pelas lentes das noções “tratamento da informação” e da “organização da informação” no português (variante brasileira).

No plano procedimental, algumas observações metodológicas justificam e orientam nossa travessia. A primeira, fundada na própria condição da tensão terminológica contextualizada. O pressuposto da investigação, a partir das barreiras identificadas pela literatura especializada do domínio, nos coloca como observadores oculares da disputa pelo sentido. Em segundo lugar, o fato anterior é respondido por uma ferramenta epistemológica: a filosofia da linguagem estruturada na pragmática, tendo como uma das bases a linhagem de reflexão sobre o significado como uso a partir da segunda fase do pensamento de Ludwig Wittgenstein (1979) e de Bárbara Cassin (2018). A partir desta última, nossa linha de reflexão segue os indícios do uso dos termos para a tentativa de compreensão de suas circunstâncias de aplicação e, logo, de potencial significação. Por fim, da pragmática à teoria social da linguagem, a partir de Bakhtin (2006) e Bourdieu (2008), propomos a discussão sobre a movimentação destes termos (e dos possíveis “parceiros”) pela literatura do campo dentro da realidade geopolítica na qual estão inseridos.

No âmbito dos lugares de fala e do recorte da amostra panorâmica, nosso olhar parte dos dilemas da realidade brasileira que interagem no domínio em direção ao contexto internacional. Ao analisar o domínio pela via de dois conceitos, nosso dilema terminológico já está localizado em um conjunto de construtos sobre o modo como, na própria Ciência da Informação brasileira – expressiva representante da produção científica no âmbito quantitativo da lusofonia internacional do campo –, os fenômenos de imprecisão se deram. Assim, partindo do ponto de vista do léxico da lusofonia brasileira em Ciência da Informação, recorreremos às evidências de produção científica via Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) e do ponto de vista internacional da *Library and Information Science Abstracts* (LISA). O recorte temporal para as assertivas foi integral, ou seja, a procura estava por toda a repercussão da BRAPCI (de 1972 a 2019) e da LISA (de 1969 a 2019), com foco na presença dos conceitos nas figurações pré-textuais da produção científica do campo, a saber, título, resumo e palavras-chave. Dada a condição qualitativa do estudo, o levantamento não visou uma demarcação quantitativa do corpus, ou seja, não nos interessa a expressão fenomênico-numérica de tal produção, mas, sim, o retrato da produção inicial, sua profusão no século XXI e os contextos de mutação para fins de reflexão crítica. As expressões quantitativas (demarcadas no período de levantamento até o mês de abril de 2019) são adotadas como aproximações panorâmicas ao contexto de crise observado, permitindo as primeiras reflexões sobre a inquietação teórica e os modos de vivência do domínio através dos conceitos “tratamento da informação” e “organização da informação”. Além do cenário inicial de uso dos termos nos anos 1960 e sua multiplicação posterior, recorreremos a algumas fontes históricas do passado anterior ao desenvolvimento da produção científica formalizada em institutos de pesquisa e programas de pós-graduação, encontrando ecos do discurso sobre “tratamento da

informação” e “organização da informação” entre o fim do século XIX e o contexto pré-segunda-guerra-mundial na obra de Paul Otlet. Os recursos históricos nos permitem reconhecer os indícios de uso das noções, principalmente pela presença isolada ou não dos termos “organização” e “tratamento” nos primórdios do campo.

Em linhas gerais, com a proposta, à luz das problematizações de tradução conceitual (com foco na discussão das relações de significados, e não na especificidade da tradução de significantes) atentadas pela filósofa Bárbara Cassin, em seu movimento pragmático de leitura dos conceitos filosóficos, estamos aqui menos atentos às definições, e mais interessados nos potenciais de compreensão de um termo dentro de sua vivência (conforme a noção wittgensteiniana), de sua circulação e seu uso na realidade tecida através da linguagem. As interrogações, deste modo, são aqui mais úteis do que as assertivas.

2. Os nomes: definir para usar ou usar para definir?

Nossa primeira tentativa está em dialogar com possíveis territórios fronteiriços dos termos. A movimentação pelas “fronteiras”, aproximando-se dos possíveis centros de gravitação do sentido de cada conceito se dá em razão do exposto anteriormente. Não estamos necessariamente diante um olhar diacrônico sobre as sincronias de produção de significante e de significados no campo e no domínio em questão, aptos a construir “a” definição. É justamente a história do campo – sua geopolítica e sua configuração epistemológica, as crises identificadas até aqui –, que demonstra a impossibilidade essencialista do sentido, sendo o domínio em foco o exemplo máximo de tal crise. Isso se justifica por ser exatamente o agregado de teorias e de métodos, da própria práxis existente sob as noções de “tratamento da informação” e “organização da informação”, res-

ponsável, dentre outras tarefas teórico-metodológicas, por reduzir a entropia da multiplicidade de vocabulários, suas ambiguidades, suas sobreposições, suas inconsistências.

Em alguns casos, as noções mencionadas “se encontram” na mesma busca por uma representação via linguagem das palavras, seres e coisas. É o caso, por exemplo, de Ribeiro (2012, grifo dos autores).

Os marcos da evolução da Biblioteconomia e da Arquivística firmam-se pela componente técnica das disciplinas relacionadas com o estudo e as práticas de **organização da informação**. [...]. O desenvolvimento de disciplinas preocupadas em estudar e aplicar saberes que se relacionam com o **tratamento da informação** são uma criação recente, projetando novos paradigmas informacionais e científicos.

Em outros contextos, observamos a relação articulada dos termos em discussão, como em Amaral e Kobashi (2017, grifo dos autores).

A **organização e tratamento da informação** nos arquivos é um processo essencial na prática arquivística, que comporta um conjunto de atividades encadeadas para propiciar pesquisas e acesso aos documentos.

O caso do termo “tratamento da informação” desdobra-se de modo lento entre os anos ainda que sua presença possa ser identificada de modo mais recorrente nos anos 1970 (quando aparecem as primeiras 14 ocorrências) e 1980, multiplicando-se com o avanço das pós-graduações e da produção científica em periódicos após os anos 1990 no Brasil. Já no primeiro número do periódico inaugural do campo no país, a revista Ciência da Informação, de 1972, do então Instituto Brasileiro de Bibliografia e Documentação (IBBD), transformado em 1976 em Instituto Brasileiro de Informação em

Ciência e Tecnologia (IBICT), o termo se faz presente, como o caso do clássico trabalho de Celia Zaher e Hagar Espanha Gomes, “Da bibliografia à ciência da informação: um histórico e uma posição”, publicado na revista *Ciência da Informação*.

No percurso dos últimos 30 anos encontramos 358 (trezentos e cinquenta e oito) ocorrências do termo na produção científica do campo, atravessando relações de “tratamento da informação” com tecnologia, arquivos, bibliotecas, periódicos, produção científica, mediação, gestão documental, pontos de acesso autorizados, fotografia, acesso aberto, inteligência artificial, política de indexação, comunicação científica, indexação automática, classificação, representação cartográfica, recuperação da informação, ética, preservação digital, arquitetura da informação, repositório, cinema, fontes de informação, saúde, gestão do conhecimento, folksonomia, ontologia, competência em informação, semiótica, mapa conceitual, curadoria digital, análise de citação, regime de informação, divulgação científica, inteligência competitiva, dentre muitos outros. A lista panorâmica de relações, longe de exaustiva, nos demonstra a profusão de uso, a pluralidade e a flexibilidade do conceito “tratamento da informação” em sua performance no léxico informacional como método, como prática ou como construto teórico no campo.

Verifica-se, nesse percurso de uso do termo “tratamento da informação” no cenário brasileiro, as seguintes variações terminológicas: “tratamento da informação”, “tratamento informacional”, “tratamento temático da informação”. Esse último, na produção do campo, configura-se, em suas ocorrências, com o delineamento orientado para a representação de assuntos em documentos (nos seus mais diferentes formatos e suportes), apontando potencialmente para uma precisão mais clara.

Em obra repercutida internacionalmente no campo bibliotecológico, na virada da década de 1960 para 1970, Antony Charles

Foskett, preocupado com o distanciamento entre processos de catalogação de assuntos, classificação e indexação, propôs o uso de um termo mais abrangente que reaproximasse os estudos relativos a tais processos – *Subject Approach to Information*. A ideia de Foskett (1969, edição consultada, 1986) era proporcionar aos estudantes de Biblioteconomia uma visão mais ampla que colocasse em relevo a real dimensão do fazer técnico-biblioteconômico. É importante frisar que, em uma perspectiva britânica, como, por exemplo, encontrado em Lancaster (2004), classificação e indexação, embora concebam produtos e serviços distintos, são frutos da mesma atividade intelectual, ou seja, trabalhar conceitos e assuntos contidos em conteúdos documentais. Nesta perspectiva, a proposta de Foskett (1969, edição consultada, 1986) era adotar o termo *Subject Approach to Information* para se referir ao amplo quadro teórico e metodológico que abarcava tanto a catalogação de assuntos quanto a classificação e a indexação.

Em 1973, Antonio Agenor *Briquet de Lemos traduziu a obra de Foskett para o Brasil, definindo a tradução do referido termo* como Abordagem Temática da Informação. Décadas depois, Guimarães (2008, 2009), influenciado pela obra de Foskett, adotou o termo Tratamento Temático da Informação (TTI) para destacar a interlocução entre algumas correntes teóricas preocupadas com o tratamento da informação no quesito temático. Nos referidos trabalhos, Guimarães (2008, 2009) colocou sob a ótica do TTI as seguintes correntes teóricas: catalogação de assuntos, de matriz norte-americana, preocupada com a geração de produtos de bibliotecas (catálogos e cabeçalhos de assuntos); indexação, de matriz inglesa, destinada ao desenvolvimento de instrumentos de classificação e indexação (esquemas classificatórios e índices); acrescentou ainda, a corrente francesa da análise documentária, preocupada, por sua vez, com o desenvolvimento teórico-metodológico dos processos atinentes ao TTI (Guimarães, 2009).

É importante deixar claro que tanto Foskett quanto Guimarães lançaram o olhar para uma dimensão específica do tratamento da informação, uma vez que, tradicionalmente, o tratamento da informação (TI) se manifesta em duas dimensões: uma dimensão descritiva, responsável pela representação formal dos documentos (catalogação), e uma dimensão temática, responsável pela representação do conteúdo dos documentos (catalogação de assuntos, indexação, classificação e análise documentária). Nesse sentido, o tratamento temático da informação (TTI) se refere apenas à dimensão temática do TI.

Esta visão de que há um tratamento da informação que se manifesta ora pelo tratamento descritivo, ora pelo tratamento temático, teve início na Biblioteconomia e hoje encontra espaço tanto na Ciência da Informação brasileira, conforme pode ser observado no âmbito da Associação Nacional de Pesquisa e de Pós-Graduação em Ciência da Informação (ANCIB), quanto na *knowledge organization* internacional, como verificado no âmbito da *International Society for Knowledge Organization* (ISKO).

No plano internacional, a dificuldade tende a ser acrescida pelo problema da tradução linguística e da tradução terminológica. Dada a hegemonia da língua inglesa na definição de um idioma científico total ao longo do século XX, centralmente em sua segunda metade, cobrindo um universo de linguagens e suas singularidades, é sob a anglofonia que buscamos a extensão da problemática lusófona brasileira.

A tradução literal (tratamento como *treatment*) não aponta para o turvo espelhamento terminológico. Outros termos em língua inglesa contextualizados no léxico da Ciência da Informação são claramente adotados na correlação, respondendo pelas semelhanças com os sentidos de “tratar”, “manejar”, “manipular”. É o caso, centralmente, de *information processing* – aqui, *processing* responde por “processamento”, “tratamento”. O termo “processamento da

informação” aparece em língua portuguesa, porém com ocorrências sobrepostas ao termo “tratamento da informação”, e com expressão quantitativa inferior na BRAPCI a este último. Seu uso – ou seja, a tradução literal de *information processing* por “processamento da informação” guarda inúmeras dificuldades, aqui não trabalhada pelo foco da reflexão. Por essa razão, detemo-nos aqui na relação (e não na tradução), de “tratamento da informação” e *information processing*, presente no plano linguístico e terminológico.

No plano internacional, *information processing* representa uma classe para a LISA, ou, no plano documentário do tesouro, um “termo geral”, acompanhado dos seguintes termos específicos, *data collection, data processing, distributed processing, multiprocessing, multitasking, online transaction processing, parallel processing*. A nota de escopo do tesouro é objetiva: “o processamento da informação é, em Psicologia e Ciência da Informação, a aquisição, o armazenamento e a manipulação da informação para torná-la útil”. O conceito se relaciona (ou seja, os “termos relacionados” dados pela estrutura do tesouro da LISA) com algoritmos, distúrbio do processamento auditivo, cognição e raciocínio, desenvolvimento cognitivo, Ciência da Computação, tolerância ao erro, eurística, sobrecarga de informação, tecnologia da informação, teoria da informação, reconhecimento de padrões, velocidade de processamento, enfileiramento, tempo real, roteamento, semântica, análise de sistemas, gerenciamento de sistemas. A profusão de relações de domínio, de métodos e de campos científicos dada na amostra do tesouro sugere uma correlação de semelhança (ainda que não de identidade) com o uso amplo no território lusófono de origem brasileira (como pode ser observado acima na lista de termos relacionados pela BRAPCI ao conceito “tratamento da informação”).

No plano do significante, problemas também são encontrados na LISA, assim como no caso da língua portuguesa. As três expressões a seguir aparecem no jogo de semelhanças: *information proces-*

sing, processing information e processing of information. Dado o mapeamento internacional da produção em língua inglesa, e reconhecida a pluralidade dos modos de produção do idioma em cada contexto, a variação pode sugerir, a princípio, uma problemática de construção do “inglês” em cada território. Apenas uma pesquisa exaustiva ou com amostragem empiricamente representativa poderia nos responder, de todo modo, este questionamento e as possíveis distinções entre as três ocorrências.

O que nos coloca, sobremaneira, no foco da crise terminológica é que, como demonstrada, a profusão de significados no contexto internacional só pode se multiplicar em espectros geométricos ainda mais vastos, dado o resultado levantamento simples na LISA. Encontramos o total de 39.342 (trinta e nove mil, trezentos e quarenta e dois) itens indexados pelo termo *information processing* na base. Lembremos sempre a questão geopolítica de limitação de um grupo específico de periódicos internacionais, a saber, a hegemonia da língua inglesa e os critérios de seleção de publicações estruturados em lógicas econômico-bibliométricas. Essas ressalvas nos fazem recordar a limitação destes dados oriundos da LISA, ou seja, estamos aqui reconhecendo que, mesmo em língua inglesa, esse resultado de cerca de 40 mil registros é ínfimo e se multiplicaria ainda mais fora da base de dados consultada, em outros contextos de produção anglófona em Ciência da Informação.

Retomando a potencialidade e os riscos da relação terminológica (e não uma tradução), faz-se necessário recontextualizar a aproximação entre “tratamento da informação” e *information processing*. Os confrontos entre lusofonia e anglofonia na produção de sentidos dos termos demonstram um uso do segundo como repercussão direta do processamento eletrônico nos anos 1960 (com ecos no Brasil já nos anos 1970). Os primeiros registros presentes na LISA datam de 1968, estruturalmente vinculados à questão eletrônica, à recuperação da informação através dos computadores. Por sua vez,

via a lusofonia tomada como amostra através das lentes do Brasil, a perspectiva em construção a partir dos anos 1970 sob o termo “tratamento da informação” está orientada para a manifestação da representação, através de palavras, de artefatos bibliográficos, seus padrões, suas especificidades. O afastamento de sentido não se torna, todavia, um modelo. Os mesmos casos relatados como singulares se sobrepõem nos dois horizontes, ou seja, como já demonstrado, na literatura brasileira encontramos a relação com o eletrônico e o digital no uso do termo, assim como na dimensão internacional, via LISA, encontramos o uso do *information processing* aplicado ao estudo de bibliotecas, gestão de coleções, registros bibliográficos e teorias da representação da realidade através de vocabulários.

Movimentemo-nos agora pelo conceito de “organização da informação” para, a seguir, retomarmos ao jogo pragmático do uso de ambos os termos em nossa literatura especializada. Sua construção sólida parece-nos dada na produção científica do campo, sob o horizonte inicial no cenário brasileiro, a partir dos anos 1990. O recurso de compreensão via BRAPCI demonstra que, até os anos 1980, ou seja, na primeira década de produção científica no Brasil (a década anterior), o termo “organização da informação” era praticamente inexistente. Registra-se apenas um caso de aparição na produção de descritores e metarrepresentação temática (títulos, resumos e palavras-chave). É em 1973, na pesquisa de Lélia Cunha sobre “normalização de originais”, onde encontramos o único uso do termo, porém no contexto da expressão “organização da informação especializada”. Vê-se aqui que a aplicação da noção, em seu contexto inicial, não está vinculada aos domínios como “representação temática”, “representação descritiva” e “linguagens documentárias”, mais comuns nas futuras apropriações do conceito nas décadas seguintes. Seu foco está no âmbito da normalização. Entre os anos 1990 e a década em curso o número de ocorrências do conceito se multiplica geometricamente.

Se no caso lusófono pela via da BRAPCI o termo “organização da informação” ganhará uma expressão quantitativa vasta, principalmente nos últimos trinta anos, seu uso no contexto internacional, via LISA, é consideravelmente distinto. É preciso considerar, neste ponto, que a base LISA não contempla só a totalidade do campo, apesar de sua representatividade internacional. Ao mesmo tempo, outro limite é preciso ser considerado: a capacidade da base chegar até as condições de representação específica do domínio da organização do conhecimento. Ressalvadas as questões, atentamos inicialmente para a singularidade da ausência do termo no tesouro da base de dados: *information organization* é um conceito ausente como “termo geral”. De uma margem interpretativa de fundo analítico, poder-se-ia afirmar: no arcaibouço mapeado pela LISA, a expressão não representa um conceito. Por esta razão, não é possível considerar o resultado de uma busca simples na base como parte de uma compreensão inicial do uso da expressão, isto é, “organização da informação” não nos permite, no maior canal de repercussão da produção científica mundial da Ciência da Informação, uma representação conceitual da realidade da investigação científica no campo. Na verdade, os 100.065 (cem mil e sessenta e cinco) resultados na LISA via o levantamento por *information organization* respondem por uma miríade de usos com centenas de milhares de apropriações, incluindo (mesmo no título), a expressão na íntegra, ou seja, *information organization*. Os primeiros casos aparecem junto do nascimento da maior base internacional de resumos em Ciência da Informação: o uso do termo já está presente em publicações de 1968 e 1969, relacionado às pesquisas científicas e abordagens institucionais, como da *American Library Association* (ALA).

Esses cem mil registros, entretanto, nos dão algumas direções interessantes. Através do demonstrativo de refinamento de busca na LISA via o enquadramento geral por “assuntos”, compreendemos

que a expressão está aplicada aos seguintes construtos: “estudos” (maior vinculação, com 11.761 registros, porém mais aberta e sem uma clara relação de mútua exclusividade perante os demais “assuntos”), “bibliotecas” (segunda maior correlação, com 10.382 ocorrências), seguidos por “administração”, “organização”, “bibliotecários”, “bibliotecas universitárias”, “trabalho informacional”, “gestão do conhecimento”, “tecnologia da informação”, “bibliotecas públicas”, “internet”, “administração de bibliotecas”, “websites”, “estudos de não ficção”, “organizações”, “pesquisa propriamente dita”, “cooperação”, “sistemas de informação”, “gestão da informação”, “serviços”, “Estados Unidos”, “conferências”, “Reino Unido”, “redes sociais” e “bibliotecas digitais”. Estes são os principais “assuntos” dentro dos quais o termo *information organization* – incluindo suas variações – é aplicado (sendo que nos limitamos, aqui, a demonstrar os assuntos que possuem um mínimo de duas mil ocorrências cada, dada a quantidade de outros descritores abaixo de dois mil registros recuperados pela base).

A profusão de uso e de suas relações dá-nos, partindo agora do plano internacional para o plano lusófono, um conjunto de indicadores para a compreensão de sua apropriação e de seus significados. No plano epistemológico, uma das revisões mais críticas do domínio em questão, desenvolvida em 2008 por Birger Hjørland (2008), recorrendo a Anderson (2003), procura representar de modo muito breve e objetivo o uso do termo: no campo informacional (ou seja, *in the context of library and information science*), o termo “organização do conhecimento” é uma forma de redução da expressão *knowledge resources organization*, o que é “frequentemente” (*often*) chamado de *information organization*. O percurso epistemológico apontado por Anderson (2003) nos coloca, pois, diante de um dilema inverso: se a maior base de dados internacional em Ciência da Informação, a LISA, não legitima o conceito “organização da informação”, por sua vez, na visão da *International Encyclopedia of Information and*

Library Science, trata-se de um dos mais importantes macroconceitos do campo, manipulado a partir de outros significantes.

Inicialmente, o que nos parece ser o mais importante no esclarecimento de Anderson (2003) não é a correlação entre os termos. O termo “frequentemente” aplicado ao debate sugere um conflito com os resultados recuperados na LISA para *information organization*, assim como sua presença na BRAPCI como “organização da informação”. A representação das evidências aponta para o uso intensivo da noção, ou seja, como um dos mais identitários conceitos da própria Ciência da Informação, representando na integralidade o domínio que reuniria, dentre inúmeros termos, aquele aqui abordado, “tratamento da informação”.

Voltemos à correlação de Anderson (2003) e ao diálogo estabelecido por Hjørland (2008). Do ponto de vista de Anderson (2003), por uma via pragmática, ou seja, por uma explicação tradicionalmente orientada pelo e para o uso, encontramos a verossimilhança entre “organização da informação”, “organização de recursos do conhecimento” e, por fim, “organização do conhecimento”. Em outras palavras, são relações dadas no campo que demarcam a ausência de distinção de tais termos na esfera de sua vivência, sendo suas possíveis diferenças fundadas apenas na hipotética observação de caso a caso, na subjetividade não demarcável de cada autoria e de cada apropriação. No contexto brasileiro, Brascher e Café (2010) delineiam uma proposta preliminar dos conceitos de “organização do conhecimento” e “organização da informação”. Segundo as autoras, “temos dois tipos distintos de processos de organização, um que se aplica às ocorrências individuais de objetos informacionais – o processo de organização da informação, e outro que se aplica a unidades do pensamento (conceitos) – o processo de organização do conhecimento. A organização do conhecimento compreende, também, a organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções, neste caso, temos a

organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos. A organização do conhecimento, por sua vez, visa à construção de modelos de mundo que se constituem em abstrações da realidade. ” (Brascher & Café, 2010, p. 93).

Do ponto de vista de Hjørland (2008), o interesse está em justamente discutir o último conceito, “organização do conhecimento”, reconhecida sua expressão como orientadora geral do domínio em discussão, ou seja, já partindo de sua hegemonia no campo. Não é senão o mesmo foco, ainda que por intenções distintas, de Dahlberg (2006), na procura pela compreensão não mais de um domínio, mas de uma ciência geral chamada *knowledge organization*. Ambos, Hjørland (2008) e Dahlberg (2006), partem (principalmente a última) de uma abordagem *top down* no plano epistemológico e investigam o domínio não legitimando a construção histórica da noção “organização do conhecimento”. O panorama institucional histórico da *International Society of Knowledge Organization* (ISKO), a partir de 1989, contribui para o entendimento desta força do conceito que adota o termo “conhecimento” em sua predicação. Porém, não explica, do ponto de vista do uso, sua relação com a “organização da informação”, restando a explicação de Anderson (2003) mais prudente para uma historiografia de nosso léxico pela estrada pragmática, na correlação (ainda) com o termo “organização dos recursos do conhecimento”. A discussão nos leva ao aprofundamento da reflexão epistemológica e sociopolítica, objeto de nosso próximo passo.

3. Das configurações geopolíticas em uma epistemologia histórica: como relacionar “organização”, “tratamento” e “informação”?

Como podemos discutir as condições geopolíticas históricas que condicionaram a construção e a aplicação dos conceitos “organização

da informação” e “tratamento da informação”? A “redução” do potencial de sentido do termo “organização” nos processos históricos e sua retomada de crescimento – do século XIX até o desdobramento do *Classification Research Group* através da ISKO, com Ingetraut Dahlberg – tem inúmeras possibilidades de entendimento a partir da filosofia. A noção de “ordem”, de “controle”, de luta contra a entropia não representa uma expressão moderna constituída pela termodinâmica. Trata-se de uma clássica reflexão da filosofia antiga orientada para o estudo do conceito de “simples”, correlacionado com a unidade do conceito de “ser”. Da filosofia ao território epistemológico da Ciência da Informação, a complexidade do conceito de “organização”, por sua vez, em fontes históricas do domínio onde transitam os termos “tratamento da informação” e “organização da informação” é vasta. “Organização” significa a totalidade das práticas institucionalizadas e racionalizadas no campo hoje dito informacional, como o caso da “enciclopédia especializada” *Public Libraries in the United States of America* (1876), nas cartas da década de 1890 entre Melvil Dewey (1978) e Paul Otlet (1978), e no clássico *Traité de Documentation* deste último (Otlet, 1934) – neste último caso, encontramos as manifestações de “organização da paz”, “organização de redes”, “organização de instituições” – como “bibliotecas”, “arquivos” e “museus” – “organização racional”, “organização internacional do livro e da documentação”, “organização racional dos livros e dos documentos”.

Essas fontes históricas, dispersas em 60 anos de reflexão anterior à 2ª Guerra Mundial, nos dão indícios do modo como os conceitos se desenvolveram no tempo, demonstrando um distanciamento não demarcável nos sentidos futuros pós-formalização das atividades científicas do campo no mundo. Por exemplo, no final do século XIX nos Estados Unidos, a noção de “tratamento” é usada já para definir padrões de uniformidade de processos e produtos dentro de bibliotecas. Na discussão sobre classificação e arranjo de livros em bibliotecas, encontramos a afirmação:

It would, moreover, simplify the conscientious fulfilment of the librarian's duties, so that it would no longer be necessary for each librarian to have detailed accounts regarding his **treatment of library science**. (EUA, 1876, p. XXIV, grifo dos autores)

O termo *treatment*, aqui usado de modo generalista, para toda a “ciência biblioteconômica”, apontava, ali, já em 1876, para os usos contemporâneos. Isto fica claro quando o tratado estadunidense de 1876 aponta, na discussão sobre o modo de trabalho das bibliotecas científicas, o método de *treatment* da informação especializada. Em outro caso, a expressão literal *treatment of books* é adotada.

Na mesma obra, o termo “organização” aparecerá em dezenas de ocorrências, com diferentes manifestações, incluindo a noção de *library organization*. Em dado contexto, a noção de *management* é colocada como sinônimo de *organization and service*, ora como dois conceitos centrais, como no título do vigésimo quinto capítulo da obra, “The organization and management of public libraries”. O que a evidência da obra permite interpretar é que, mais de cem anos antes da ISKO e do papel estratégico-epistemológico que o termo *organization* ganha junto aos termos *information* e *knowledge*, o mesmo já era adotado como macroconceituação no campo, abrangendo das práticas de arranjo, à classificação e à catalogação, incluindo ainda seleção e circulação de livros, ou seja, já se candidatava a um (co)termo epistemológico, um macroconceito (EUA, 1876).

O caso em Otlet (1934) também nos ajuda a compreender o percurso do conceito. Da travessia dos diálogos com Melvil Dewey na última década do século XIX até a publicação de seu tratado, é reconhecida uma vasta produção conceitual. O termo francês *traitement* aparecerá no teórico belga como parte do “tratamento lógico de assunto” no plano da organização da produção científica, tendo sua forma estabelecida claramente na expressão “*principes*

généraux de traitement de documents bibliographiques.” (Otlet, 1934, p. 193, grifo dos autores)

O termo *organisation*, por sua vez, compõe a estrutura do pensamento otletiano. Intitulando o extenso capítulo teórico-empírico “Organisation rationnelle des Livres et des Documents”, o termo aparecerá em muitos outros contextos da obra. São estabelecidas, por exemplo, a partir de *organisation*, as aproximações e as relações entre “organização e racionalização”, “organização internacional (de instituições documentárias e de documentos)”, “organização universal (bibliográfica)”. Ao todo, são 231 (duzentas e trinta e uma) ocorrências que permitem demonstrar o peso epistemológico do termo, também antecipando os futuros usos anglófonos e lusófonos na pesquisa científica avançada em Ciência da Informação.

As fontes históricas pré-segunda-guerra-mundial parecem-nos não sugerir grande distinção para os futuros usos das noções “organização da informação” e “tratamento da informação”, sendo esta última orientada para o processamento de elementos de um dado artefato com vistas a torná-lo acessível, reapropriável no espaço-tempo. Já “organização da informação”, principalmente pela carga epistemológico-histórica do termo “organização” demonstra-se como um macroconceito que abriga teoria e aplicação, uma espécie de teoria geral com suas formulações empiricamente demonstráveis através de abordagens como, poderíamos neste ponto afirmar, o “tratamento da informação”.

Chegamos ao aspecto da questão da “informação”. Se o termo se coloca como original, para alguns relatos, a partir da 2ª Guerra Mundial, principalmente pela apropriação da Teoria Matemática da Comunicação de Claude Shannon e Warren Weaver, desde 1948, seu uso também não tende a revelar grandes mutações. Ao contrário, em muitos dos casos, como a própria literatura sobre a crise terminológica do campo aborda, demonstra um vasto território de ampliação das dificuldades de comensurabilidade entre pares

(ou seja, o termo “informação” viria trazer ainda mais problemas iniciais no âmbito da precisão lexical da ciência que assumiu seu nome próprio a partir do mesmo, ou seja, a Ciência da Informação). O fato histórico-político revelou-nos que, dentre tantos conceitos da teoria shannon-weaveriana, como “ruído”, “canal”, “emissor”, “receptor”, “redundância”, “seleção”, aquele que parecia não ser, nem de longe, a noção central, ganhou um peso tal no plano epistemológico e geopolítico que tomou de assalto a cena social, dentro e fora das comunidades científicas. “Informação” se tornou a palavra candidata a grande vocábulo do século XX.

Seu grau de ambiguidade, de sobreposição e de ausência de uma semântica mínima sob o olhar, por exemplo, de uma filosofia analítica, é desmedido em qualquer levantamento em bases como BRAPCI e LISA. Uma busca por *information* não nos permite, em um primeiro olhar, nenhum resultado passível de delimitação lógica, tamanho o grau de redundância na literatura do campo. No entanto, no âmbito das fontes históricas, lá estava o termo, isto é, a problemática não se inicia após a 2ª Guerra Mundial. Otlet (1934) dará um destaque epistemológico-político para o vocábulo, incluindo-o na expressão *Réseau Universel d'Information et de Documentation*. Na verdade, a primeira e clássica frase do Capítulo 0 do *Traité* contém o termo “informação”. Ali, no capítulo de fundamentação epistemológica do *Traité de Documentation*, tal frase menciona os famosos questionamentos futuros sobre o excesso de registros no mundo, ou seja, a velha questão inerente à *quantité d'information* presente desde o final do século XIX. O termo aparecerá ao longo da obra otletiana, apesar da expressão “organização da informação” não figurar em sua forma no léxico do teórico belga.

Décadas antes da publicação do *Traité de Documentation*, na geração que inspirou Paul Otlet, o coletivo consagrado nos Estados Unidos no ano de 1876, igualmente o termo *information* é comum no vocabulário das ações institucionais, teóricas e empíricas. Deste

modo, no século XIX, noções e expressões como *demands for information, complete statistical information, local information, to collect information, satisfactory information, mass of information, dissemination of information, trustworthy information* (informação confiável), *means of information, sources of information*, além de dezenas de outras aparecem como forma de uso corrente do termo e de sua manipulação como conceito (EUA, 1876).

O recuo às fontes históricas demonstra que a complexidade e a aplicação do termo *information* é claramente anterior, no campo, ao desenvolvimento científico formalizado após a 2ª Guerra Mundial e à teoria matemática da comunicação. A “aplicação” do termo como forma de predicação na relação com os vocábulos “organização” e “tratamento” responde, pois, de um lado, por um lastro histórico e geopolítico no âmbito da epistemologia do campo, de outro lado, por um processo de subordinação de tal complexidade a determinados aspectos ligados ao processamento de artefatos, identificação e retirada de descritores tangíveis e intangíveis. Estas relações demonstram, a nosso ver, o longo percurso da epistemologia histórica que concebe a abertura e as tentativas de demarcação fronteiriça das expressões “organização da informação” e “tratamento da informação”.

4. Considerações finais: o percurso como vivência

O desenvolvimento terminológico do campo informacional a partir das lentes dos conceitos “tratamento da informação” e “organização da informação” demonstra um conjunto de nuances específicas de cada tradição e de cada espaço-tempo. A investigação comprova, a partir das bases BRAPCI e LISA, o rápido desenvolvimento, a partir dos anos 1960, do uso dos conceitos na literatura em Ciência da Informação. É exatamente, pela via pragmática, que o uso se revela

como força propulsora do significado. Suas formas de apropriação permitem remontar condicionantes de uma luta geopolítica, de suas configurações econômicas e das transformações tecnológicas.

O percurso nos demonstra uma dada vivência dos conceitos entre a substancialização e racionalização no período que avança nas últimas seis décadas, junto ao desenvolvimento intensivo, a partir dos anos 1990, do conceito “organização do conhecimento”. As dificuldades de demarcação fronteiriça dos conceitos “tratamento da informação” e “organização da informação” representam, pois, o conjunto de indícios de cada modalidade de uso e de busca por afirmação terminológica, impulsionando a emancipação de coletivos científicos e de formas de organização teórica.

Referências Bibliográficas

- Alvarenga, L. (2003). Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaço digitais. *Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.*, 8(15), 18-40. doi:10.5007/1518-2924.2003v8n15p18
- Amaral, C., & Kobashi, N. Y. (2017). Organização da informação nos arquivos: classificação. In *Anais do XVIII Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), Marília, UNESP, Brasil*. Recuperado de http://enancib.marilia.unesp.br/index.php/XVIII_ENANCIB/ENANCIB/paper/view/81/1242
- Anderson, J. D. (2003). Organization of knowledge. In J. Feather & P. Sturges (Eds.), *International encyclopedia of information and library science* (2nd ed., pp. 471-490). London: Routledge.
- Bakhtin, M. (2006). *Marxismo e filosofia da linguagem*. São Paulo: Hucitec.
- Bourdieu, P. (2008). *A Economia das trocas linguísticas: o que falar quer dizer*. São Paulo: EdUSP.
- Brascher, M. (1986). Terminologia Brasileira em Ciência da Informação: Uma Análise. *Ciência Da Informação*, 15(2), 135-142. Recuperado de <http://revista.ibict.br/ciinf/article/view/234>
- Brascher, M., & Café, L. (2010). Organização da Informação ou Organização do Conhecimento? In M. L. G. de Lara & J. W. Smit (Eds.), *Temas de pesquisa em Ciência da Informação no Brasil* (pp. 85-103). São Paulo: Escola de Comunicações e Artes da USP.
- Cabré, M. T. (1993). *La terminología: teoría, metodología, aplicaciones*. Barcelona: Editorial Antártida/Empúries.

- Cabré, M. T. (1999). *La terminología: representación y comunicación*. Barcelona: Institut Universitari de Lingüística Aplicada.
- Café, L. (2003a). Contribuições da Gramática Funcional na delimitação de segmentos descritores de informação. In G. M. Rodrigues & I. L. Lopes (Eds.), *Organização e representação do conhecimento na perspectiva da ciência da informação* (pp. 118-140). Brasília: Thesaurus.
- Café, L. (2003b). Terminologia: aplicação do (re)modelo de Simon Dik. In E. Faulstich & S. P. Abreu (Eds.), *Lingüística aplicada à terminologia e à lexicologia: cooperação internacional: Brasil e Canadá* (pp. 59-82). Porto Alegre: UFRGS, Instituto de Letras, NEC.
- Cambridge Scientific Abstracts Internet. (2004). Library and Information Science Abstracts (LISA) Thesaurus [Data base]. Recuperado 10 de setembro de 2019, de www.csa.com/csa/
- Cassin, B. (Coord.) (2018). *Dicionário dos intraduzíveis: Um vocabulário das filosofias. Vol. 1: Línguas*. São Paulo: Autêntica.
- Cunha, L. G. C. da (1973). Normalização de Originais. *Ciência Da Informação*, 2(1), 59-63. Recuperado de <http://revista.ibict.br/ciinf/article/view/24/24>
- Dahlberg, I. (2006). Knowledge Organization: A New Science? *Knowledge Organization*, 33(1), 11-19.
- Dewey, M. (1978). Letter from Dewey to Olet, 29 June 1895, 29 June 1895. In S. K. Vann (Ed.), *Melvil Dewey: his enduring presence in Librarianship* (pp. 191-192). Littleton: Libraries Unlimited.
- EUA. (1876). *Public libraries in the United States of America: their history, condition, and management: special report*. Washington: Government Printing Office.
- Foskett, A. C. (1973). *A abordagem temática da informação*. São Paulo: Polígono; Brasília: Ed. UnB.
- Foskett, A. C. (1986). *The subject approach to information* (5th ed.). London: Clive Bingley.
- Gilchrist, A. (2003). Thesauri, taxonomies and ontologies - an etymological note. *Journal of Documentation*, 59(1), 7-18. doi:10.1108/00220410310457984
- González de Gomez, M. N. (1993). A representação do conhecimento e o conhecimento da representação: algumas questões epistemológicas. *Ciência Da Informação*, 22(3), 217-222. Recuperado de <http://revista.ibict.br/ciinf/article/view/479>
- Guimarães, J. A. C. (2008). 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*, 1(1), 77-99. doi:10.26512/rici.v1.n1.2008.940
- Guimarães, J. A. C. (2009). Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. *Ibersid: Revista de Sistemas de Información y Documentación*, 3, 105-117. Recuperado de <https://www.ibersid.eu/ojs/index.php/ibersid/article/view/3730>
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2-3), 86-101. doi:10.5771/0943-7444-2008-2-3-86
- Kobashi, N. Y. (2007). Los problemas terminológicos de la organización y del acceso a la información. In C. Naumis Peña (Ed.), *Memoria del I Simposio Internacional*

- sobre *Organización del Conocimiento: bibliotecología y terminología* (pp. 87-102). Cidade do México: UNAM.
- Lancaster, F. W. (2004). *Indexação e resumos: teoria e prática* (2ª ed.). Brasília: Briquet de Lemos.
- Lara, M. L. G. de (2004). Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. *Ciência Da Informação*, 33(2), 91-96. Recuperado de <http://revista.ibict.br/ciinf/article/view/1050>
- Naves, M. M. L., & Kuramoto, H. (2006). *Organização da informação: princípios e tendências*. Brasília: Briquet de Lemos.
- Otlet, P. (1934). *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelas: Editiones Mundaneum.
- Otlet, P. (1978). Letter from Paul Otlet, 24 March 1895. In S. K. Vann (Ed.), *Melvil Dewey: his enduring presence in Librarianship* (p. 189). Littleton: Libraries Unlimited.
- Ribeiro, F. (2012). Organização e uso da informação: conhecer bem para bem representar. *IRIS – Revista de Informação, Memória e Tecnologia*, 1(1), 7-16. Recuperado de <http://hdl.handle.net/20.500.11959/brapci/93394>
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press.
- Wittgenstein, L. (1979). *Investigações filosóficas* (2ª ed.). São Paulo: Abril Cultural.

Financiamento

Esta pesquisa recebeu o financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – Brasil.

CAPÍTULO 2

**TRATAMENTO TEMÁTICO DA INFORMAÇÃO
A PARTIR DOS TRABALHOS PUBLICADOS NOS
ANAIS DO CAPÍTULO BRASILEIRO DE ISKO:
Pontos de partida, identidade nacional
e agentes epistêmicos**

**SUBJECT REPRESENTATION IN KNOWLEDGE
ORGANIZATION FROM THE WORKS PUBLISHED
IN THE BRAZILIAN CHAPTER OF ISKO:
Starting points, national identity
and epistemic agents**

Daniel Martínez-Ávila

Universidad Carlos III de Madrid (UC3M)

dmartine@bib.uc3m.es

ORCID: 0000-0003-2236-553X

Luciana de Souza Gracioso

Universidade Federal de São Carlos (UFSCar)

luciana@ufscar.br

ORCID: 0000-0002-6320-4946

Resumo: Este trabalho propõe-se a identificar, descrever e analisar as diferentes definições e relações aferidas ao Tratamento Temático da Informação, no contexto da Organização do Conhecimento e

da Informação na Ciência da Informação brasileira. Para tanto, são consultados e analisados os trabalhos publicados sobre o assunto nos anais da ISKO (*Information Society for Knowledge Organization*), capítulo Brasil. Como resultado, procura-se reconhecer, no âmbito do Tratamento Temático da Informação, quais são seus pontos de partida, suas origens internacionais, sua identidade nacional e seus atuais agentes epistêmicos, entendendo que estes elementos são substanciais para futuras análises e prospecções da área. Evidencia-se a partir da análise desenvolvida que Guimarães é o autor cuja produção sobre o tema é mais citada, o que demonstra por outra via que a escola francesa, nos estudos informacionais de Tratamento Temático da Informação, teriam maior influência no desenvolvimento da pesquisa brasileira.

Palavras-Chave: Tratamento Temático da Informação. Tratamento da Informação. Organização do Conhecimento. Organização da Informação. ISKO.

Abstract: This paper aims to identify, describe, and analyze the different definitions and relationships of the Portuguese term “Tratamento Temático da Informação” (subject representation) in the context of knowledge organization in Brazil. We analyzed the papers published in the proceedings of the ISKO Brazil meetings. As a result, we aim to reveal the starting points, international origins, national identity, and current epistemic agents involved in this concept, as we understand that these elements are essential for future analysis and research on the area. The results reveal that Guimarães is the most cited author on the topic, also showing that the French tradition of knowledge organization was a great influence in Brazil.

Keywords: Subject representation. Information processing. Knowledge organization. Information organization. ISKO.

1. Introdução

O Tratamento da Informação pode ser entendido, de modo geral, como um conjunto de ações, vinculadas a Organização do Conhecimento e da Informação, voltadas a análise, representação, descrição e síntese da informação, em ambientes sistêmicos. Mas esta definição não é unívoca e nem consensual no campo da Ciência da Informação (CI). Há variações substanciais tanto no uso de termos para descrever as ações de Tratamento da Informação, como na construção de seus significados. Estas variações promovem, por um lado, um conjunto dinâmico e rico de significações e usos, e por outro podem sugerir distorções e incompatibilidade comunicativa entre pesquisadores, professores e profissionais do campo que se dedicam ao tema. Neste cenário e diante dos crescentes movimentos de produção e internacionalização da Ciência, é que se identificou a necessidade de averiguar o quanto e como tem sido tecida a rede de significação do Tratamento da Informação, em especial o Tratamento Temático da Informação, no âmbito da produção científica especializada neste domínio. Este mapeamento não teria como objetivo servir como ponto de partida para uma proposta de homogeneização ou unificação conceitual sobre o assunto. Ao contrário, intentaria identificar e reconhecer as convergências e as divergências conceituais que se estabeleceram no desenvolver das pesquisas e das práticas de Tratamento da Informação na Ciência da Informação. Com isto, este trabalho se propôs a apresentar um conjunto de variações terminológicas e de significação sobre o assunto, procurando espelhar a amplitude e a potencialidade das ações de Tratamento Temático da Informação e ao mesmo tempo, pode servir como recurso e fonte para promoção de intercâmbio mais consistente no desenvolvimento de pesquisas e ações de ensino sobre o tema.

O objetivo específico deste capítulo então é o de identificar, descrever e analisar as diferentes definições e relações aferidas ao

Tratamento Temático da Informação, no contexto da Organização do Conhecimento e da Informação, pormenorizando as variações terminológicas produzidas na Ciência da Informação brasileira. Ao mesmo tempo, reconhece-se que emergem, no ensejo das análises do Tratamento Temático, as discussões que dizem respeito a própria delimitação conceitual e operacional do que se assume enquanto Organização do Conhecimento e Organização da Informação e tais discussões, em alguma medida, serão tecidas e contextualizadas na perspectiva do Tratamento Temático.

Metodologicamente, o trabalho se desenvolveu a partir de revisão de literatura em pesquisas publicadas nos anais da ISKO (*International Society for Knowledge Organization*) do Brasil. Neste universo, a partir de diferentes estratégias de busca (que são descritas oportunamente em cada seção do texto), foram recuperados os trabalhos que compuseram a amostra para análise de conteúdo.

A partir destas análises apresenta-se um panorama geral sobre as definições, os limites e alcances do Tratamento Temático da Informação, procurando reconhecer seus pontos de partida, suas origens internacionais, sua identidade nacional e seus atuais agentes epistêmicos. Acreditamos que este panorama nos possibilita construir uma visão ampla sobre a configuração deste assunto, permitindo, a partir disto, que sejam feitas prospeções sobre futuros desdobramentos e alinhamentos do campo.

2. Análise dos conceitos de Tratamento Temático e Tratamento Temático da Informação nos congressos de ISKO Brasil

A *International Society for Knowledge Organization* (ISKO) é uma organização científica criada em 1989 na Alemanha com o objetivo de promover a pesquisa e discussão na organização do conhecimento. A ISKO está composta por capítulos nacionais ou regionais (Brasil,

Canadá/Estados Unidos, China, França, Alemanha/Áustria/Suíça, Índia, Itália, Irã, Maghreb, Polônia, Singapura, Espanha/Portugal e Reino Unido) que organizam congressos bianuais nos anos ímpares e de forma itinerante celebram congressos internacionais nos anos pares. O capítulo brasileiro de ISKO foi criado em 2007 e até o momento da presente análise realizou quatro congressos com os seguintes temas: Desafios e perspectivas científicas para a Organização do Conhecimento na atualidade (Brasília, 2011), Complexidade e Organização do Conhecimento: desafios de nosso século (Rio de Janeiro, 2013), Organização do Conhecimento: passado, presente e futuro em um contexto de diversidade cultural (Marília, 2015) e Memória, Tecnologia e Cultura na Organização do Conhecimento (Recife, 2017).

Os trabalhos dos congressos do capítulo brasileiro da ISKO são publicados na série “Estudos Avançados em Organização do Conhecimento”. Estes volumes têm sido utilizados como fontes em diversos estudos sobre a organização do conhecimento já que constituem um reflexo do domínio brasileiro da organização do conhecimento, como por exemplo Gomes, Evangelista, Martínez-Ávila e Grácio (2017) e Guimarães, Martínez-Ávila, Tognoli e Milani (2018). O primeiro volume da série, “Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade” (Guimarães & Dodebei, 2012), incluiu as contribuições ao I Congresso Brasileiro de Organização e Representação do Conhecimento celebrado em Brasília – DF nos dias 20-22 de outubro de 2011. Decorrentes deste primeiro congresso de ISKO Brasil se publicaram 40 trabalhos (incluindo uma introdução e considerações finais), todos eles no campo da organização do conhecimento/informação por estarem inseridos no contexto da ISKO. Destes trabalhos, só oito capítulos (20%) discutem o usam os termos “tratamento temático da informação” ou “tratamento da informação”. Dos 40 capítulos também cabe ressaltar que apenas

15 trabalhos (37,5%) usam a expressão tratamento, nem sempre adjacente ao termo informação e muitas vezes como sinônimo geral de processamento.

Na introdução do livro, José Augusto Chaves Guimarães e Vera Dodebei (2012) revisam a história da ISKO e a organização do conhecimento (sem utilizar em nenhum momento o termo organização da informação) e colocam o termo “tratamento temático da informação” no contexto da trajetória da fundadora da ISKO Ingetraut Dahlberg.

Trazendo a questão para o universo específico da Ciência da Informação, destaca Dahlberg (1995) três fases que vem se desenvolvendo ao longo do tempo: uma fase classificatória original (ou notacional), em que a tônica residia no *organizar para achar* (Dewey, LC, etc.); uma fase tesáurica (ou alfabética), marcada pela ruptura com os sistemas tradicionais de classificação, buscando em referenciais da Linguística soluções para o tratamento temático da informação, e, mais recentemente, uma nova fase classificatória (ou fase classificatória revisitada, como define Dahlberg), em que se reconhece a complementaridade da organização lógica de conceitos e de sua representação lingüística, mormente quando a discussão acerca das ontologias assume maior ênfase. (Guimarães & Dodebei, 2012, p. 13)

O texto citado de Dahlberg (erroneamente citado como de 1995, mas realmente publicado em 1993) mesmo falando de organização do conhecimento não usa a expressão “tratamento temático da informação” (ou nenhuma de suas possíveis variações em inglês) em nenhum momento. Isto pode indicar que os autores estão extrapolando um significado global em uma terminologia própria com o intuito de fundamentar sua proposta terminológica. Neste sentido cabe destacar que o termo “tratamento temático da informação”

(ou inclusive tratamento da informação) não tem um equivalente claro em inglês como resultado desta apropriação. Segundo explica Guimarães (2008, p. 78):

A referida denominação encontrou seu maior impulso notadamente a partir da década de 70, com destaque para a obra de Anthony Charles Foskett, *The subject approach to information*, originalmente publicada em 1969 e que, no Brasil, foi traduzida por A.A. Briquet de Lemos como “A abordagem temática da informação” (FOSKETT, 1973). Tal aspecto surtiu efetivos reflexos, inclusive, no ensino de Biblioteconomia brasileiro, quando se estabeleceram as denominações de sub-áreas *representação temática* e *representação descritiva* no Controle Bibliográfico dos Registros do Conhecimento (em substituição a denominações anteriores como Classificação e Catalogação), no âmbito do Currículo Mínimo de Biblioteconomia de 1982 (BRASIL..., 1982).

De fato, Guimarães (2008) traduziu o termo por “Subject Approach to Information” no *abstract*, enquanto seu equivalente claro em espanhol “tratamiento temático de la información” foi traduzido ao inglês por Mario Barité (2000) como “subject representation” nas *keywords*. Contudo, expressões de difícil sentido em inglês como “thematic treatment of information” têm sido utilizadas até hoje exclusivamente por autores brasileiros em foros nacionais e internacionais, como no último congresso da ISKO Internacional em Portugal. A expressão “thematic treatment of information” tem sido analisada por Smiraglia (2018) como coocorrência de termos em títulos de trabalhos citados traduzidos (usando Google Translate) e aparecendo em um *cluster* sobre políticas de bibliotecas universitárias (aparentemente junto a “documentary reading”, “subject cataloging”, “indexing policy”, “university libraries” e possivelmente “verbal protocol”).

Sobre os demais trabalhos do primeiro volume de ISKO Brasil, no capítulo “Concepções de Análise Documental: Um estudo comparativo a partir dos universos acadêmicos brasileiro e espanhol em Ciência da Informação” (Guimarães & Sales, 2012), o termo é utilizado na primeira frase da introdução citando um texto de Guimarães de 2008: “Em que pese o desenvolvimento teórico e aplicado da área de “tratamento temático da informação” – TTI, observa-se que a dimensão conceitual da área ainda não se encontra efetivamente sedimentada, na medida em que coexistem, de forma nem sempre muito nítida, três correntes teóricas distintas: catalogação de assunto, indexação e análise documental (GUIMARÃES, 2008)” (Guimarães & Sales, 2012, p. 39). Neste texto se liga o termo, sem fazer uma relação explícita, com a área da Análise Documental de matriz francesa baseada nas concepções de Gardin. Os autores continuam dizendo:

O TTI, na Biblioteconomia, apresenta natureza mediadora visto propiciar interlocução entre os contextos de produção e uso da informação, Para tanto, centra-se nas questões atinentes: “à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação” em cujo âmbito desenvolvem-se processos, valendo-se de instrumentos para a geração de produtos (BARITÉ, 1997, p. 124). (Guimarães & Sales, 2012, p. 39)

A referência citada de Mario Barité (outro dos autores do primeiro volume que utilizam a expressão “tratamento temático da informação”, mas em espanhol) está intitulada precisamente “Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica” (e na verdade foi publicado em 1998). Por outro lado, a referência citada de Guimarães de 2008 se intitula “A dimensão teórica do tratamento temático da informação

e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO)”. O texto de Guimarães e Sales (embora não nestes trechos) também cita uma referência de Guimarães (2009) intitulada “Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental”. A continuação, os autores, esclarecem (e repetem) que o TTI, como área de estudos

[...] apresenta-se, historicamente, sob três vertentes teóricas: catalogação de assunto, de matriz norte-americana, mais centrada na geração de produtos; indexação, de matriz inglesa, com maior ênfase nos instrumentos; e análise documental – AD, de matriz francesa, mais preocupada com ênfase no desenvolvimento de referenciais teórico-metodológicos relativos aos procedimentos envolvidos na identificação e representação de conteúdos documentais, em busca de parâmetros que lhes conferissem alguma cientificidade, por meio da explicitação dos procedimentos ou mecanismos lógico-linguísticos envolvidos. (Guimarães & Sales, 2012, p. 39)

Portanto se desprende que o conceito de “tratamento temático da informação” é um constructo desenvolvido no Brasil (e/ou talvez também em países hispano-falantes) que inclui influências de três correntes internacionais, nomeadamente Estados Unidos, Inglaterra e França (portanto não só do livro de Foskett), concluindo que:

[...] as três abordagens teóricas que concebem o arcabouço teórico-metodológico do tratamento temático da informação (*subject catalog*, *indexing* e *analyse documentaire*) coadunam-se não somente na prática profissional, mas também nas descrições teórico-conceituais da área. (Guimarães & Sales, 2012, p. 42)

No capítulo intitulado “Análise documental de obras de ficção: A indexação na perspectiva do percurso gerativo de sentido” de Antonio e Moraes (2012), a expressão “tratamento temático da informação” é utilizada na introdução como objetivo da “Análise Documental” e “[...] para fins de análise do conteúdo temático dos documentos e sua representação (Guimarães, 2003, p. 103)” (Antonio & Moraes, 2012, p. 45). A preferência desta terminologia pelos autores poderia ser explicada pelo fato de ter familiaridade e compartilhar afiliação (a Universidade Estadual Paulista) com o autor citado, José Augusto Guimarães. No capítulo “Ciência da informação: Reflexões sobre a constituição de um campo científico”, de Raquel Barioni Abdalla e Nair Yumiko Kobashi (2012, p. 56), se indica que: “Nessa medida, as linguagens de especialidade (terminologias) são valorizadas de forma crescente na indexação e construção de ferramentas de tratamento da informação (Hjørland, 2008)”. Certamente é difícil identificar no texto citado de Hjørland o que Abdalla e Kobashi se referem por “tratamento da informação” (já que não se trata de uma citação direta), mas pode interpretar-se talvez que as autoras estão utilizando o termo como sinônimo de “organização do conhecimento”.

No capítulo de Mario Barité (2012, p. 265), a expressão “Organización y tratamiento de la información” é utilizada como nome da área de trabalho onde se encaixam os estudos de graduação e pós-graduação em Organização do Conhecimento nos Encontros de Diretores e Docentes de Escolas de Biblioteconomia do Mercosur, realizados entre 1996 e 2006.

Pode concluir-se das publicações do primeiro congresso do capítulo brasileiro de ISKO Brasil que o uso e discussão desta terminologia está baseada principalmente em dois autores: José Augusto Guimarães e Mario Barité, da Universidade Estadual Paulista, no Brasil e da Universidad de La República, no Uruguai, respectivamente.

O segundo volume da série, “Complexidade e Organização do Conhecimento: Desafios do nosso século” (Guimarães & Dodebei,

2013), contempla 46 trabalhos dentre os apresentados no II Congresso de ISKO Brasil no Rio de Janeiro, além de uma introdução. Destes 46 trabalhos, só oito (17%) usam “tratamento temático da informação” ou “tratamento da informação”, e em pelo menos na metade dos casos estas expressões são usadas de forma anedótica e sem ser o foco principal da discussão. Entre os trabalhos que discutem o assunto destacamos dois com autoria da Professora Mariângela Spotti Lopes Fujita (Dal’Evedove, Tartarotti, Fujita & Boccato, 2013; Fujita, 2013), e um trabalho em espanhol de Diana Comesaña e Alicia Díaz (2013). Em todos os casos, estes autores são colegas de instituição de José Augusto Guimarães ou Mario Barité e citam seus respectivos trabalhos.

Dal’Evedove, Tartarotti, Fujita, e Boccato (2013) reforçam a ideia de Guimarães (2008) de que o “tratamento temático da informação” está composto por três diferenciadas ramas (catalogação de assunto, indexação e análise documental) provenientes de três tradições internacionais diferentes que colocam na língua original que:

Embora as três abordagens teórico-metodológicas do Tratamento Temático da Informação (*subject cataloguing, indexing e analyse documentaire*) serem, por vezes, interligadas no que tange ao arcabouço teórico-conceitual e prático-aplicado (Guimarães, 2008), optou-se por investigar apenas a indexação com o propósito de tornar a pesquisa possível operacionalmente. (Dal’Evedove, Tartarotti, Fujita, & Boccato, 2013, pp. 66-67)

Fujita (2013) também desenvolve em mais profundidade estas ideias com base em Guimarães (2008):

Com história recente, a Organização do Conhecimento, em sentido específico, teve origem na institucionalização da ISKO, entretanto, tanto a indexação quanto a classificação vistas como

áreas de atividades da Biblioteconomia e da Ciência da Informação, tem origens mais distantes e estão atreladas, na concepção de Guimarães (2008), ao desenvolvimento do Tratamento Temático da Informação (TTI). (Fujita, 2013, p. 148)

Não enquanto, esta distinção entre Organização do Conhecimento e os processos de catalogação de assunto, indexação e análise documental que supostamente conformam o “tratamento temático da informação” não está clara na literatura internacional de diversos países como Dinamarca ou Estados Unidos. Por um lado, segundo Hjørland (2016), a história da Organização do Conhecimento como disciplina tem uma origem anterior à institucionalização da ISKO, estando situada sobre o 1900 com o trabalho de Charles A. Cutter, W. C. Berwick Sayers e Ernest Cushing Richardson. Por outro lado, considerando que a Organização do Conhecimento está composta tanto de sistemas de organização do conhecimento como de processos de organização do conhecimento (incluindo a catalogação, indexação, classificação, etc.), Hjørland também aponta que alguns aspectos que estão incluídos na história da Organização do Conhecimento são a história destes sistemas e a história dos processos (também no contexto da Biblioteconomia). Portanto, resulta difícil sustentar com base na literatura internacional a distinção entre o conceito de organização do conhecimento e outro conceito diferenciado que possa incluir os mesmos processos como a catalogação e indexação. Mesmo assim, Fujita continua explicando as características do “tratamento temático da informação” com base em Guimarães (2008):

O Tratamento Temático da Informação é parte das atividades e operações do tratamento da informação que envolve o conhecimento teórico e metodológico disponível quanto ao tratamento descritivo do suporte material da informação e ao tratamento temático de conteúdo da informação. Esta dicotomia que se

apresenta no tratamento da informação é explicada, de um lado, pelo desenvolvimento teórico e metodológico distinto alcançado pelas duas áreas e de outro, pela diferença existente entre os aspectos da informação – o material e o conteúdo, que exigem tratamento diferenciado.

O tratamento temático do documento realiza uma descrição do conteúdo mediante representação condensada com vistas à sua acessibilidade temática e, nesse sentido, as operações de elaboração de resumos, classificação, catalogação de assuntos e indexação possuem graus de condensação nas representações documentais (verbais, não verbais ou codificadas), de modo que um resumo tem menor grau de condensação que palavras-chaves, cabeçalhos de assunto ou um número de classificação. (Fujita, 2013, p. 149).

Neste comentário se revela a visão de que o “tratamento temático da informação” é considerado um subconjunto dicotômico de outro conceito maior chamado de Tratamento da Informação, sendo que o “tratamento temático da informação” aborda as questões temáticas de conteúdo e o outro subconjunto as questões descritivas do suporte material. De novo, esta dicotomia não é reconhecida de forma tão clara nos principais trabalhos internacionais (por exemplo, no citado trabalho de Hjørland, 2016, ou nos principais manuais sobre Organização do Conhecimento/Informação como Taylor & Joudrey, 2009).

No trabalho de Comesaña e Díaz (2013, p. 189) parece-se indicar que o “tratamento temático da informação” é sinônimo da Organização do Conhecimento:

Si entendemos la ORC como el “área de Tratamiento temático de la información que persigue la identificación, el procesamiento y la disponibilidad del contenido informacional de los documentos” (Guimarães (1998) en Barité, 2001: 38), queda implícita que

la función última de la ORC es permitir recuperar la información en sus dos dimensiones: el acceso y la apropiación del contenido.

Outra interpretação deste trecho seria que a Organização do Conhecimento é uma área ou domínio enquanto o tratamento temático da informação é simplesmente como um conjunto de processos dedicados à análise de conteúdo.

No terceiro volume da série, “Organização do Conhecimento e Diversidade Cultural” (Guimarães & Dodebei, 2015), se incluem 81 trabalhos do III Congresso de ISKO-Brasil celebrado na Universidade Estadual Paulista, Marília, incluindo a apresentação ao volume e os resumos de trabalhos selecionados para publicação em revista. Destas 81 referências, só 20 trabalhos (24,7%) usam a expressão “tratamento da Informação” ou “tratamento temático da informação”. Embora muitos destes trabalhos não discutam os conceitos especificamente, eles frequentemente assumem sua existência e reforçam as dicotomias apresentadas anteriormente. Por outro lado, é importante ressaltar que este terceiro volume foi traduzido oficialmente ao inglês em 2017 (Guimarães & Dodebei, 2017) incluindo diversas e incoerentes traduções dos termos. Por exemplo, Araujo e Crippa (2015, p. 37) listam uma série de trabalhos “sobre o tratamento descritivo e temático da informação”, sendo traduzido o trecho no volume em inglês por “descriptive and thematic processing of information”. Outras ocorrências das expressões “tratamento da Informação” e “tratamento temático da informação” foram traduzidas por “information processing”, “processing of information”, “thematic processing of information”, “thematic treatment of information”, “information treatment”, “thematic information treatment”, “information subject treatment” e “subject information treatment” (as duas últimas expressões de forma rara).

Um exemplo desta problemática terminológica é ilustrado no trabalho de Sales (2015), onde se indica que:

A primeira década do século XXI revelou a predominância de uma perspectiva que define a OC como um “fazer” de natureza “operacional” (GARCIA, OLIVEIRA, LUZ, 2000; GREEN, 2002; GÁRCIA GUTIÉRREZ, 2002), cujos objetos de investigação são, principalmente, os conceitos e as estruturas conceituais (KENT, 2000; GREEN, 2002; OHLY, 2008, SMIRAGLIA, 2010), formalizadas instrumentalmente nos sistemas de organização do conhecimento – SOC (KENT, 2000; GREEN, 2002; ZHEREBCHEVSKY, 2010; SOUZA, TUDHOPE e ALMEIDA, 2010). Nesse sentido, percebemos uma OC ligada às atividades atinentes à Ciência da Informação e à Biblioteconomia, mais especificamente ao tratamento temático da informação (FOSKETT, 1973). (Sales, 2015, p. 74).

A expressão “tratamento temático da informação” ligada a Foskett e apresentada na tradução oficial (Guimarães & Dodebei, 2017, p. 70) como “information processing”, possivelmente por descuido e pelo fato de que o trabalho de Foskett, listado nas referências, é a tradução para o português intitulado “A abordagem temática da informação” (Foskett, 1973). Mesmo assim, é importante resenhar que o título da versão original do livro de Foskett (1969) é “The subject approach to information” e no seu conteúdo não se mencionam as expressões “information processing”, “thematic treatment” ou outra semelhante que possa justificar esta escolha terminológica. Na literatura internacional em inglês, para referir aos processos de análise de assunto na mesma linha que o trabalho de Foskett, a terminologia utilizada predominante seria “subject analysis” (Taylor & Joudrey, 2009; Hjørland, 2016), expressão que poderia ter sido traduzida ao português como “análise de assunto”.

Posteriormente, Sales também discute as relações entre a Organização do Conhecimento, Ciência da Informação e Biblioteconomia, identificando três correntes internacionais baseadas respectivamente em Foskett, Hjørland e Dahlberg. Neste sentido se indica que a

tradição fundamentada em Foskett, própria da Biblioteconomia e da Ciência da Informação, “[...] prefere abordar a OC como fundamentalmente discutida no âmbito do tratamento temático da informação” (Sales, 2015, p. 77). Esta frase estaria alinhada com a afirmação citada anteriormente de Comesaña e Díaz (2013) na qual se considera a Organização do Conhecimento como uma parte do “tratamento temático da informação”.

Alves e Moraes (2015) discutem “A dimensão epistemológica da análise documental de conteúdo de obras de ficção na organização do conhecimento” e em diversas ocasiões qualificam o Tratamento Temático da Informação e a Organização do Conhecimento como duas áreas diferentes (por exemplo, na introdução, colocando-as portanto no mesmo nível):

Esses aspectos epistemológicos incorporados à Análise Documental de Conteúdo colaboram para seu embasamento teórico-metodológico e desenvolve áreas em que mantém interlocução teórica, como o Tratamento Temático da Informação, Linguística Documental, Textologia Documental e Organização do Conhecimento. (p. 117)

Posteriormente, Alves e Moraes (2015) também aprofundam e discutem o conceito de “tratamento temático da informação” com base em Guimarães:

[...] conforme Guimarães (2008, 2009), [a Análise Documental de Conteúdo] consiste em corrente teórica do Tratamento Temático da Informação, de origem francesa, teve início no final da década de 1960 com estudos de Gardin e Coyaud, também é denominada de Análise Documental (*analyse documentaire*). Segundo Guimarães et al. (2012), a Análise Documental de Conteúdo na França e no Brasil designa o Tratamento Temático da Informação,

enquanto que na Espanha se insere no Tratamento Temático (descrição de forma e conteúdo). (p. 118)

Neste trecho e posteriores se esclarece que, segundo Guimarães, o “tratamento temático da informação” tem origem na França e seria o mesmo que a Análise Documental/*Analyse Documentaire* (confirmando, portanto, a não ligação direta e exclusiva com a tradição de Foskett). As outras duas partes do conceito citadas anteriormente contribuiriam externamente, segundo a narrativa de Guimarães, para sua consolidação teórica e metodológica:

O Tratamento Temático da Informação, segundo Guimarães (2008), contém mais duas correntes teóricas em sua base epistemológica, a catalogação de assunto (subject cataloguing) norte-americana e a indexação (indexing) de origem inglesa, que contribuem para sua consolidação teórica e metodológica.

Por exemplo, nos Estados Unidos e Inglaterra os estudos foram voltados para produtos e instrumentos de informação, e na França houve a “[...] preocupação com o desenvolvimento de referenciais teórico-metodológicos para o *processo* de TTI propriamente dito”, pois, os procedimentos até então utilizados eram guiados apenas pelo bom senso (Guimarães, 2009, p. 109). Nesse contexto:

[...] tratamento temático da informação estuda a descrição e representação de forma e conteúdo dos documentos para recuperação, se insere no Tratamento da Informação, uma subárea da Biblioteconomia e Documentação que mantém interlocução teórica com a Organização do Conhecimento (BARITÉ, 2001). (Alves & Moraes, 2015, p. 118-119).

Na tradução oficial de todos estes trechos o termo utilizado é “Thematic Treatment of Information” (Guimarães & Dodebei, 2017), talvez com intuito de afastar-se e afastar a genealogia do conceito

da terminologia e tradição anglo de “subject”. Aqui é interessante remarcar que mesmo no contexto de ISKO Brasil nenhuma tradução ao inglês tem usado este termo. Em outros contextos, Barité tem usado a tradução “subject representation” muito mais em conexão com a tradição e realidade internacional. Por outro lado, se deve enfatizar de novo que a expressão “thematic treatment”, embora sendo uma escolha terminológica deliberada e pessoal, não faz muito sentido em inglês, como demonstrado em Smiraglia (2018) e comunicações pessoais.

A dicotomia entre “tratamento temático da informação” e tratamento descritivo da informação, como subconjuntos de um conceito mais amplo de tratamento da informação, é dissentida por Lima e Cunha (2015), que entendem, com base em outros autores, que o “tratamento temático da informação” inclui ambos os processos de descrição física e de conteúdo e é parte de um conjunto maior denominado ciclo documentário:

O tratamento temático da informação pode ser entendido como o processo que envolve a descrição física e de conteúdo dos objetos informacionais, o produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico (BRASCHER; CAFÉ, 2008).

O tratamento temático da informação é responsável pela mediação entre o produtor e o usuário informação, exercendo a tarefa fundamental de dar acesso ao conteúdo documental (NARUKAWA; SALES 2012, p. 154). Esse tratamento é parte de um conjunto de operações denominado de ciclo documentário. (Lima & Cunha, 2015, p. 274).

Esta visão estaria fundamentada, já não tanto em Guimarães, mas no trabalho de Brascher e Café (citado como 2008 mas na realidade

publicado em 2010). No entanto, o trabalho de Brascher e Café (de uma alta qualidade e interesse) não usa a expressão “tratamento temático da informação” em nenhum momento, pelo que a escolha terminológica inicial corresponde unicamente com o critério de Lima e Cunha. Por outra parte, Marisa Brascher também contribui com um capítulo em coautoria (Medeiros, Vital, Linden, & Brascher, 2015) no qual se fala de “tratamento descritivo e temático” (p. 498), e, de forma mais relevante, de “representação de assunto” e “tratamento de assunto” (no contexto da autora canadense francófona de Quebec, Louise Gagnon-Arguin), mostrando uma maior coerência com a corrente internacional inclusive de influência francesa. No entanto, na tradução ao inglês deste trabalho (Guimarães & Dodebei, 2017, p. 476) foram utilizados os termos “subject representation” e “treatment of subject” respectivamente.

A visão de Guimarães e Barité do “tratamento temático da informação” como subárea da Organização do Conhecimento é repetida de novo por Tartarotti, Dal’Evedove, e Fujita (2015), autoras que possuem, como já mencionado, algum tipo de vinculação com a Universidade Estadual Paulista:

No âmbito da Organização e Representação do Conhecimento (ORC), destaca-se o Tratamento Temático da Informação (TTI) como uma subárea fundamental ao lidar com as questões acerca da “análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação” (BARITÉ, 1997, p. 124). Guimarães (2009, p. 106) esclarece que em busca de uma fundamentação teórica e estabelecimento como campo de pesquisa, historicamente o TTI apresenta-se sob três correntes teóricas: a análise documental (*analyse documentaire*, de concepção francesa); a catalogação de assunto (*subject cataloguing*, de concepção norte-americana) e a indexação (*indexing*,

de concepção inglesa). (Tartarotti, Dal'Evedove, & Fujita, 2015, p. 329)

Na tradução ao inglês deste trabalho, curiosamente, se usa a expressão “Information Subject Treatment”, sendo o único capítulo do livro em que se utiliza desta expressão.

No trabalho de Pinho, Nascimento e Marinho (2015), sobre música, literatura e audiovisual, a visão de Guimarães se sumariza de forma bastante clara:

O tratamento da informação envolve duas especialidades identificadas como: tratamento descritivo e tratamento temático. O primeiro consiste em identificar os elementos descritivos de um documento, tais como: autor, título, ano de publicação, etc. O segundo recai no propósito de identificar o documento a partir do seu conteúdo temático, ou seja, o assunto (p. 625).

Não entanto, na sua versão em inglês, o texto se expressa em termos mais bizarros:

Information treatment involves two specialties identified as: descriptive treatment and thematic treatment. The first aims to identify the descriptive elements of a document, such as: author, title, year of publication, etc. The second aims to identify the document from its thematic content, that is, the subject (Guimarães & Dodebei, 2017, p. 591).

Nesta tradução chama a atenção o uso do termo “treatment” em três ocasiões/conceitos diferentes. Deve também ser ressaltado que Fabio Pinho, o primeiro autor do texto, é doutor pela Universidade Estadual Paulista e foi orientando de José Augusto Guimarães.

Gercina Lima, em seu estudo da Organização do Conhecimento no Brasil (2015), sumariza a narrativa do artigo de Guimarães (2008) no seguinte trecho:

Também no ano de 2008, Guimarães apresentou o artigo intitulado “A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da *International Society for Knowledge Organization* (ISKO)”, em que faz uma análise da interlocução da organização do conhecimento, dentro do universo da ISKO, a partir das três correntes: catalogação de assunto, de influência norte americana; indexação, de influência inglesa; e análise documentária, de influência francesa. Dentro desse contexto, o autor faz uma análise das pesquisas e estudos apresentados no GT2 da ANCIB, e do capítulo brasileiro da ISKO, e conclui que os estudos realizados nas escolas brasileiras de BCI, nesta área, têm apontado para uma equidade na realização de estudos e pesquisas, sem perder de vista as abordagens internacionais. Como necessidade de desenvolvimento na área, Guimarães (2008) aponta: (1) necessidade de reflexão sobre os aspectos teóricos e metodológicos que contemplem a dimensão interdisciplinar sem perder seu foco principal de tratar a informação para se ter uma recuperação eficiente; (2) reflexão sobre a atuação profissional diante dessa necessidade, e (3) sugere que dentro do contexto da ISKO não se deve perder de vista como a área de Tratamento Temático da Informação (TTI) vem construindo seu referencial teórico-metodológico (Lima, 2015, p. 681).

O problema desta narrativa em relação ao “tratamento temático da informação” e a ISKO Brasil é que em 2008, quando o artigo de Guimarães foi publicado, ainda não se tinha apresentado nenhuma pesquisa ou estudo no contexto de ISKO Brasil propriamente dito

(somente no GT2 do ENANCIB), já que o primeiro congresso do capítulo brasileiro de ISKO foi organizado em 2011. De fato, no estudo de Guimarães poderia inclusive interpretar-se uma equivalência quase absoluta entre os termos “tratamento temático da informação” e organização do conhecimento no relativo à história da ISKO, sinalando portanto a emergência e origem do discurso. Este trabalho de Guimarães (2008) demonstrou-se extremamente influente entre os autores de ISKO Brasil (por exemplo, Bufrem, 2015), pelo que se pode aceitar esta análise não como um reflexo da pesquisa publicada até esse momento e sim como algo programático. Neste sentido, se poderia interpretar a composição das três áreas do “tratamento temático da informação” como uma confirmação, preferência e reconhecimento da influência francesa de Gardin (como, por exemplo, em Smit, 2015) na proposta de Guimarães.

O quarto volume dos congressos de ISKO Brasil, “Memória, tecnologia e cultura na organização do conhecimento” (Pinho & Guimarães, 2017), inclui 16 de 46 trabalhos (34,78%) que usam ou discutem as expressões “tratamento da informação” ou “tratamento temático da informação”. Desta forma, os números de trabalhos totais que usam estas expressões mostram um aumento desde o primeiro ano, mas um decréscimo no último ano (Figura 1), enquanto a quantidade em porcentagens de trabalhos gerais mostra um decréscimo no segundo ano mas um crescimento nos dois últimos anos (Figura 2).

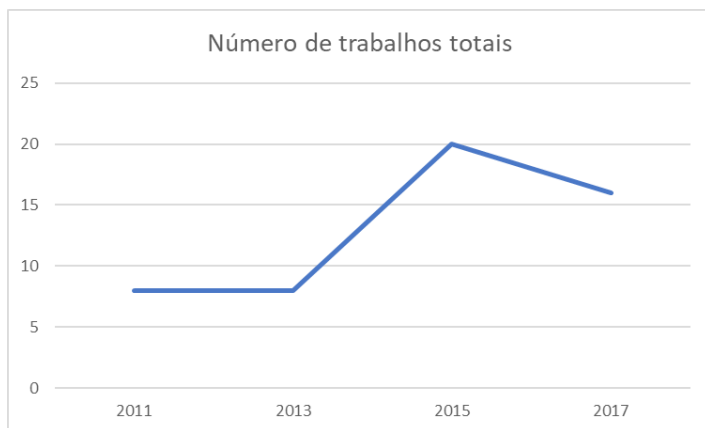


Figura 1: Número de trabalhos totais que usam as expressões analisadas

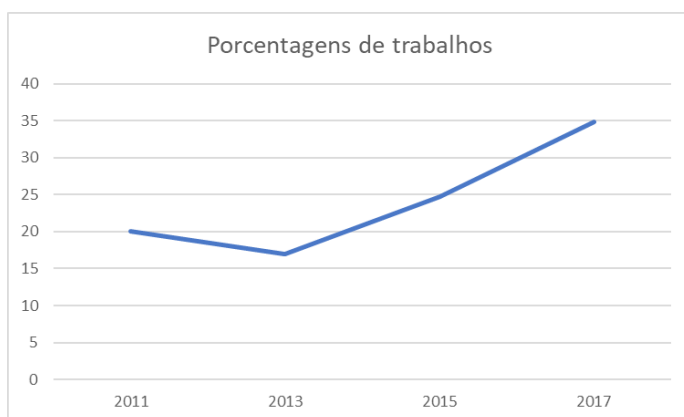


Figura 2: Porcentagens de trabalhos nos congressos que usam as expressões analisadas

No que se refere ao conteúdo deste quarto volume, diversos trabalhos (muitos deles de autores com afiliação na Universidade Estadual Paulista ou que se formaram nela) utilizam ou definem a expressão como parte da introdução ou conceitos preliminares, frequentemente citando os trabalhos de Guimarães (2008, 2009). Rodrigo de Sales (2017) é um exemplo de trabalho relevante na temática que continua na linha do seu trabalho de 2015. Felipe e

Pinho (2017), por outro lado, colocam a ideia de que o “tratamento temático da informação” é parte da Organização do Conhecimento (“A descrição temática do conteúdo dos documentos está ligada ao Tratamento Temático da Informação, aspecto da Organização do Conhecimento, no qual o conhecimento é organizado por meio de atividades de análise e descrição do conteúdo do mesmo”, p. 148), contradizendo a visão citada anteriormente de que a Organização do Conhecimento é uma parte do Tratamento Temático da Informação.

Neste quarto volume se revelam também ideias que se escapam da narrativa original de Guimarães, como por exemplo, a preferência de uso do termo “tratamento informacional” quando se refere a aspectos mais tecnológicos (por exemplo, Ninin & Simionato, 2017, e Sá & Campos, 2017) e o uso da expressão “tratamento temático da informação” como metodologia (por exemplo, Marinho & Santos, 2017, e Bezerra & Santos, 2017). No trabalho de Marinho e Santos, mesmo definindo o conceito de uma maneira heterodoxa na linha de Guimarães (“O tratamento da informação envolve duas especialidades identificadas como tratamento descritivo e tratamento temático. O primeiro consiste em identificar os elementos descritivos de um documento, tais como: autor, título, ano de publicação, etc. O segundo recai no propósito de identificar o documento a partir do seu conteúdo temático, ou seja, o assunto” (p. 319)), também apresenta-o como método (“No âmbito da Organização da Informação (OI), é possível estabelecer relações de significação referentes ao domínio do ciclo junino, que se constitui a partir de elementos representativos e simbólicos por meio das técnicas de análise de assunto, análise de domínio e tratamento temático da informação” (p. 321)).

No segundo trabalho citado nesta linha, também de Raimundo Nonato Macedo dos Santos da Universidade Federal de Pernambuco, se apresenta o conceito na seção de metodologia nos seguintes termos: “A Análise de Assunto (AA) e o Tratamento Temático da

Informação (TTI), sob o viés da Organização da Informação, técnicas metodológicas utilizadas para proceder a extração de termos ou “microestruturas” das letras das músicas” (Bezerra & Santos, 2017, p. 345). Posteriormente, também é explicado o método sem se afastar da visão de Guimarães:

Para Dias e Naves (2013, p. 7): “[...] o tratamento temático tem uma forte carga subjetiva pois, como o nome indica, visa caracterizar o documento do ponto de vista do seu conteúdo”. O TTI, envolve três etapas principais: a análise de assunto; a extração / seleção de conceitos e a determinação do assunto.

Estas três etapas correspondem a um procedimento cognitivo e muitas vezes, de caráter intuitivo, faz parte do ciclo de operações documentais, conforme aponta Guimarães (2008, p. 79):

[...] o TTI integra aquilo que se convencionou denominar ciclo de operações documentais, uma vez que ocupa posição intermediária entre a coleta e a difusão de documentos e, destarte, caracteriza-se por atividade de processamento, tanto sob a ótica do suporte material – tratamento descritivo – quanto do conteúdo – tratamento temático.

O TTI refere-se a fase essencial do processo da análise documental, realizada na pesquisa aqui apresentada. (Bezerra & Santos, 2017, p. 346).

Talvez esta visão metodológica seja propiciada e pela decorrência da consideração do conceito de “tratamento temático da informação” não tanto como uma área mas como um processo.

Embora, até a data de hoje, não exista uma tradução oficial do volume ao inglês como no caso do anterior congresso, as traduções proporcionadas pelos autores nos trabalhos foram as seguintes: “informational treatment” (tradução de “tratamento informacional” no *abstract*, em dois trabalhos, p. 170 e p. 297), “process of thematic

treatment of the documents” (tradução de “processo de tratamento temático dos documentos” no *abstract* de um trabalho, p. 217), “Thematic Treatment of Information” (tradução de “Tratamento Temático da Informação” no título e *keywords* de um trabalho, p. 341) e “Tematic Treatment of Information” (tradução de “Tratamento Temático da Informação”, no *abstract* e *keywords*, p. 162). Este último trabalho, de Cândido et al. (2017), mostra não só o uso de um termo (“treatment”) que não faz sentido em inglês neste contexto, mas também uma palavra (“tematic”) em duas ocasiões que nem sequer existe em inglês. Estas traduções são um reflexo do sentido que tem sido dado ao termo “Tratamento Temático da Informação” entre a comunidade discursiva da ISKO-Brasil, o qual se afasta do espírito original da obra de Foskett (1973) “The subject approach to information” e também das traduções originais de Guimarães (2009) (“subject approach to information”) e Barité (2000) (“subject representation”). Esta última conclusão pode ver-se refletida no seguinte trecho de Raimunda Fernanda dos Santos e Renato Fernandes Corrêa:

Guimarães (2009, p. 1) afirma que no âmbito da organização da informação dois universos se descortinam: “o primeiro, ligado ao acesso aos documentos e o segundo, de natureza mais complexa, voltado para o acesso ao conteúdo informacional, genericamente denominado de tratamento temático da informação” ou representação temática da informação, como é designada na literatura da área de organização da informação.

Nessa perspectiva, verifica-se que o documento deve ser tratado em seu todo levando em consideração tanto as operações que contemplam a descrição física (tratamento descritivo ou representação descritiva), como a descrição do conteúdo temático (tratamento temático ou representação temática) haja vista que a diferenciação entre a forma e o conteúdo “[...] reside na busca do

o que (materialização) e do sobre o que (teor) que convivem no âmbito do documento” (GUIMARÃES, 2009, p. 1).

Nesse entendimento, a representação temática da informação centra-se nas questões relativas ”à análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação (BARITÉ, 1999, p. 124). (Santos & Corrêa, 2017, p. 251).

Neste significativo trecho se revela primeiro, o reconhecimento da expressão “tratamento temático da informação” (citada no trecho de Guimarães) como sinônimo da expressão portuguesa “representação temática” (utilizada, segundo os autores, na literatura sobre organização da informação e sendo, sem dúvida, uma tradução mais direta na linha de Foskett e outros autores internacionais); segundo, se revela o reconhecimento implícito da expressão “tratamento temático” como sinônimo de “organização da informação”, já que se indica que tanto a representação temática como a representação descritiva são os dois universos que conformam a organização da informação. Sobre este último aspecto, não existe consenso entre a comunidade de ISKO Brasil tal como foi observado anteriormente e nos seguintes trechos de outros trabalhos do mesmo volume que também se baseiam em Guimarães e Barité (e com afiliação na Universidade Estadual Paulista):

A denominada Organização e Representação do Conhecimento (ORC), a OC constitui uma vertente teórica do Tratamento Temático da Informação (TTI), que por sua vez dialoga com outras vertentes como a catalogação de assunto (de matriz estadunidense), a indexação (de matriz inglesa) e a análise documental (de matriz francesa), constituindo, destarte, um domínio de conhecimento (Guimarães, 2008). (Almeida, Grácio, & Castanha, 2017, p. 334);

e:

Já no contexto profissional, a Organização do Conhecimento preocupa-se em apresentar “[...] subsídios teóricos (e realimentar-se) com tudo que é relativo ao tratamento da informação, particularmente com o tratamento temático da informação, e de um modo menos específico - mas não menos importante - com a gestão do uso social da informação” (BARITÉ, 2001, p. 38, tradução nossa). (Casarin, Ferreira, & Milani, 2017, p. 362).

Como conclusão desta seção podemos indicar que a terminologia utilizada em relação ao “tratamento temático da informação” (e o tratamento da informação) na comunidade de ISKO Brasil está muito influenciada pela obra e narrativa de Guimarães (e talvez em menor medida de Barité), que ao mesmo tempo tentou assimilar as influências francesas de autores como Gardin na corrente internacional de outros autores internacionais.

3. Conclusões

Este trabalho buscou identificar, descrever e analisar as diferentes definições e relações aferidas ao “tratamento temático da informação”, no contexto da Organização do Conhecimento e da Informação, na Ciência da Informação brasileira. Longe de se propor a identificar ou construir uma univocidade em sua definição, almejou-se reconhecer os movimentos, as nuances, as fronteiras e as origens que desenham a concepção do Tratamento da Informação, em especial o “tratamento temático da informação”, na literatura científica, tendo como propósito identificar suas origens internacionais, sua identidade nacional e seus atuais agentes epistêmicos, entendendo

que estes elementos são decisivos para estabelecermos análises e prospecções sobre o tema.

A conclusão mais geral que pode ser feita em um primeiro momento é a de que há uma consolidação científica em construção sobre o assunto. A criação do capítulo ISKO Brasil, também é uma ação concreta e efetiva que demonstra e confirma o fortalecimento e o crescimento de produção do conhecimento sobre Organização do Conhecimento e da Informação especialmente no Brasil.

Quanto às variações terminológicas identificadas, no bojo dos trabalhos analisados, para denominar as ações que compõe os processos de “tratamento temático da informação”, pode-se confirmar um comportamento já reconhecido em outros momentos, de que não há consenso sobre o uso destas expressões. Assim, Tratamento da Informação, Representação da Informação, Representação de Assunto, Tratamento de Assunto, Análise de Assunto são expressões frequentemente utilizadas para nomear, com algumas nuances, o que se reconhece como Tratamento Temático na Ciência da Informação brasileira. No entanto, há uma predominância no uso da expressão “tratamento temático da informação” nesta literatura, inclusive considerando-a como uma das ações de Tratamento da Informação.

O agente epistêmico que se destaca, na construção da definição sobre o assunto, é José Augusto Chaves Guimarães, vinculado a Universidade Estadual Paulista. Os trabalhos do autor foram os mencionados com mais frequência nos trabalhos que investigamos. Em especial, suas publicações de 2008 e 2009, intituladas respectivamente “A dimensão teórica do Tratamento Temático da Informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO)” e “Abordagens teóricas do Tratamento Temático da Informação (TTI): catalogação de assunto, indexação e análise documental”, são as mais frequentemente citadas.

A definição mais recorrente nos trabalhos que se dispuseram a descrever e conceituar o Tratamento Temático é deste modo, a proposta por Guimarães (2009), em especial, a que elucida e distingue a catalogação de assunto, a indexação e análise documental, sendo a primeira de origem norte-americana, a segunda de ordem inglesa e a terceira francesa. A partir desta distinção, as práticas que compõem este Tratamento, estariam diretamente relacionadas a análise documental. A Organização do Conhecimento e da Informação, embora frequentemente mencionados nos trabalhos sobre o tema enquanto campo maior que acomoda estas práticas, não foram necessariamente e suficientemente definidas neste capítulo. Alguns posicionamentos atrelam o Tratamento Temático a Organização da Informação, e outros, a do Conhecimento. No entanto, definir e distinguir estas disciplinas extrapolariam o contexto deste nosso capítulo. Sugerimos como possibilidade de leitura e compreensão maior sobre esta distinção, o trabalho de Pando (2018).

Enfim, reconhecemos os limites que circundam trabalhos desta natureza metodológica, uma vez que há, ainda, toda uma literatura nacional (artigos de revistas científicas, teses e dissertações) e internacional que merecem ser investigadas para solidificar as indicações que fizemos. O movimento desenvolvido neste capítulo, a partir dos recortes estabelecidos, teve como intuito geral estabelecer um reconhecimento de conceitos e atores que estão agindo e produzindo conhecimento sobre o tema, em especial no Brasil, sendo possível arriscar a sinalizar que neste país já haveria uma construção relativamente sólida sobre o Tratamento Temático, que por sua vez potencializa a construção de novos conceitos e de novas práticas que operacionalizam e fundamentam o “tratamento temático da informação” de modo global.

Referências Bibliográficas

- Abdalla, R. B., & Kobashi, N. Y. (2012). Ciência da informação: Reflexões sobre a constituição de um campo científico. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (pp. 54-58). Marília: ISKO-Brasil: FUNDEPE.
- Almeida, C. C. de, Grácio, M. C. C., & Castanha, R. G. (2017). Colaboração Científica na Temática Ética em Organização e Representação do Conhecimento: Um Estudo no Periódico Knowledge Organization. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 334-340). Recife, PE: Ed. UFPE.
- Alves, R. C. V., & Moraes, J. B. E. (2015). A dimensão epistemológica da análise documental de conteúdo de obras de ficção na organização do conhecimento. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 117-124). Marília: ISKO-Brasil: FUNDEPE.
- Antonio, D. M., & Moraes, J. B. E. (2012). Análise documental de obras de ficção: A indexação na perspectiva do percurso gerativo de sentido. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (pp. 45-48). Marília: ISKO-Brasil: FUNDEPE.
- Araujo, A. V. F., & Crippa, G. (2015). A dimensão histórico-epistemológica da Organização do Conhecimento: contributos da taxonomia Gesneriana, Séc. XVI. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 37-46). Marília: ISKO-Brasil: FUNDEPE.
- Barité, M. (1998). *Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica*. Porto Alegre: ABEED.
- Barité, M. (2000). Los conceptos y su representación: Una perspectiva terminológica para el tratamiento temático de la información. *Scire*, 6(1), 31-53. Recuperado de <https://www.iberid.es/ojs/index.php/scire/article/view/1123>
- Barité, M. (2001). Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In K. Carrara (Ed.), *Educação, universidade e pesquisa* (pp. 35-60). Marília: UNESP.
- Barité, M. (2012). La producción científica en Organización del Conocimiento vista a través de la Sección Literatura de la revista Knowledge Organization: invisibilidad y transparencia. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (pp. 262-272). Marília: ISKO-Brasil: FUNDEPE.
- Bezerra, E. M. B. L., & Santos, R. N. M. (2017). A Música, Informação e Política: Tratamento Temático da Informação na Produção Musical de Chico Buarque no Período do AI-5 (1968-1978). In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 341-350). Recife, PE: Ed. UFPE.
- Brascher, M., & Café, L. (2010). Organização da Informação ou Organização do Conhecimento?. In M. L. G. de Lara & J. W. Smit (Orgs.), *Temas de pesquisa em Ciência da Informação no Brasil* (pp. 85-103). São Paulo: Escola de Comunicações e Artes da Universidade de São Paulo.

- Bufrem, L. S. (2015). Perspectivas da pesquisa sobre organização do conhecimento no Brasil. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 709-724). Marília: ISKO-Brasil: FUNDEPE.
- Cândido, G. G., Barros, T. H. B., & Redigolo, F. M. (2017). Indexação e Descrição Arquivística: Relações Histórico-conceituais. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 162-169). Recife, PE: Ed. UFPE.
- Casarin, H. C. S., Ferreira, T. G., & Milani, S. O. (2017). As Crianças como Comunidades de Usuários e os Sistemas de Organização do Conhecimento: Uma Análise da Literatura. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 361-367). Recife, PE: Ed. UFPE.
- Comesaña, D., & Díaz, A. (2013). Análisis de Dominio Como Abordaje Teórico en la Construcción de Ontologías. Estudio de Caso, Desarrollo de Ontologías Para la IDE-Uruguay. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Complexidade e organização do conhecimento: desafios de nosso século* (pp. 189-194). Marília: ISKO-Brasil: FUNDEPE.
- Dahlberg, I. (1993). Current trends in knowledge organization. In F. J. García Marco (Org.), *Organización del conocimiento en sistemas de información y documentación* (pp. 7-26). Zaragoza: Universidad de Zaragoza.
- Dal'Evedove, P. R., Tartarotti, R. C. D., Fujita, M. S. L., & Boccato, V. R. C. (2013). A produção científica na temática indexação: Análise bibliométrica de teses e dissertações no período de 2002 a 2011. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Complexidade e organização do conhecimento: desafios de nosso século* (pp. 66-72). Marília: ISKO-Brasil: FUNDEPE.
- Dias, E. W., & Naves, M. M. L. (2007). *Análise de assunto: teoria e prática*. Brasília: Thesaurus.
- Felipe, C. B. M., & Pinho, F. A. (2017). Mapeamento da Produção Brasileira sobre Indexação de Imagens. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 146-153). Recife, PE: Ed. UFPE.
- Foskett, A. C. (1973). *A abordagem temática da informação* (A. A. Briquet de Lemos, trad.). São Paulo: Polígono; Brasília: Ed. UnB.
- Fujita, M. S. L. (2013). A importância teórica e prática da indexação na fundamentação científica da organização e representação do conhecimento. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Complexidade e organização do conhecimento: desafios de nosso século* (pp. 147-159). Marília: ISKO-Brasil: FUNDEPE.
- Gomes, P. H. C., Evangelista, I. V., Martínez-Ávila, D., & Grácio, M. C. C. (2017). A Dimensão Cultural da Organização do Conhecimento: Análise das Comunidades Epistêmicas a partir dos Congressos da ISKO-Brasil. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 383-393). Recife, PE: Ed. UFPE.
- Guimarães, J. A. C. (2008). 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*, 1(1), 77-99. doi:10.26512/rici.v1.n1.2008.940
- Guimarães, J. A. C. (2009). Abordagens teóricas do Tratamento Temático da Informação (TTI): catalogação de assunto, indexação e análise documental. *Ibersid: Revista*

- de Sistemas de Información y Documentación*, 3, 105-117. Recuperado de <https://www.iberid.eu/ojs/index.php/iberid/article/view/3730>
- Guimarães, J. A. C., & Dodebei, V. (Orgs.) (2012). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade*. Marília: ISKO-Brasil: FUNDEPE.
- Guimarães, J. A. C., & Dodebei, V. (Orgs.) (2013). *Complexidade e organização do conhecimento: desafios de nosso século*. Marília: ISKO-Brasil: FUNDEPE.
- Guimarães, J. A. C., & Dodebei, V. (Orgs.) (2015). *Organização do Conhecimento e Diversidade Cultural*. Marília: ISKO-Brasil: FUNDEPE.
- Guimarães, J. A. C., & Dodebei, V. (Orgs.) (2017). *Knowledge Organization and Cultural Diversity*. Marília: ISKO-Brasil.
- Guimarães, J. A. C., & Sales, R. (2012). Concepções de Análise Documental: Um estudo comparativo a partir dos universos acadêmicos brasileiro e espanhol em Ciência da Informação. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (pp. 39-44). Marília: ISKO-Brasil: FUNDEPE.
- Guimarães, J. A. C., Martínez-Ávila, D., Tognoli, N. B., & Milani, S. O. (2018). Presence of Francophone Pioneers in the Brazilian Authors of Knowledge Organization. In W. M. El Hadi (Org.), *Fondements épistémologiques et théoriques de la science de l'information-documentation: Actes du 11e colloque ISKO France 2017, 11 et 12 juillet 2017, Siege de l'Unesco, Paris* (pp. 156-166). London: ISTE Editions.
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(3/2), 86-101. doi:10.5771/0943-7444-2008-2-3-86
- Hjørland, B. (2016). Knowledge organization (KO). *Knowledge Organization*, 43(6), 475-484. doi:10.5771/0943-7444-2016-6-475
- Lima, G. A. (2015). Organização do conhecimento: pesquisa e desenvolvimento. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 670-687). Marília: ISKO-Brasil: FUNDEPE.
- Lima, M. F. S., & Cunha, F. A. P. (2015). As linguagens documentárias na descrição arquivística. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 272-282). Marília: ISKO-Brasil: FUNDEPE.
- Marinho, A. C. M., & Santos, R. N. M. (2017). A Música como Recurso Informacional: Novas Perspectivas de Estudo na Organização da Informação. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 317-324). Recife, PE: Ed. UFPE.
- Medeiros, G. M., Vital, L. P., Linden, L. L., & Brascher, M. (2015). A representação de assunto no escopo da Arquivologia: uma análise de artigos científicos nacionais e internacionais. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 498-506). Marília: ISKO-Brasil: FUNDEPE.
- Ninin, D. M., & Simiionato, A. C. (2017). Sustentabilidade de Acervos de Digitais: Reflexões sobre Linking Open Data para Coleções de Patrimônios Culturais. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 154-161). Recife, PE: Ed. UFPE.
- Pando, D. A. (2018). *Epistemologia da Organização da Informação: uma análise de sua cientificidade no contexto brasileiro* (Tese de doutorado em Ciência da

- Informação, Faculdade de Filosofia e Ciência, Universidade Estadual Paulista, Marília, Brasil). Recuperado de <http://hdl.handle.net/11449/153395>
- Pinho, F. A., & Guimarães, J. A. C. (2017). *Memória, tecnologia e cultura na organização do conhecimento*. Recife, PE: Ed. UFPE.
- Pinho, F. A., Nascimento, F. A., & Marinho, A. C. M. (2015). Música, literatura e audiovisual: as contribuições da organização do conhecimento (OC) nas relações de interseccionalidade das obras de Dorival Caymmi e Jorge Amado. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 621-629). Marília: ISKO-Brasil: FUNDEPE.
- Sá, I. P., & Campos, M. L. A. (2017). Aportes teóricos e metodológicos da Organização e Representação do Conhecimento na Ciência Aberta: subsídios para a implantação de sistemas de gestão de informação de pesquisa (CRIS) nos laboratórios da Fundação Oswaldo Cruz. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 170-179). Recife, PE: Ed. UFPE.
- Sales, R. (2015). A relação entre organização do conhecimento e ciência da informação na comunidade científica brasileira: uma investigação no âmbito da ISKO-Brasil. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 73-84). Marília: ISKO-Brasil: FUNDEPE.
- Sales, R. (2017). Ciência da Informação e Organização do Conhecimento no Brasil: Uma Interface entre o GT2 da ANCIB e a ISKO-Brasil. In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 120-127). Recife, PE: Ed. UFPE.
- Santos, R. F., & Corrêa, R. F. (2017). Organização da Informação em Repositórios Digitais: Uma Abordagem sobre a Política de Indexação da Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). In F. A. Pinho & J. A. C. Guimarães (Orgs.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 249-261). Recife, PE: Ed. UFPE.
- Smiraglia, R. P. (2018). ISKO 15's Bookshelf: Dispersion in a Digital Age - An Editorial. *Knowledge Organization*, 45(5), 343-357. doi:10.5771/0943-7444-2018-5-343
- Smit, J. W. (2015). A busca pela eficiência na representação da informação e do conhecimento – desdobramentos posteriores no pensamento de Gardin. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 726-735). Marília: ISKO-Brasil: FUNDEPE.
- Tartarotti, R. C. D., Dal'Evedove, P. R., & Fujita, M. S. L. (2015). Atuação bibliotecária na análise de assunto de teses nas dimensões teóricas da catalogação de assunto e indexação. In J. A. C. Guimarães & V. Dodebei (Orgs.), *Organização do Conhecimento e Diversidade Cultural* (pp. 329-337). Marília: ISKO-Brasil: FUNDEPE.
- Taylor, A. G., & Joudrey, D. N. (2009). *The Organization of Information* (3rd ed.). Westport, CT: Libraries Unlimited.

CAPÍTULO 3
CATALOGAÇÃO E METADADOS:
REFLEXÕES SOBRE CONCEPÇÕES, PERSPECTIVAS
E TENDÊNCIAS
CATALOGING AND METADATA: REFLECTIONS
ON CONCEPTS, PERSPECTIVES AND TRENDS

Cíntia de Azevedo Lourenço
Universidade Federal de Minas Gerais (UFMG)
cal@eci.ufmg.br
ORCID: 0000-0002-2172-7300

Zaira Regina Zafalon
Universidade Federal de São Carlos (UFSCar)
zaira@ufscar.br
ORCID: 0000-0002-4467-2138

António Tavares Lopes
Universidade de Coimbra (UC)
Faculdade de Letras/CEIS20
atlopes@fl.uc.pt
ORCID: 0000-0002-8876-5387

Resumo: A catalogação passou por um grande período de estagnação, sendo resgatada no final do século XX. Nesse novo universo híbrido, onde sistemas de informação analógicos convivem com sistemas de informação digitais, novas perspectivas conceituais, as normas de catalogação e os objetivos comunicacionais do catálogo cada vez mais centrados na satisfação do usuário são o foco deste trabalho. Outro enfoque abordado é a relação da catalogação com os metadados e suas aplicações em sistemas de representação da informação. Nas considerações finais, são levantados vários pontos e conceitos que necessitam uma maior reflexão neste novo contexto que surge no fazer da catalogação.

Palavras-Chave: Catalogação. Metadados. Recurso informacional.

Abstract: Cataloging went through a long period of stagnation, being rescued at the end of the 20th century. In this new hybrid universe, where analogical information systems coexist with digital information systems, new conceptual perspectives, cataloging norms and communicational objectives of the catalog, increasingly centered around user satisfaction, are the focus of this work. Another approach addressed is the relationship of cataloging with metadata and its applications in information representation systems. In the final considerations, several points and concepts are raised that need further reflection on the new contexts of cataloging.

Keywords: Cataloging. Metadata. Informational resource.

1. Introdução

Os ambientes informacionais digitais com os quais a Ciência da Informação se compromete na atualidade, quer seja do ponto de vista dos estudos, quer seja da aplicação e das proposituras, centram-se nos usuários da informação e se delinham a partir do entendimento de que suas necessidades informacionais são conhecidas.

A partir dessa concepção, o ponto de partida é o questionamento central de como os aspectos teórico-conceituais da organização e representação da informação se consolidam nas unidades de informação de modo a garantir a efetividade de suas ações sob a perspectiva de comunicação centrada no usuário.

Comprometidos com esse viés propõe-se a discussão de aspectos teórico-conceituais da organização e representação da informação e a sua correlação com os aspectos práticos que envolvem os ambientes informacionais digitais. O delineamento conceitual inerente à representação da informação será feito a partir da discussão do objeto da Ciência da Informação e da catalogação como confluência de aspectos descritivos, temáticos, contextuais e de acesso à informação de modo que a interconexão entre eles esteja assegurada.

Nesta oportunidade, faz-se primordial esclarecer aspectos que são considerados inerentes à expressão “representação descritiva” e à perspectiva de dissociação de saberes voltados ao ensino e aos conteúdos disciplinares canalizados aos instrumentos e aos métodos específicos do tratamento descritivo e temático da informação, mas que não refletem a prática biblioteconômica, com produtos consubstanciados em catálogos das mais diferentes unidades de informação, bibliotecas, arquivos e museus, mas também em outros sistemas de recuperação da informação como repositórios e bases de dados.

A compreensão que se tem de “representação descritiva” como tautologia centra-se, de modo simplista e posto como argumento inicial, que, para representar algo, é necessário descrevê-lo; que toda descrição é, também, representação; que o processo de representação é visto como colocar algo que é descrito pelo fato de não estar ali, disponível, naquele momento (as representações, nesse caso servem como meio de se chegar ao que se quer). É importante considerar que outra prática biblioteconômica que faz uso da descrição é a referência bibliográfica que prescinde de aspectos cruciais e presentes na catalogação: os aspectos temáticos, de

localização e de contexto, tratados logo mais. Tanto a referenciação quanto a catalogação são importantes e necessárias na comunicação científica, porém, cada uma tem, em seu cerne, objetivos e metas diferentes e específicas.

A proposta de discussão feita aqui quanto à “representação descritiva” considera os argumentos apresentados por Sfez (2000), em *Crítica da Comunicação*, com a proposição do tautismo na comunicação, um misto de tautologia e de autismo, uma comunicação autista; noutras palavras, “[...] o tautismo utiliza a tautologia como única verificação: se repito, provo.” (Sfez, 2000). Há tantos anos se repete “representação descritiva” tornando-se necessário entendê-la como prova, como se o conceito em si se esvaziasse em uma surdez!

Esse fenômeno tautístico pode ser descrito da seguinte maneira: não se toma mais o real como representado. O real tampouco é o que se inventou com esse nome ao exprimi-lo. No tautismo, toma-se a realidade representada por uma realidade expressa. Toma-se o representado pelo representante. [...] como se a cadeia dos intermediários, que extraíram a informação e produziram seu enquadramento, sua apresentação ao receptor, fosse bruscamente suprimida. (Sfez, 2000, pp. 77-78).

Daí que, ao adotar-se “representação descritiva” omite-se pela repetição, ou seja, faz a representação tornar-se surda, faz com que o aspecto de apresentação, inerente ao processo comunicativo, seja suprimido. Frente a essa realidade, optou-se por adotar, no presente trabalho, o termo catalogação que, no sentido amplo, é o processo que culmina objetivamente pela apresentação de produtos voltados ao estabelecimento do processo comunicativo entre pessoas, máquinas e unidades de informação. Um conceito que serve como mote para argumentar que a catalogação conjuga aspectos descritivos,

temáticos, contextuais e de acesso à informação de modo que a interconexão entre eles esteja assegurada.

Se, por um lado, ao longo dos anos, catalogação foi o termo usado para os estudos e as atividades do aspecto descritivo dos documentos, nesta discussão que se propõe fazer, compreende-se como catalogação o termo que expressa o conceito de representação da informação e, sob esse prisma, a catalogação descritiva conjuga-se à catalogação de assuntos, aos dados de localização e ao contexto em que se insere.

Faz-se, neste universo, de suma importância o enfoque no papel que a catalogação exerce na construção de catálogos eletrônicos, bibliotecas digitais, repositórios digitais e bases de dados.

A catalogação, portanto, associa elementos de metadados e seus padrões, tanto estruturais quanto de valor, na potencialização da capacidade de interoperabilidade por eles proporcionadas, e a consequente geração e integração das unidades de informação em cadeias de valor que ultrapassam o âmbito mais restrito dos produtos bibliográficos e dos sistemas biblioteconômicos, visto que se concretizam no contexto dos recursos distribuídos e em redes de organização do conhecimento.

2. Aspectos teórico-conceituais da organização e representação da informação

Inicialmente, para falar de organização e representação da informação, é primordial que se entenda o conceito de documento: “[...] qualquer objeto informacional, independente de seu suporte, que forneça a um usuário de informação um dado ou uma informação relevante de forma diferenciada, de acordo com suas características físicas ou intelectuais específicas.” (Feitosa, 2006, pp. 17-18). Em 1934, Otlet (2018, p. 580) já considerava documento como “[...]”

representação da realidade em forma literária (o bíblion, o escrito, o texto), gráfica ou plástica (o ícone, a imagem)”.

Vinculam-se no documento, portanto, características físicas e intelectuais, ou, como Smiraglia (2002) apresenta, conteúdos ideacionais realizados em uma expressão alfanumérica, musical, sonora, imagética, etc. Dentre as características físicas de um documento, destacam-se: material, natureza, tamanho, peso, forma de produção, suporte, etc.; como características intelectuais: conceitos relacionados ao objetivo, conteúdo, assunto, tipo de autor, fonte, forma de difusão, originalidade, etc.

É importante frisar que nos estudos de organização e representação da informação, são as características intelectuais que irão permitir que se identifique os tópicos de interesse contidos em um documento, seu público-alvo específico e seu valor informacional para os usuários de informação; as características do suporte documental dão condições de localizar o documento em que tais características intelectuais estão contidas, de modo que seja possível ter acesso ao conhecimento nele registrado.

Quanto aos conceitos relacionados à representação da informação é necessário ter, preliminarmente, uma visão mais abrangente do conceito de representação, pois é um termo repleto de polissemia, por possuir significados distintos em diversas áreas do conhecimento. Nesse sentido, Kobashi (1996) esclarece que entre todos os significados da palavra representação, os mais relevantes para a área da Ciência da Informação são oriundos das áreas da linguística e semiótica, da ciência da comunicação e da documentação. É no escopo dessas três áreas que se dá a abordagem de representação na Ciência da Informação, que envolve características físicas e circunstanciais dos documentos e características temáticas. Nesse sentido, Lima (1998, p. 56) nos apresenta uma definição do fazer da representação da informação:

[...] representar o conhecimento acumulado por um domínio é apenas uma parte do processo de tratamento documentário para que este conhecimento transformado em informação possa ser acessado pelo usuário de um Sistema de Informação e efetivar assim a comunicação documentária.

A representação na Ciência da Informação acontece, portanto, em dois momentos claramente definidos: o da produção do registro do conhecimento, e o do trabalho intelectual e material de organização da informação.

A representação, quando da produção do registro do conhecimento, ocorre com a intenção autoral, ainda na mente do criador, que considera suas impressões e conceitos ideacionais e avalia qual a melhor forma para apresentá-los. Isso se reflete na organização textual, na construção de sumários e de índices alfabéticos materializados no documento.

A representação, no segundo momento, consagra-se por conta da estruturação e organização de sistemas de informação que favorecerão os processos de busca e recuperação de maneira mais eficiente. Neste processo intelectual e material são requeridos métodos, técnicas e instrumentos para que o produto desse processo possa ser apresentado aos usuários, quer seja humano ou maquínicos. Tais métodos, técnicas e instrumentos de representação podem ser classificados segundo sua finalidade: organização da informação e organização do conhecimento. Brascher e Café (2008) recorrem à abordagem de informação e conhecimento dado por Fogl (1979) para indicarem que a organização da informação recorre a aspectos morfológicos do documento e a organização do conhecimento aos aspectos conceituais do documento.

O registro, nos sistemas de informação, é, portanto, o produto gerado pela aliança de métodos, técnicas e instrumentos específicos da organização da informação e da organização do conhecimento

que permitem que os documentos possam ser buscados e identificados: pela definição dos dados de localização (quer seja pelo número de chamada, quer seja por meio de links) que faz com que os documentos possam ser adquiridos, obtidos ou acessados; e pelo contexto, que dá ao usuário, meios para encontrar similitudes e distinções entre um documento e outro, seja pelas várias expressões de uma mesma obra, seja pelas associações entre as mesmas pessoas ou instituições, ou, ainda, pelo assunto ou interpretações.

A partir da afirmação de Kobashi (1996) sobre a contribuição dos estudos da linguística para a Ciência da Informação, retoma-se a análise feita por Zafalon (2014, p. 54) sobre os aspectos sintáticos e semânticos do registro bibliográfico, produto da catalogação:

[...] a) sintaxe [...] estuda a forma, o arranjo e a disposição de cada elemento descrito na representação do recurso informacional e, com base em Saussure (2010), compreende-se que a sintaxe estabelece as relações formais e síncronas entre *significado*, *significante*, *obra* e *manifestação*;

b) em registros bibliográficos a semântica, estabelecida a partir da teoria dos níveis do estudo estruturalista dos registros bibliográficos, com base em Hjelmslev (1991), ocorre em três momentos: a. subsemântico – estudo da teoria das unidades semânticas, ou dos elementos semanticamente relacionados presentes em um registro bibliográfico; significado estabelecido entre cada um dos elementos de um registro e os valores de representação que assume; b. semântico – descreve as relações estruturais estabelecidas entre o registro bibliográfico e o recurso informacional; c. supersemântico – relação semântica entre os vários registros bibliográficos representados em um catálogo, por meio de identidade de significados (semelhanças) e de oposição de significados (diferenças).

Está posto, afinal, que para a Ciência da Informação, a representação acontece em duas frentes de tratamento da informação:

- O tratamento descritivo da informação, que envolve o trabalho de representação das características morfológicas do documento a partir de um item documental;
- O tratamento temático da informação, que envolve a representação do conteúdo ideacional de um documento.

Outro ponto importante a destacar é a percepção de Klim (1981, pp. 55) que identifica três funções da representação que vão delinear seu objetivo principal, a saber:

- A função de informação, que consiste primeiramente em prover informação sobre a existência de um documento específico e, em segundo lugar, em comunicar diversos tipos de informação sobre o documento;
- A função de identificação, que consiste em assegurar que a representação e o documento a que ela corresponde sejam idênticos; daí a formulação prévia dos elementos de representação, com número suficiente de elementos que permitam sua identificação;
- A função de recuperação, que consiste no fato de que a representação é um meio de se recuperar um documento, qualquer que seja o sistema de recuperação da informação.

Para a determinação do conjunto perfeito de elementos da representação em cada caso específico [...] precisamos considerar não somente as funções da representação [...], mas também toda a gama de fatores que afetam esta descrição, tais como tipo de organização da informação, tipo e tamanho do SRI, fonte da descrição, demandas dos usuários, etc. (Klim, 1981, pp. 56)

Além desses conceitos básicos apresentados até o momento, também é de suma importância o entendimento dos processos que envolvem a catalogação.

Podem ser identificadas como processos em catalogação, tanto em ambientes tradicionais quanto digitais, as seguintes atividades de organização e tratamento da informação:

- Leitura técnica, que permite a identificação tanto dos elementos intrínsecos (descritivos) quanto extrínsecos (temáticos) do documento;
- Identificação dos pontos de acesso ao documento, definidos no momento da leitura técnica, que permitirão a construção de índices a serem utilizados como recurso nas atividades de recuperação da informação;
- Descrição bibliográfica propriamente dita, transcrição dos elementos representativos do documento em um sistema de recuperação da informação.

É a partir desses processos iniciais que a catalogação pode gerar diversos produtos que servirão de aporte para os sistemas de recuperação da informação, dentre os quais podem ser citados: catálogos analógicos, eletrônicos, online, de acesso público; bibliotecas digitais; repositórios de informação digital, temáticos e institucionais; portais de informação; bases de dados referenciais, tanto na forma de *index* quanto de *abstract*, e de texto completo, além do controle bibliográfico.

Tendo em vista a finalidade dos processos e produtos da organização e representação da informação o próximo tópico trata do processo comunicativo com os usuários.

3. Perspectiva de comunicação centrada no usuário

Nos primórdios dos estudos da organização da informação o foco das instituições estava no armazenamento, na organização dos documentos e na preservação do conhecimento humano. Poucas pessoas tinham acesso às letras e, por conseguinte, o acesso aos acervos das bibliotecas era restrito. Nesse período, portanto, pouca atenção era dada ao usuário; concentrava-se nas atividades de organização e aos cuidados de preservação.

Com o passar dos tempos, o conhecimento humano passou de uma produção de registros demorada e onerosa para uma produção em larga escala, resultante do desenvolvimento da pesquisa científica e da invenção da imprensa. Nesse período foram iniciados os primeiros estudos focados, principalmente, na organização dos acervos de uma forma mais dinâmica e eficiente, e na melhoria dos recursos para desenvolvimento de sistemas de recuperação da informação em formatos analógicos.

Nesse contexto de evolução cultural e de produção de conhecimento, um evento marca os estudos de organização da informação: a Revolução Francesa. Mey (2009) afirma que nesse período há o confisco de grandes bibliotecas particulares pertencentes a nobres e ao clero e é dado o acesso público às bibliotecas. Apesar do intento, dar acesso ao conhecimento não bastava, pois a maioria da população ainda era analfabeta.

Foi finalmente no final do século XIX que teóricos plantaram as primeiras sementes dos estudos da organização da informação: Cutter, Jewett, Panizzi, Bliss, Dewey, Otlet, La Fontaine, entre outros. Apesar de os usuários já terem sido indicados nos objetivos bibliográficos por alguns destes teóricos, o foco maior ainda estava na melhoria das técnicas de organização da informação e do conhecimento de modo a universalizar o acesso à informação por meio de atividades

de recuperação pautada em técnicas e métodos mais sofisticados de tratamento da informação.

Nesse período observou-se o grande *boom* da Biblioteconomia e o início dos estudos de Ciência da Informação: desenvolvimento de códigos e normas de catalogação e descrição bibliográfica, sistemas de classificação de assuntos para a organização dos documentos na estante de forma sistemática, e linguagens de controle de vocabulário, como listas de cabeçalhos de assuntos.

No entanto, o primeiro grande marco para a catalogação se deu com a realização da Conferência de Paris, em 1961, evento em que foram estabelecidos princípios a serem considerados na catalogação e na definição de códigos, normas e formatos de intercâmbio para a catalogação cooperativa por computador. Nesta década, começam a ser discutidos o Código de Catalogação Anglo-Americano (AACR2), o Padrão Internacional para Descrição Bibliográfica (ISBD) e o formato MARC. Todos esses instrumentos foram sendo revistos e atualizados ao longo do século XX, para se adequarem aos novos suportes da informação que surgiam, como as mídias de som e imagem, ao avanço dos recursos tecnológicos que modernizaram os catálogos de bibliotecas, e às novas necessidades de informação dos usuários, seja quanto à forma de se relacionar com as bibliotecas e outras unidades de informação, seja nos comportamentos de busca nos sistemas de recuperação da informação.

Porém, é com o advento da informatização das bibliotecas que uma nova realidade começa a se delinear na área de organização da informação: a necessidade de adaptação dos processos de organização e tratamento da informação ao ambiente tecnológico que se expandia apressadamente.

Nesse quadro de mudanças e inovações tem início, no Seminário sobre Registros Bibliográficos, realizado em Estocolmo, em 1990, estudos para a compreensão da função dos registros bibliográficos nesse novo ambiente, o digital, e para o desenvolvimento de novos

modelos de organização da informação baseados nas tecnologias da informação e da comunicação. Esses questionamentos são marcantes para a catalogação, pois é quando se iniciam os estudos de modelagem conceitual do universo bibliográfico.

O primeiro modelo conceitual apresentado é o *Functional Requirements for Bibliographic Records* (FRBR), em 1998. A partir de reuniões, iniciadas em 2013, do Grupo de Revisão dos FRBR com o grupo do Comitê de Documentação do Conselho Internacional de Museus (ICOM), os FRBR, propostos com modelagem entidade-relacionamento, são reformulados para o modelo orientado a objetos, adotado pelo ICOM no Modelo de Referência Conceitual (CIDOC CRM), o que resulta na publicação, em 2009, do FRBRoo (FRBR orientado a objetos). Neste mesmo ano é publicado o *Functional Requirements for Authority Data* (FRAD), e, em 2011, o *Functional Requirements for Subject Authority Data* (FRSAD), responsáveis, respectivamente, pela descrição de dados de autoridade e dados de assunto. A aproximação do FRAD e do FRSAD ao FRBRoo, dá origem à segunda versão do FRBRoo.

Desde o início, em 2010, grupos de revisão trabalharam para a consolidação dos modelos conceituais propostos até então para as bibliotecas. Em 2016, o novo modelo, inicialmente chamado de Modelo de Referência Bibliotecária FRBR, passa a ser identificado como Modelo de Referência Bibliotecária da IFLA (IFLA LRM). Sua aprovação e publicação aconteceram em 2017. Esses modelos foram a base para o desenvolvimento de uma nova norma de catalogação, a RDA, mais adequada para o universo de informação digital.

É também no final do século XIX que a preocupação com o atendimento aos usuários pelos sistemas de recuperação de informação e pelos serviços e produtos das bibliotecas começa a florescer, tendo nas cinco leis de Ranganathan (2009) sua expressão mais conhecida até os dias de hoje. O teórico propõe um olhar atento aos recursos informacionais e seu uso; ao fato de que os usuários fazem bus-

cas específicas; cada documento deve ser visto com base no perfil dos usuários; os produtos e serviços oferecidos devem valorizar o tempo disponível dos usuários; as unidades de informação devem-se atualizar quanto aos seus produtos, serviços, acervos, capacitação de pessoal e tecnologias.

Ao longo do século XX buscou-se meios para identificar e conhecer os usuários, atender e prever as necessidades informacionais dos usuários e adequar a organização dos objetos informacionais, tanto em bibliotecas tradicionais quanto em bibliotecas digitais, de maneira a melhor atender seus usuários, postos agora como foco das unidades de informação.

Nos estudos de organização e representação da informação o usuário comparece, especificamente, nos objetivos bibliográficos definidos por Cutter, em 1876, e com Svenonius (2000). Mas é a partir dos estudos iniciados no Seminário de Estocolmo (apresentado anteriormente) que surge um enfoque mais moderno e atualizado para a organização da informação: centrado no usuário. Os modelos conceituais da família FR e, conseqüentemente, a norma RDA, trazem, já no início do século XXI, um novo olhar para uma organização focada nas necessidades dos usuários de informação.

Nos catálogos manuais, o foco no usuário era estruturado com base na criação de pontos de acesso que facilitariam aos usuários conhecer o acervo da biblioteca dentro de suas necessidades individuais de informação. O catálogo era construído de forma a serem elaborados índices alfabéticos para que os usuários pudessem localizar um documento por seus autores, pelo título do documento ou da coleção, ou por assuntos.

Com o advento das novas tecnologias, que contribuiu para a informatização dos catálogos de bibliotecas, para o estabelecimento de bibliotecas e repositórios digitais, e bases de dados referenciais e com texto completo em ambiente eletrônico e acessíveis pela Internet, percebeu-se a necessidade de ampliar esse escopo de recuperação

da informação, de maneira que os usuários pudessem localizar um recurso informacional por outros critérios que não somente aqueles dos convencionais (autor, título e assunto), mas também por outros elementos da descrição bibliográfica, como editores e produtores, tipo de suporte de informação, idioma, ano de publicação. Também começou a ser possível ao usuário navegar pelo catálogo, por meio de critérios e associações de metadados, e fora dele, localizando obras relacionadas àquela área de seu interesse, ampliando o seu universo de recuperação da informação, sem, no entanto, ampliar o universo de revocação em suas buscas.

De modo simplificado e direto apresenta-se o quadro abaixo, com a sistematização das tarefas dos usuários segundo os modelos conceituais da IFLA:

Quadro 1: Tarefas dos usuários nos modelos conceituais da IFLA

FRBR ¹	FRAD ²	FRSAD ³	LRM ⁴
encontrar	encontrar	Encontrar	encontrar
identificar	identificar	Identificar	identificar
selecionar	contextualizar	Selecionar	selecionar
obter	justificar	Explorar	obter
			explorar

Fonte: Os autores (tradução nossa).¹²³⁴

Com o modelo conceitual IFLA LRM, as tarefas definidas para o FRBR, FRAD e FRSAD são avaliadas, discutidas e recompiladas, de modo que a tarefa *encontrar* tem o objetivo de reunir informações sobre um ou mais recursos de interesse do usuário pesquisando em qualquer critério relevante; a tarefa *identificar* permite ao usuário compreender claramente a natureza dos recursos encontrados e

¹ International Federation of Library Associations and Institutions (1998).

² International Federation of Library Associations and Institutions (2009).

³ International Federation of Library Associations and Institutions (2011).

⁴ International Federation of Library Associations and Institutions (2017).

distinguir entre recursos semelhantes; a tarefa *selecionar* faz com que o usuário determine a adequação dos recursos encontrados e habilite-se para aceitar ou rejeitar recursos específicos; a tarefa *obter* é definida para que o usuário acesse o conteúdo do recurso; e a tarefa *explorar* tem o propósito de dar ao usuário a possibilidade de descobrir recursos usando as relações entre eles e, assim, colocar os recursos em um contexto (*International Federation of Library Associations and Institutions [IFLA]*, 2017).

De acordo com Mering (2014, p. 2), “[...] todos os usuários compartilham uma necessidade em comum: catálogos que executem de forma eficiente, eficaz e com êxito suas tarefas de busca de informação.” Nesse sentido, estas tarefas definidas para o usuário no uso dos catálogos tornam-se peça fundamental na catalogação.

No universo atual de organização e representação da informação, as reavaliações e reconsiderações dos métodos, processos, instrumentos e produtos da catalogação têm ocorrido com vistas à melhor adequação às necessidades do usuário, à sua linguagem, além de estar mais adaptável ao surgimento de novos recursos informacionais e sujeitos às adequações às tecnologias da informação emergentes, discussões que serão feitas no tópico a seguir.

4. Catalogação e metadados: concepções, confluências e perspectivas

A catalogação assume para si a preocupação da análise criteriosa do recurso informacional – na tentativa de uma avaliação pela perspectiva do usuário, que, por vezes, não se tem contato proximal, que permita uma aproximação e que favoreça o ‘dar ouvidos’ –, o conhecimento de instrumentos, de sua apropriação e uso, que permitam a representação dos dados inerentes, explícitos ou implí-

ritos, a eles, tendo em vista os processos de busca e recuperação por metadados, para que seja acessado.

Concebe-se a catalogação como

[...] atividade fundamental para o estabelecimento do processo comunicativo nos catálogos de instituições de patrimônio cultural tendo em vista a ponte a ser consolidada entre os recursos de que tais unidades dispõem e os usuários ou grupos ou outras instituições a serem atendidas em suas necessidades informacionais. [...] A representação documental requer, portanto, conhecimento do público a que se destina, das especificidades do gênero e do formato dos recursos informacionais, e dos padrões, métodos e instrumentos a serem utilizados. (Zafalon, 2017, pp. 131-132).

Com o propósito de tecer, por meio de um discurso racional-reflexivo, uma avaliação dos conceitos da catalogação e dos metadados, pretende-se identificar, ou não, confluências, e traçar, ou se arriscar, a arquitetar perspectivas.

Como processo de elaboração de metadados, a catalogação se faz presente nos mais variados discursos institucionais, desde os de arquivos, bibliotecas, museus e galerias de arte, até naqueles que se apropriam deste conceito para designar um método de descrição, de categorização, de elaboração de listagens em outras instituições, como clínicas médicas, lojas de departamentos, agências de turismo, empresas de comunicação, rádio e televisão, etc.

Enquanto nos estudos que envolvem as primeiras instituições a catalogação recorre a teorias, métodos, instrumentos e produtos delineados científica e criticamente com vistas ao usuário de tais registros, nas seguintes o processo é naturalizado, elaborado, experimentado e reavaliado pelos próprios usuários.

Perspectivas diferentes, motivações heterogêneas, fim comum: servir para pesquisa e indagação em sistemas de informação em

busca de respostas, quer sejam aquelas de recursos informacionais, quer sejam as relativas a outros dados, e/ou fluxos. A finalidade, o “para quê” é, portanto, comum às instituições de patrimônio cultural e àquelas de uso cotidiano.

O que, então, precisa ser olhado de um modo diferente, visto que a finalidade de busca, recuperação e acesso de documentos, e até a outros dados, são inerentes ao processo de elaboração de metadados por instituições de patrimônio cultural e além delas (aquelas gerais)?

Ao considerar as questões “onde” e “quando”, a contestação se dá no nível político-institucional, visto que recorrer ao processo de catalogação e de elaboração de metadados obedece a diretrizes lógico-administrativas voltadas a dirimir dúvidas e orientar tomadas de decisão em situações correlatas ou congêneres. Isto significa que definir o escopo da catalogação, em cada “lugar em que se faça presente”, abarca a definição de seus objetivos e finalidades, bem como de recursos informacionais a serem descritos, visto que farão parte do rol de documentos/objetos do acervo; a escolha do nível de detalhamento da descrição, da estrutura a ser utilizada como “campos” da descrição, das tecnologias adotadas nos sistemas de informação; o controle de termos que melhor explicitem os conceitos identificados no documento/objeto; a delimitação de responsabilidades e papéis dos atores implicados no processo.

A relação entre catalogação e metadados, ou “o quê há entre”, “qual o limite entre um e outro”, se dá em uma perspectiva direta, apesar de sistêmica; são diferentes, mas complementares; com vida própria, e, ao mesmo tempo, interdependentes. A catalogação, como processo mental, consolida-se, pelos metadados, enquanto elementos-meio, em um produto, o *representamen*. O registro é, portanto, simbólico, é resultado representativo de algo que não está ali, mas que a ele remete. Daí a compreensão que a catalogação

conjugue aspectos intelectuais e técnicos. Intelectuais por conta dos processos; técnico pelos instrumentos.

Chega-se ao “como se faz”, ao ponto de inflexão que traz à tona o desacordo com o senso comum da catalogação popular, para se chegar à perspectiva especializada, científica e crítica, discutida nos estudos de Organização e Representação da Informação e do Conhecimento.

A proposição da discussão se faz pela perspectiva científica interdisciplinar e pelo valor circunstancial. Estudos que se propõem a discutir a interdisciplinaridade na e da Ciência da Informação já estão marcados em pesquisas⁵ a isso dedicadas. Denotam-se relações da Ciência da Informação com as dimensões humana, social e tecnológica, configuradas nos mais variados domínios. A questão que importa neste estudo centra-se no fato de que a relação interdisciplinar é dada em um determinado contexto, com o domínio que for conveniente para sua delimitação e delineamento. Por conta disso, o objeto de estudo da catalogação se configurou, majoritariamente, nos recursos informacionais, ou seja, nos documentos, nos objetos que são fruto da realização mental e que se consubstanciam ideal e fisicamente⁶. Encaixam-se, nessa ótica, objetos que fizeram parte, ao longo dos anos, dos estoques de arquivos, bibliotecas e museus, ou seja, aqueles que expressam o patrimônio cultural material e institucionalizado. Considera-se, porém, que é possível também conjecturar aqueles “[...] objetos com potencial informativo, [...] que, por conta do contexto e, por caracterizarem-se como evidência,

⁵ Sugere-se a consulta a pesquisas de Gomes (2001), Silva (2005), Riecken (2006), Pinto (2007), Zins et al. (2007), Pinheiro e Loureiro (2007), Silva, Lima e Araújo (2009), Bicalho e Oliveira (2011), Santos e Rodrigues (2013), Santos, Chacon e Veras (2015), Moraes e Carelli (2016), Nisenbaum e Pinheiro (2016), Jovanovich et al. (2017), Santos Neto et al. (2017), Lança, Amaral e Gracioso (2018).

⁶ Incluem-se aqui os documentos analógicos, eletrônicos e digitais. Opta-se por não utilizar o termo virtual visto que neste, também, é imprescindível o registro, quer seja em recursos eletrônicos ou digitais.

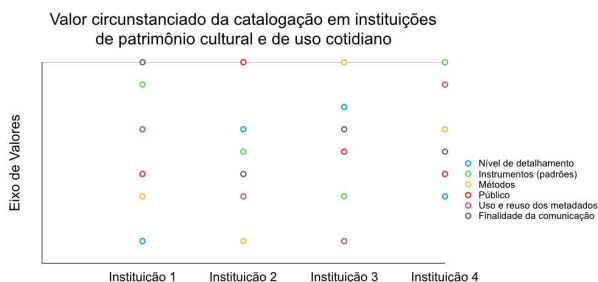
tem suas potencialidades projetadas nas ações da representação documental.” (Zafalon, 2017, p. 126).

Esta perspectiva de recursos informacionais como evidência ainda tem muito a ser discutida e considera que o olhar da catalogação igualmente deva ser convergido ao patrimônio cultural imaterial, haja vista os estudos de documentação e musealização e os processos de investigação participativa.

Quanto ao valor circunstancial dado à catalogação, considera-se que o nível de detalhamento da descrição, os instrumentos e os métodos a serem adotados, a delimitação do público a que os registros se destinam, o uso que se faz dos metadados, e o estabelecimento de como se dará o processo comunicativo nos sistemas de busca e recuperação, sejam pertinentes, mas o valor que se atribui a cada um deles, visto o peso que cada um tem diante dos objetivos institucionais, é variável.

A ideia que se quer dar para o conceito do valor circunstancial da catalogação pode ser avaliado no Gráfico 1:

Gráfico 1: Ilustração da ideia do valor circunstancial da catalogação



Fonte: Os autores.

A intenção, ao recorrer ao gráfico, é demonstrar que o valor que se dá à catalogação, de certo modo, independe do tipo de instituição (de patrimônio cultural ou do cotidiano) que define os metadados para os documentos, mas tem uma dependência direta da ponderação que a instituição faz para os quesitos inerentes à catalogação. Daí os ganhos e perdas que se tem: mais por conta da amplitude que cada quesito tem, o que define a qualidade do produto final (o registro), do que pelo tipo de instituição que faz o processo de catalogação, pela elaboração dos metadados.

Uma instituição que valorize mais os métodos de catalogação do que o uso e reúso dos metadados pode não ter tanto sucesso no compartilhamento dos metadados criados, apesar de preocupar-se, sobremaneira, com o nível de detalhamento da catalogação (caso da instituição 3, no gráfico). Por outro lado, uma instituição que evidencia o seu cuidado com os padrões de metadados e com o uso e reúso dos metadados, mesmo que não tenha tanta preocupação com o nível de detalhamento, poderá por métodos delimitados, ainda assim ter bons resultados nos processos de interoperabilidade de registros com outras instituições (caso da instituição 4, no gráfico).

Desse modo, assume-se que a catalogação, sob o prisma centrado no usuário, e que conjuga conceitos teóricos e particularidades da prática, por uma perspectiva crítica dos estudos de Organização e Representação da Informação, merece a (re)avaliação de seus aportes teóricos, quer seja revisitando autores da Ciência da Informação e da Biblioteconomia, quer seja da ponderação da relação interdisciplinar com a Computação, a Linguística, a História, e, por que não dizer, dos próprios argumentos ontológicos sobre o seu objeto de estudo e sua configuração como disciplina, como área ou como campo.

O tópico seguinte discute questões inerentes aos metadados em ambientes informacionais digitais.

5. Aplicações de metadados em ambientes informacionais digitais

Uma das mais significativas facetas no desenvolvimento dos metadados bibliográficos – aqui considerando a modelagem e o conjunto de suas aplicações – manifesta-se na diversidade de tipos de agentes cuja atenção procuram captar.

O principal produto da catalogação continua a ser o catálogo e, como principal instrumento de comunicação e de representação, o registro bibliográfico. Em um mundo informacional cada vez mais complexo, a exigir padrões de conectividade intensivos e geridos por sistemas computacionais auto-operados, ou de elevado grau de autonomia operacional, novos desafios são colocados aos sistemas de informação bibliográfica, convocando-os para fora da sua área de conforto e estimulando a sua participação em redes complexas de processamento e troca de informação.

A resposta aos desafios da interoperabilidade representa uma oportunidade imperdível para acrescentar valor aos produtos da catalogação, sobretudo por potencializar a sua capacidade para se inserir na cadeia de valor de outros sistemas, por conta do fornecimento de informação estruturada produzida com padrões de qualidade relativamente elevados, normalmente sujeitos a orientações estabelecidas e normalizadas por uma comunidade de prática de cobertura à escala global.

Para que o potencial de valorização dessa resposta possa ser percebido, a primeira proposição é começar por fazer uma revisão do conceito de interoperabilidade. Em seguida, serão enunciadas algumas das formas como ela tem se concretizar, tanto do ponto de vista conceitual como operacional, bem como algumas das dificuldades que tais formas encontram na produção, reprodução, transformação e utilização de metadados.

Esta primeira exploração servirá como guia para uma identificação exemplar não exaustiva de projetos que proponham a integração aplicacional das revisões conceituais da representação da informação bibliográfica apresentadas anteriormente neste capítulo, ao mesmo tempo em que projetam o seu valor para além do campo restrito da catalogação e das aplicações biblioteconômicas.

5.1. Metadados bibliográficos e os desafios da interoperabilidade

Uma definição comum de interoperabilidade não pode deixar de salientar a capacidade que dois sistemas distintos têm para trocar informação que conseguem integrar no seu processamento corrente, ou, mais basicamente, a capacidade que um sistema tem para operar com outro sistema, no todo ou em parte. Tal definição pressupõe o reconhecimento de um efeito positivo de ganho como resultado da transação.

Diferentes níveis de interoperabilidade, que pode ser vista sob a perspectiva de camadas, contribuem e concorrem para a percepção dessa vantagem. A um nível mais elementar,

- a *interoperabilidade de sistema* ou a *camada de transporte e troca* torna possível a comunicação de dados entre os sistemas envolvidos, em uma troca de mensagens protocolada (como acontece, por exemplo, com as mensagens HTTP);
- a um nível seguinte, a *interoperabilidade sintática* ou *camada de representação* pressupõe a capacidade de interpretação dos esquemas de codificação da informação transacionada (como o recurso do XML para a representação da informação, ou a sua codificação em um sistema de representação do UNICODE);
- a outro nível superior, a *interoperabilidade estrutural* ou *camada de espaço de atributos* designa o requisito da com-

preensão dos modelos e esquemas que informam e organizam a informação (como o esquema EAD⁷ ou MODS⁸);

- e, finalmente, com a *interoperabilidade semântica* ou *camada de espaço de valores* se desenha o reconhecimento na definição dos contextos significativos da informação e da sua inscrição em um domínio de conhecimento (como permitem a aplicação de vocabulários controlados, vocabulários de valores, e estruturas de organização de conhecimento) (Baker et al., 2003; Zeng, 2019).

O objetivo geral da sobreposição coordenada destes diferentes níveis de interoperabilidade é que os sistemas envolvidos na transação tenham, sobre os objetos concretos a que os metadados se referem, um conhecimento operacional significativo, e que com essa informação possam realizar as funções de processamento que lhes são próprias.

A preocupação e necessidade de desenvolvimento de instrumentos que facilitem a interoperabilidade de metadados é naturalmente mais relevante em ambientes heterogêneos. Mesmo a comunidade de prática da catalogação está sujeita a essa heterogeneidade que se manifesta em diferentes modelos de registros bibliográficos, ainda que, à escala mundial, sob a posição dominante do MARC21. Tal domínio não exclui a coexistência com outros esquemas para a descrição de recursos bibliográficos, por vezes até no seio de uma mesma organização local, coexistência que se pode explicar, por exemplo, pelo tipo de aplicações e sistemas que são utilizados para o controle aplicativo de diferentes componentes do sistema de informação bibliográfica de uma instituição, e pelos esquemas a que recorrem para a modelagem conceitual dessa informação:

⁷ Encoded Archival Description

⁸ Metadata Object Description Schema

MARC21, para o catálogo bibliográfico, *Dublin Core* para o repositório institucional, por exemplo.

Chan e Zeng (2006) sistematizaram as diferentes técnicas usadas na interoperabilidade de esquemas (equivalente, na nossa enumeração, aos níveis superiores de *interoperabilidade estrutural e semântica*), identificando os processos de *derivação*, a definição de *perfis de aplicação*, a *transposição* (*crosswalk*) e a *comutação* (*switching-across*).

Com a *derivação*, um novo esquema de metadados é criado a partir de um esquema estabelecido, seja por simplificação, por enriquecimento, ou ambos. O exemplo dado por Chan e Zeng (2006) da derivação do MARC21 em MODS merece uma atenção particular, porque se um processo de derivação enquadra a origem do esquema MODS, a sua evolução e contexto de aplicação foi separando-o do esquema-fonte original, tornando mais complexos os processos de interoperabilidade dos modelos e das suas instâncias (isto é, da aplicação do modelo à representação de um objeto informacional concreto), aproximando-os mais do modelo de *transposição*, a ser apresentado a seguir.

Os *perfis de aplicação* permitem definir um conjunto de regras estruturais que se adequam ao uso particular de um esquema, ou grupos de esquemas, em particular as regras de cardinalidade dos diversos elementos do esquema, criando para um uso específico – dado pelo tipo de objetos, ou de comunidade, ou seguindo outro critério relevante – um modelo *ad hoc* mais adequado e próximo da aplicação que pretende cobrir.

A *transposição*, ou *crosswalk*, procede a um mapeamento entre os esquemas envolvidos, nos seus diversos elementos, procurando representar no esquema-destino a informação registrada no esquema-fonte, sem que seja alterada, pelo menos completamente, a carga semântica e outros atributos. Este tipo de processo envolve um esforço muito significativo na concepção e manutenção dos

instrumentos computacionais que o executam, requerendo um conhecimento aprofundado dos esquemas envolvidos, a nível sintático e semântico, para além do conhecimento mais próximo da instanciação dos esquemas. Para além disso, um ambiente muito rico de esquemas obriga a conceber um número de transpositores em escala de dificuldade (por exemplo, quatro diferentes esquemas exigem 12 transpositores para garantir a transposição em dois sentidos entre os quatro esquemas; com cinco, aquele número passa para 20, e assim por diante).

Um último tipo de processo apresentado por Chan e Zeng (2006), o da *comutação*, pretende diminuir o custo da criação dos transpositores, oferecendo a ideia de um esquema comutador central, com o qual os restantes se mapeiam. A ideia da comutação está na origem do desenvolvimento do UNIMARC, um modelo para a estruturação de um registro bibliográfico que teve por primeiro objetivo a facilitação da troca da informação bibliográfica a nível internacional (International Federation of Library Associations and Institutions [IFLA], 2008), e que depois evoluiu para suporte a outras funções, entre as quais a de basear a modelagem de dados de sistemas de informação bibliográfica.

Haslhofer e Klas (2010) concentraram-se na compreensão das diferentes abordagens para a interoperabilidade de metadados, começando por salientar a vantagem da *normalização de modelos* (que, no universo do controle bibliográfico, não encontra uma entidade central com força suficiente para determinar a aprovação de um modelo sobre os outros), mas recuperando ainda a possibilidade dessa normalização se verificar através de *metamodelos* construídos com linguagens formais de modelagem.

Na falta de uma e outra abordagem – falta explicada, na maior parte dos casos, pela inexistência de estímulos e capacidades suficientes para ultrapassar os custos e a sua percepção na adoção de soluções de maior funcionalidade (Arms et al., 2002) – é ainda

possível recorrer à *reconciliação de modelos*. Esta *reconciliação* permite, a um custo de maior complexidade (porque deve resolver as heterogeneidades nos diferentes níveis em que se executa), estabelecer processos de *transposição* ou *comutação*:

- Pelo *mapeamento de linguagens* usadas na expressão dos metadados;
- Do *mapeamento de esquemas*, tal como identificados por Chan e Zeng (2006);
- E da *transformação da instância*, em que os metadados sob um modelo-fonte, e referentes a um determinado objeto de qualquer nível de agregação, são reinstanciados para o mesmo ou para diferentes níveis de agregação sob um modelo-alvo (Haslhofer & Klas, 2010).

Por trás destas sistematizações pode encontrar-se um conjunto de dificuldades de implementação, que se verificam quer nas fases de mapeamento conceitual, ou de esquema, quer na fase de instanciação dos modelos na transposição.

Estes problemas tornam expectável a degradação da qualidade da informação transacionada, por perda de informação em cada fase do processo de transposição. Fatores como diferenciação no domínio terminológico ou linguístico, na granularidade, na linguagem formal de expressão da estrutura de metadados, na utilização de regras não escritas ou prescritas sem reconhecimento na transposição, afetam negativamente os processos de interoperabilidade.

5.2. Aplicações em desenvolvimento

Coexistindo com procedimentos que podem ser classificados como tradicionais, ancorados em sistemas de informação particularmente

orientados para a catalogação de recursos impressos, assiste-se ao surgimento de projetos desenvolvidos com o propósito de integrar os conceitos da família FR e, ao mesmo tempo ou alternativamente, de publicar a informação bibliográfica de acordo com os requisitos da Web Semântica. Ou seja, projetos que têm em vista a *frbrização* e/ou *rdfização* do catálogo.

O suporte do modelo FRBR tem servido de meta para várias iniciativas, de escala diferenciada (Decourselle, Duchateau & Lumineau, 2015; Hickey & O'Neill, 2009). Estas experiências são ainda acompanhadas por uma tentativa de entidades normalizadoras, como é o caso da IFLA, de tornar a sua gama de formatos UNIMARC mais amigável à hospedagem de informação FRBR, como se pode verificar nas atualizações dos formatos.

Uma das experiências que tem conseguido apresentar um desenvolvimento consistente, e promovido oficinas de trabalho abertas ao público por todo o mundo, é o projeto RIMMF (*RDA in Many Metadata Formats*), um protótipo de um sistema de catalogação baseado em RDA mas igualmente capaz de importação de registos MARC21 (Dunsire, Fritz & Fritz, 2016).

A perspectiva de publicar a informação bibliográfica como *Linked Data* tem motivado o interesse da comunidade de catalogação, que aqui vê uma oportunidade para tornar a sua informação, o resultado do seu esforço, mais reutilizável por outras comunidades de prática e de conhecimento, através de um conjunto de procedimentos normalizados que facilitam a integração e a reutilização (Baker et al., 2011; Berners-Lee, 2009; Borst, Fingerle, Neubert & Seiler, 2010). Ao mesmo tempo, podem os repositórios de dados ligados servir como *comutadores*, resolvendo ou ajudando a resolver o problema de troca de informação no próprio interior da comunidade.

A iniciativa BIBFRAME, promovida por um consórcio de organizações liderado pela Library of Congress, faz convergir um e outro movimento, propondo uma revisão da informação bibliográfica

modelada e expressa usando os instrumentos da Web Semântica, enquanto, pelo modelo RDA, se inspira indiretamente no modelo conceitual dos FRBR para organizar as relações entre as diversas entidades referenciadas na catalogação (Kroeger, 2013).

Os principais instrumentos para o processo de remodelagem continuam a ser a *transposição* a partir do MARC, o que se explica bem por um conjunto de fatores relevantes:

- O peso da base de conhecimento da comunidade de prática;
- A indisponibilidade de sistemas integrados para a gestão da informação bibliográfica que recorram ao BIBFRAME;
- A dimensão apreciável da informação bibliográfica registrada em MARC;
- O processo de amadurecimento do BIBFRAME, que ainda não terminou (Szeto, 2013; Xu, Hess & Akerman, 2018).

Assiste-se, neste momento, a uma reavaliação dos modelos de catalogação, que antecipam alterações do papel do catalogador do futuro.

Esse papel, no século XXI, será sobretudo o de *catalisador*, alguém responsável pela identificação dos vocabulários relevantes para a descrição de um recurso, e eventualmente pela sua manutenção, que integrará a responsabilidade de organização de estruturas do conhecimento cujo impacto apenas agora se começa a entrever.

6. Considerações finais

Atualmente as unidades de informação do mundo todo estão disponibilizando e dando acesso à informação, tanto em meio analógico (bibliotecas tradicionais) quanto por meio eletrônico ou digital (bibliotecas, repositórios e bases de dados online).

Nesse universo híbrido de organização e acesso às informações, muitas teorias futuristas exageradas e extremistas surgiram durante todo o século XX, dentre as quais cabe citar: o fim do registro do conhecimento utilizando o papel como suporte, o acesso de todo e qualquer documento em formato eletrônico disponível em bibliotecas e repositórios digitais, a dispensa de intermediários nas consultas às bases de dados bibliográficas e a autonomia do usuário na busca de informações, a transformação dos catálogos em versões mais eficientes do que as fichas catalográficas, entre outras que, com o tempo, percebeu-se que eram imprecisas e exageradas.

Assim, no início do século XXI, com base nas evidências dessa realidade híbrida, onde documentos tradicionais convivem com aqueles em formato digital, surgem novas previsões, agora relacionadas ao potencial das bibliotecas e repositórios digitais, pois os aspectos que interferem na disponibilização de documentos em bibliotecas e repositórios digitais ainda são os mesmos que impediram as previsões anteriores de acontecer: direitos autorais, preferência de alguns usuários pelo formato tradicional e o problema da preservação da informação, entre outros.

Pode-se dizer que, atualmente, o problema básico da Ciência da Informação está em facilitar o acesso à informação pelos usuários, reconhecendo a importância das bibliotecas, tanto as tradicionais quanto as digitais, que selecionam e reúnem informação visando uma comunidade específica, e dos sistemas de recuperação da informação que precisam ser projetados e desenvolvidos de forma a facilitar uma localização mais eficiente das informações contidas nos documentos de uma dada coleção.

Outro ponto importante nesse contexto, é reconhecer que existem pontos em comum entre o ambiente analógico e digital, pois ambos realizam as funções básicas de selecionar e adquirir; tratar, organizar e armazenar; dialogar com o usuário sobre uma consulta, entre outras, que contribuirão para um atendimento das necessidades dos

usuários de uma maneira mais efetiva. Isso porque, independente da terminologia utilizada pela Ciência da informação, o objetivo principal sempre será facilitar o acesso às coleções que já existiam há muito tempo, com variada dificuldade de acesso, e a coleções nativo-digitais, com suas próprias dificuldades de acesso que podem ser as mesmas verificadas no universo analógico ou outras, relacionadas à forma de construção informática do ambiente digital. Uma coisa é certa neste cenário: ter registros do conhecimento disponíveis na Internet não significa que a sua disponibilidade de acesso será para todos.

Atualmente, com os resultados das modelagens conceituais da família FR, o lançamento do RDA e as discussões do BIBFRAME, alguns questionamentos são necessários. Primeiramente, percebe-se uma facilidade de interoperabilidade entre as arquiteturas de software para gerenciamento de informações, tornando o uso de formatos de intercâmbio baseados na norma ISO 2709, como o MARC21, menos relevantes (apesar de amplamente utilizados ainda).

Outra questão importante está relacionada à adequação dos formatos de intercâmbio, projetados para substituir a ficha catalográfica permitindo a catalogação por suporte físico da informação, de acordo com o código de catalogação vigente, o AACR2r, a essa nova realidade híbrida.

Apesar de algumas bibliotecas por todo o mundo estarem em fase inicial de mudança para a norma RDA em seus catálogos, implementando essa nova norma do formato MARC, em suas recomendações, o Working Group on the Future of Bibliographic Control (2008, p. 24) afirma que “[...] o MARC, é baseado em técnicas de quarenta anos para gerenciamento de dados e está fora de sintonia com os estilos de programação de hoje.” Ainda de acordo com esse grupo de trabalho da Library of Congress, “[...] o conjunto Z39.2 / MARC existente não é um ponto de partida apropriado para um novo formato de dados bibliográficos, devido às limitações impos-

tas pelos formatos do passado.” (Working Group on the Future of Bibliographic Control, 2008, pp. 24).

Nesse contexto, a Library of Congress tem trabalhado em duas propostas de padrão de metadados para catalogação pela norma RDA: o MODS (*Metadata Object Description Schema*), uma iniciativa desenvolvida pela *Network Development and MARC Standards Office at the Library of Congress*; e o BIBFRAME, que utiliza os recursos de *Linked Data* para a implementação da RDA em catálogos, bibliotecas e repositórios digitais e bases de dados.

O BIBFRAME tem sido visto como uma possibilidade de substituição do MARC, servindo como um padrão de codificação não só para RDA como também para outros padrões de conteúdo. Desta forma, não é apenas uma substituição do MARC, tendo em vista os objetivos que cada um assiste, mas também é um ambiente / modelo / ontologia com foco em dados bibliográficos (Li, 2018).

O que é especialmente importante nessa mudança de paradigma da catalogação é a discussão e a análise, com o merecido cuidado, sobre a continuidade do formato MARC com a implementação da RDA. Os sistemas baseados em MARC são estruturados para que a entrada de dados aconteça por tipo de suporte de informação, incompatível com a proposta da RDA de catalogar com base nas entidades do modelo FRBR. Além disso, a nova norma também trabalha com a representação dos diferentes relacionamentos entre as entidades do FRBR, possibilitando uma navegação mais dinâmica do usuário no catálogo da biblioteca. Com a implementação destes relacionamentos no formato MARC, alguns poderão deixar de dar a resposta adequada ao RDA. Outro ponto central é a discussão do RDA a partir do IFLA LRM, visto que este conjuga aspectos da descrição bibliográfica e da descrição museológica.

Ao final da discussão proposta, compreende-se que os instrumentos, amplamente discutidos nas comunidades de prática, têm servido, também, para o vínculo dos aspectos teórico-conceituais

da organização e representação da informação e consolidado, ainda que de um modo não muito transparente, claramente perceptível e naturalizado, a perspectiva de comunicação centrada no usuário.

Referências Bibliográficas

- Arms, W. Y., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J., ... Van de Sompel, H. (2002). A Spectrum of Interoperability. *D-Lib Magazine*, 8(1). doi:10.1045/january2002-arms.
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., ... Zeng, M. (2011). *Library Linked Data Incubator Group Final Report*. Cambridge, MA: W3C. Recuperado 4 de agosto de 2017, de <http://www.w3.org/2005/Incubator/lld/XGR-lld/>
- Baker, T., Blanche, C., Brickley, D., Duval, E., Heery, R., Johnston, P., ... Sugimoto, S. (2003). *Principles of Metadata Registries* (A White Paper of DELOS Working Group on Registries). Recuperado de <https://pdfs.semanticscholar.org/01ea/e200c915fbb38faf2584e87230bb15d2d683.pdf>
- Berners-Lee, T. (2009). *Linked Data: Design Issues*. Recuperado de <http://www.w3.org/DesignIssues/LinkedData.html>
- Bicalho, L. M., & Oliveira, M. (2011). Aspectos conceituais da multidisciplinaridade e da interdisciplinaridade. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 16(32), 1-26. doi:10.5007/1518-2924.2011v16n32p1
- Borst, T., Fingerle, B., Neubert, J., & Seiler, A. (2010). How do Libraries Find their Way onto the Semantic Web? *Liber Quarterly*, 19(3-4), 336-343. doi:10.18352/lq.7970.
- Brascher, M., & Café, L. (2008). Organização da informação ou organização do conhecimento? In *Anais do IX Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), São Paulo, Brasil*. Recuperado de <http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142>
- Chan, L. M., & Zeng, M. L. (2006). Metadata Interoperability and Standardization – A Study of Methodology: Part I: Achieving Interoperability at the Schema Level. *D-Lib Magazine*, 12(6). doi:10.1045/june2006-chan.
- Decourselle, J., Duchateau, F., & Lumineau, N. (2015). A Survey of FRBRization Techniques. In S. Kapidakis, C. Mazurek, & M. Werla (Eds.), *Research and Advanced Technology for Digital Libraries: Proceedings of the 19th International Conference on Theory and Practice of Digital Libraries (TPDL), Poznań, Poland* (pp. 185-196). Cham, Switzerland: Springer. doi:10.1007/978-3-319-24592-8_14.
- Dunsire, G., Fritz, D., & Fritz, R. (2016). *Instructions, interfaces, and interoperable data: the RIMMF experience with RDA*. Paper presented at IFLA WLIC 2016, Columbus, OH, USA. Recuperado de <http://library.ifla.org/1324/>.
- Feitosa, A. (2006). *Organização da informação na web: das tags à web semântica*. Brasília: Thesaurus.

- Fogel, J. (1979). Relations of the concepts 'information' and 'knowledge'. *International Forum on Information and Documentation*, The Hague, 4(1), 21-24.
- Gomes, H. F. (2001). Interdisciplinaridade e Ciência da Informação: de característica a critério delimitador de seu núcleo principal. *DataGramaZero*, 2(4), 1-7. Recuperado de <http://hdl.handle.net/20.500.11959/brapci/5176>
- Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys*, 42(2), 1-37. doi:10.1145/1667062.1667064.
- Hickey, T. B., & O'Neill, E. T. (2009). *FRBRizing OCLC's WorldCat*, (March 2014), 37-41. doi:10.1300/J104v39n03_15.
- International Federation of Library Associations and Institutions. (1998). *Functional requirements for bibliographic records (FRBR)*. Munich: K. G. Sauer. Com acréscimos e correções em 2009.
- International Federation of Library Associations and Institutions. (2008). *Manual UNIMARC: formato bibliográfico*. (R. M. Galvão & M. P. Lopes, Eds.). Lisboa: Biblioteca Nacional de Portugal.
- International Federation of Library Associations and Institutions. (2009). *Functional requirements for authority data (FRAD)*. Munich: K. G. Sauer. Com acréscimos em 2013.
- International Federation of Library Associations and Institutions. (2011). *Functional requirements for subject authority data (FRSAD) – A conceptual model*. Munich: De Gruyter Saur.
- International Federation of Library Associations and Institutions. (2017). *IFLA library reference model: a conceptual model for bibliographic information*. Den Haag, Netherlands: Author.
- Jovanovich, E. M. S., Souza, L. P. P. de, Trevisan, L. C., Ottonicar, S. L. C., & Castro Filho, C. M. de (2017). A interdisciplinaridade na constituição e evolução da Ciência da Informação. In *Anais do XVIII Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), Marília, UNESP, Brasil*. Recuperado de <http://hdl.handle.net/20.500.11959/brapci/104802>
- Kaufman, K. (2014). *Metadata Object Description Schema (MODS)* [PowerPoint Presentation]. Recuperado 4 de agosto de 2017, de <http://www.slideserve.com/kenley/metadata-object-description-schema-mods>.
- Klim, I. L. (1981). Developing a theory of bibliographic description. *Scientific and Technical Information Processing*, 1, 50-58.
- Kobashi, N. Y. (1996). Análise documentária e representação da informação. *INFORMARE – Cadernos do Programa de Pós-Graduação em Ciência da Informação*, Rio de Janeiro, 2(2), 5-27.
- Kroeger, A. (2013). The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future. *Cataloging and Classification Quarterly*, 51(8), 873-890. doi:10.1080/01639374.2013.823584.
- Lança, T. A., Amaral, R. M., & Gracioso, L. S. (2018). Multi e interdisciplinaridade nos programas em Ciência da Informação brasileiros. *Perspectivas em Ciência da Informação*, 23(4), 150-183. Recuperado de <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/3608>

- Li, K. (2013). *Introduction to BIBFRAME* [PowerPoint Presentation]. Recuperado 4 de agosto de 2017, de https://www.slideshare.net/islanderlee/introduction-to-bibframe?from_action=save.
- Lima, V. M. A. (1998). *Terminologia, comunicação e representação documental* (Dissertação de mestrado em Ciências, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, Brasil). Recuperado de <https://www.teses.usp.br/teses/disponiveis/27/27143/tde-11052004-122839/pt-br.php>.
- Mering, M. (2014). *The RDA workbook: learning the basics of Resource Description and Access*. Santa Barbara, CA: Libraries Unlimited.
- Mey, E. S. A., Silveira, N. C. (2009). *Catálogo no plural*. Brasília: Briquet de Lemos.
- Moraes, M., & Carelli, A. E. (2016). A interdisciplinaridade na Ciência da Informação pela perspectiva da análise de citações. *Em Questão*, 22(1), 137-160. doi:10.19132/1808-5245221.137-160
- Nisenbaum, M. A., & Pinheiro, L. V. R. (2016). Métodos para medição da interdisciplinaridade adotados na Ciência da Informação. In *Anais do XVII Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB)*, Salvador, BA, Brasil. Recuperado 5 de maio de 2018, de <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/3874>
- Ollet, P. (2018). *Tratado de documentação*. Brasília: Briquet de Lemos. Recuperado 13 de maio de 2019, de https://repositorio.unb.br/bitstream/10482/32627/1/LIVRO_TratadoDeDocumenta%ç%a7%ç%a3o.pdf
- Pinheiro, L. V. R., & Loureiro, J. M. M. (2007). Pilares conceituais para mapeamento do território epistemológico da Ciência da Informação: disciplinaridade, interdisciplinaridade, transdisciplinaridade e aplicações. In V. B. Pinto, L. E. Cavalcante, & C. Silva Neto (Orgs.), *Ciência da Informação: abordagens transdisciplinares, gêneses e aplicações* (pp. 71-104). Fortaleza: Edições UFC.
- Pinto, V. B. (2007). Interdisciplinaridade na Ciência da Informação: aplicabilidade sobre a representação indexal. In V. B. Pinto, L. E. Cavalcante, & C. Silva Neto (Orgs.), *Ciência da Informação: abordagens transdisciplinares, gêneses e aplicações* (pp. 105-142). Fortaleza: Edições UFC.
- Ranganathan, S. R. (2009). *As cinco leis da biblioteconomia*. Brasília: Briquet de Lemos.
- Riecken, R. F. (2006). Frame de temas potenciais de pesquisa em Ciência da Informação. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, 4(1), 43-63. Recuperado de <http://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2044/2170>.
- Santos, A. P. L. dos., & Rodrigues, M. E. F. (2013). A interdisciplinaridade e a ciência da informação: algumas reflexões. *Liinc em revista*, 9(1), 335-349. doi:10.18617/liinc.v9i1.498
- Santos, I. L. dos, Chacon, W., & Veras, J. (2015). Uma análise da Ciência da Informação brasileira no contexto da interdisciplinaridade. *Biblionline*, 11(2), 218-231. Recuperado de <https://periodicos.ufpb.br/index.php/biblio/article/view/26030>
- Santos Neto, J. A. dos, Santos, J. C. dos, Teles, P. S., & Valentim, M. L. P. (2017). Interdisciplinaridade no contexto da Ciência da Informação: correntes e questionamentos. *Em Questão*, 23(1), 9-35. doi:10.19132/1808-5245231.9-35

- Sfez, L. (2000). *Crítica da comunicação* (2ª ed). São Paulo: Edições Loyola.
- Silva, A. K. A., Lima, I. F. de, & Araújo, C. A. Á. (2009). Desvelando a interdisciplinaridade da ciência da informação: o enfoque dos alunos do PPGCI/UFGM. *Ciência da Informação*, 38(1), 31-44. Recuperado de <http://revista.ibict.br/ciinf/article/view/1253>
- Silva, R. J. (2005). Faces da pesquisa e da interdisciplinaridade em ciência da informação no Brasil. In *Anais do VI Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), Florianópolis, SC, Brasil*. Recuperado 9 de abril de 2018, de <http://enancib.ibict.br/index.php/enancib/vienancib/paper/view/1688/825>.
- Smiraglia, R. P. (2002). Further Reflections on the Nature of 'A Work': An Introduction. *Cataloging & Classification Quarterly*, 33(3/4), 1-11. doi:10.1300/J104v33n03_01.
- Szeto, K. (2013). Positioning Library Data for the Semantic Web: Recent Developments in Resource Description. *Journal of Web Librarianship*, 7(3), 305-321. doi:10.1080/19322909.2013.802584.
- Working Group on the Future of Bibliographic Control. (2008). *On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control*. Recuperado 29 de janeiro de 2019, de <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.
- Xu, A., Hess, K., & Akerman, L. (2018). From MARC to BIBFRAME 2.0: Crosswalks. *Cataloging & Classification Quarterly*, 56(2-3), 224-250. doi:10.1080/01639374.2017.1388326.
- Zafalon, Z. R. (2014). *Scan for MARC: conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico*. São Paulo: Editora Unesp. Recuperado de <http://hdl.handle.net/11449/113734>.
- Zafalon, Z. R. (2017). Recurso informacional e representação documental. In Z. R. Zafalon, P. R. Dal'Evedove (Orgs.), *Perspectivas da representação documental: discussões e experiências* (pp. 125-144). São Carlos: CPOI.
- Zeng, M. L. (2019). Interoperability. *Knowledge Organization*, 46(2), 122-146. doi:10.5771/0943-7444-2019-2-122.
- Zins, C., Debons, A., Beghtol, C., Buckland, M. K., Davis, C. H., Dodig-Crnkovic, G., ... & Smiraglia, R. P. (2007). Mapa do conhecimento da Ciência da Informação: implicações para o futuro da área. *Brazilian Journal of Information Science*, 1(1), 3-32. Recuperado 10 de abril de 2017, de <http://www.brapci.ufpr.br/brapci/index.php/article/view/0000008751>.

CAPÍTULO 4
PERSPECTIVAS DE MODELAGEM CONCEITUAL
NA ELABORAÇÃO DE SISTEMAS
DE ORGANIZAÇÃO DO CONHECIMENTO:
SUBSÍDIOS PARA O ENSINO
PERSPECTIVES FOR CONCEPTUAL MODELING
IN THE DEVELOPMENT OF KNOWLEDGE
ORGANIZATION SYSTEMS:
SUBSIDIES FOR TEACHING

Gercina Ângela de Lima

Universidade Federal de Minas Gerais

limagercina@gmail.com

ORCID: 0000-0003-0735-3856

Maria Luiza de Almeida Campos

Universidade Federal Fluminense

Programa de Pós-Graduação em Ciência da Informação

maria.almeida@pq.cnpq.br

ORCID: 0000-0002-9253-3706

Nair Kobashi

Universidade de São Paulo

nairkobashi@gmail.com

ORCID: 0000-0002-5282-7363

Resumo: Toda representação de conhecimento está sujeita às tecnologias disponíveis em um dado momento/espço temporal. Essa afirmativa se insere perfeitamente no âmbito dos estudos de organização do conhecimento e da recuperação da informação. A afirmativa é também verdadeira no que tange à elaboração de “Modelos de domínios”. Nesse sentido, este artigo apresenta uma sistematização dos conhecimentos que advêm da Biblioteconomia e da Ciência da Informação para a elaboração de modelos de domínios. Estes últimos podem subsidiar o ensino de teorias e métodos de construção de ferramentas de tratamento e recuperação de informação, tais como os sistemas de classificação, os tesouros, as taxonomias e as ontologias.

Palavras-Chave: Organização do Conhecimento. Modelagem conceitual; Instrumentos de representação de informação: Instrumentos de recuperação de informação.

Abstract: All knowledge representation is sensitive to the technologies available at a given time and space. This statement is perfectly within the scope of knowledge organization and information retrieval studies. It is also true in regard to the development of Domain Models. In this sense, we present a systematization of knowledge that comes from Library and Information Science for teaching of theories and methods of constructing instruments for information processing and retrieval, such as classification systems, thesauri, taxonomies, and ontologies.

Keywords: Knowledge organization; Conceptual modeling; Information representation instruments. Information retrieval instruments.

1. Considerações iniciais sobre o espaço informacional e a representação de domínios de conhecimento

Modelos conceituais de domínios são produtos obtidos a partir do reconhecimento de que os campos do saber evidenciam sistemas de conceitos que expressam de forma rigorosa as suas teorias e

métodos. Os modelos conceituais, *lato sensu*, têm por objetivo construir representações de alta qualidade de fenômenos selecionados de algum domínio do conhecimento. No âmbito da Organização do Conhecimento, os modelos resultantes pretendem facilitar a representação e a recuperação de informações em contextos diversos.

Essa questão vem preocupando o campo dos estudos informacionais desde a sua gênese, fato que se torna mais agudo com o desenvolvimento contínuo das tecnologias da informação e comunicação. Esses desafios vêm exigindo pesquisas renovadas com a consequente elaboração de novas teorias e métodos. Nesse sentido, propomos discutir, neste trabalho, diferentes perspectivas teóricas e metodológicas sobre a modelagem de domínios aplicados à elaboração de ferramentas de representação e recuperação de informação, no contexto das redes eletrônicas de informação e comunicação.

Trata-se, portanto, de revisitar teorias e métodos consagrados e também as propostas contemporâneas sobre a construção de modelos conceituais de domínios. São bases fundantes das reflexões os aspectos ontológicos, as questões terminológicas e as questões lógicas e semânticas subjacentes à classificação e à representação de conhecimentos. Esses conhecimentos, como já foi sinalizado por diferentes pesquisadores da área, são indispensáveis para construir ferramentas robustas de representação do conhecimento para viabilizar a difusão, a circulação e a apropriação social da informação. Não por acaso, dedicam-se a essas temáticas os pesquisadores da Ciência da Informação, da Inteligência Artificial, da Psicologia e das Ciências da Linguagem.

No domínio da Biblioteconomia e da Ciência da Informação, teóricos como Ranganathan (1967), Bliss (1929), Feibleman (1954), Vickery (1997) e o próprio Classification Research Group – CRG (Wilson, 1972) construíram teorias com a finalidade de representação de domínios, visando à organização de documentos, mas que, de alguma forma, apresentavam princípios que possibilitavam a representação independentemente dos próprios domínios de saber.

Nesse contexto, dos estudos relacionados à representação e à recuperação de informação, especificamente daqueles ligados às ferramentas atualmente existentes no cenário informacional para a realização de tais processos, existem dúvidas sobre a aplicabilidade dessas ferramentas em ambientes onde questões semânticas se impõem. Até os últimos anos, um dos papéis exercidos pelo profissional de informação se restringia ao desenvolvimento de ferramentas semânticas capazes de representar um dado domínio de conhecimento (por meio de linguagens documentárias como tesouros, taxonomias e esquemas de classificação), possibilitando certa precisão e compatibilidade entre a linguagem utilizada para o tratamento informacional e a linguagem utilizada pelo usuário para a recuperação da informação. Nesse cenário, impera a Web Sintática.

Como apresenta Karin Breitman, a “Internet atual pode ser definida como a Web Sintática” (Breitman, 2005). Na Web Sintática, os computadores fazem apenas a apresentação da informação, porém, o processo de interpretação fica a cargo dos seres humanos. Dessa forma, a Web Sintática é uma rede desenvolvida para a produção de sentido realizada por humanos. A questão que começou a ser colocada há alguns anos no âmbito dos estudos computacionais e que nos levou a uma nova geração da Web, ou seja, à Web Semântica, foi: por que os computadores não podem realizar o trabalho de interpretação de dados?

No cenário da Web Semântica, em que as ontologias formais se inserem, já é possível, de algum modo, não só a representação do conhecimento de um dado domínio, mas também a derivação do conhecimento sobre esse domínio por meio de uma semântica expressa em linguagens formais, possível de ser explorada computacionalmente para fins de inferência (Campos, 2012).

Dessa forma, nossa intenção neste artigo é apresentar uma sistematização inicial sobre os conhecimentos para a elaboração de

modelos de domínios que possam subsidiar o ensino de teorias e métodos de construção de diferentes ferramentas de tratamento e recuperação de informação, tais como os sistemas de classificação, os tesouros, as taxonomias e as ontologias.

2. Modelos de representação de domínio no âmbito da Web Sintática

O conhecimento é algo dinâmico, que está presente na vida do ser humano desde o nascimento e o acompanha até o último dia de sua vida. No percorrer desse caminho, vão sendo adquiridos novos conhecimentos, que sofrem alterações, sendo necessário serem lapidados na sua concepção para que sejam compreendidos, armazenados e, posteriormente, registrados e disseminados dentro de uma comunidade.

Esse processo é realizado pela mente humana, a partir do ponto de vista de como o indivíduo vê seu mundo e o representa por meio de informações sobre objetos pertencentes à sua realidade. Porém, a realidade não pode ser vista sob um único aspecto, restrito às percepções e sensações de um indivíduo. Surge aí a necessidade de se criar mecanismos para a padronização desse conhecimento, de modo a tornar possível o seu compartilhamento social

Cada estrutura de conhecimento pode ser vista como um objeto, ideia ou evento e, também, como um grupo de atributos, ligado a outra estrutura de conhecimento. De acordo com Lima (2018), à medida que aprendemos, apreendemos novas estruturas e ligações, adicionando informações às estruturas existentes, ou alterando essas estruturas por meio de processos de reestruturação. Portanto, a influência do contexto é fundamentalmente importante na categorização. Assim, qualquer classificação individual que desconsidera o contexto pode ser muito limitada e, portanto, de difícil compartilhamento.

Para organizar o conhecimento de qualquer área, desde sua representação à sua recuperação, estudam-se primeiramente os conceitos que compõem esse campo do conhecimento e as relações entre eles. Categorias e estruturas hierárquicas são a melhor maneira de se organizar o conhecimento para recuperação, pelo óbvio motivo de que a informação estruturada é mais fácil de ser recuperada do que a informação desorganizada. Portanto, pode-se dizer que o ser humano, desde os primeiros momentos de vida, percebe o seu mundo, registra, classifica e dá forma categorial às estruturas que espelham o ambiente externo. Nota-se que essa classificação vem de nossa interação com nosso o ambiente. Se não interagimos com o ambiente, não temos o que classificar; o ambiente, por sua vez, influencia o modo como categorizamos o conhecimento.

A área de Organização do Conhecimento (OC) tem suas raízes históricas nas teorias da Classificação. A Classificação pode ser vista como uma área de conhecimento que procura classificar o conhecimento filosófico; no caso da Organização do conhecimento registrado, procura-se criar sistemas de classificação bibliográfica para a representação do conteúdo temático dos documentos, de modo a aumentar a eficácia da recuperação de informações. As duas abordagens da classificação procuram ligar teoria e prática nos estudos e pesquisas realizadas dentro da área da Biblioteconomia e Ciência da Informação (BCI).

As bases da classificação remontam a Aristóteles, em que um todo pode ser dividido em partes, sucessivamente, formando classes e subclasses. Essa lógica de divisão para representar o conhecimento parte do princípio classificatório de gênero e espécie para formar uma estrutura conceitual hierárquica.

Aristóteles foi um dos primeiros estudiosos a pensar o mundo e categorizá-lo (e mesmo antes, com Platão), há mais de dois milênios, com sua inquietude em saber como o homem poderia representar o seu conhecimento. A teoria clássica de Aristóteles é considerada por

alguns como a hierarquia perfeita, na qual as categorias são definidas apenas pelas propriedades comuns a todos os seus membros. Destarte, veio da Filosofia, por meio dos estudos de Aristóteles, a primeira preocupação em sistematizar o conhecimento com o intuito de torná-lo padronizado, com a utilização de símbolos que representem a realidade, tornando possível a manifestação dessas ideias com um significado convencional. Portanto, ideias e coisas estariam integradas de forma a materializar o pensamento humano:

Destarte, Aristóteles, em seu *Órganon* (Aristóteles, 2010), sistematiza o saber humano propondo um sistema filosófico representado por dez categorias, a partir de proposições que se referem à determinação formal do ser, a saber: Substância, Quantidade, Qualidade, Relação, Duração, Lugar, Ação, Paixão ou Sofrimento, Maneira de ser e Posição. Outras contribuições deixadas por Aristóteles foram as quatro distinções, entre o gênero, o próprio, o acidente e a definição (Morujão, 2006, p. 304) e o esquema tripartitivo de classificação, no qual o conhecimento se divide em ciências teóricas, ciências práticas e ciências poéticas, representando o pensar, o agir e o produzir (Piedade, 1983). Esses estudos de Aristóteles influenciaram o desenvolvimento das teorias de classificação do conhecimento, sendo considerados o embasamento teórico para as classificações bibliográficas que começaram a surgir posteriormente, no século IV. Deve-se mencionar aqui os estudos de Porfírio, que, em sua obra *Isagoge* (Introdução às Categorias), incorporou a lógica aristotélica na categoria substância, considerada como gênero supremo, resultando em cinco predicáveis¹ – gênero, espécie, diferença, propriedade e acidente, que, quando interligados, formam uma lógica sucessiva

¹ Porfírio faz distinção entre predicados e predicáveis. Por exemplo: na expressão «todos os homens são mortais», «mortais» é um predicado, mas é atribuído a «todos os homens» de modo diferente do que o seria, por exemplo, o predicado «bípedes sem penas». O predicável será, então, o modo como certo predicado é atribuído a um sujeito (Morujão, 2006, p. 304).

de subordinação, resultando em uma classificação dicotômica, que tornou-se conhecida como “Árvore de Porfírio”.

Para organizar o conhecimento de qualquer área, da sua representação à sua recuperação, estudam-se primeiramente os conceitos que compõem esse campo do conhecimento e as relações entre eles. O núcleo principal desse processo é a escolha dos conceitos e seus relacionamentos e sob quais aspectos se quer representar determinado domínio. Pode-se dizer, então, que os conceitos são a base para a construção da estrutura conceitual, ou modelagem conceitual, e são as relações entre eles que os mantêm reunidos dentro de um determinado domínio e princípio de representação. Para Green (2008), os relacionamentos são o coração da organização do conhecimento:

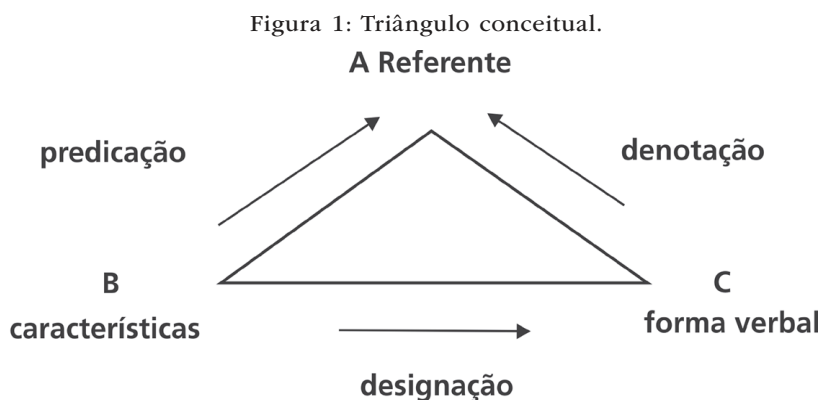
Tentamos localizar informações relacionadas para a necessidade de um usuário; tentamos localizar documentos contendo palavras que possam ser usadas para transportar a informação procurada; e/ou recuperamos documentos indexados por um descritor de assunto relevante ou classificado em uma categoria de assunto relevante. Usamos relacionamentos para navegar entre assuntos. Ajudamos a localizar documentos que juntos dizem mais do que a soma de suas partes. (p. 158, tradução nossa)²

A modelagem conceitual de uma área do conhecimento pode ser influenciada pelo contexto social, cultural, estético, político. A organização de conceitos está diretamente relacionada à capacidade de aprender, supõe a assimilação de novas informações, sua estocagem e sua acomodação dentro de um determinado contexto. Quando

² We attempt to locate information that relates to a user's need; we attempt to locate documents containing words that might be used in conveying the sought-for information; and/or we retrieve documents indexed by a relevant subject descriptor or classed in a relevant subject category. We use relationships to navigate among subjects. We help to locate documents that together say more than the sum of their parts.

se estabelecem as categorias de um domínio do conhecimento, de certa forma, elas são determinadas pela perspectiva do contexto que foi considerado. Segundo Barros (2004, p. 109): “Por contexto compreende-se o enunciado que exprime uma ideia completa, no qual o termo estudado se encontra atualizado. A identificação das características de um conceito num contexto é possível graças aos descritores”.

A definição de um conceito, por meio de suas características e predicados, representado pelo seu símbolo verbal, ou seja, o termo, faz com que ele seja contextualizado dentro de uma área específica do conhecimento. O símbolo isolado de seu contexto, mesmo possuindo propriedades semânticas relevantes – se não forem evidenciadas através de sua definição –, não poderá ser interpretado e distinguido de outros conceitos que formam o sistema conceitual daquele domínio. É necessário, portanto, conforme Dahlberg (1978a), que cada objeto seja definido para fixar os seus limites enquanto conceito pertencente àquele domínio específico. Isso pode ser melhor visualizado e simbolizado, por meio do triângulo conceitual proposto por Dahlberg (1978a), conforme Figura 1:

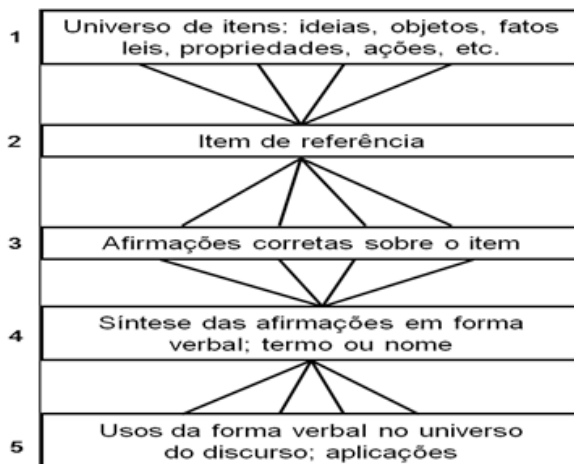


Fonte: Dahlberg, 1978a, p. 144.

O Triângulo Conceitual de Dahlberg tem no ápice o referente, ou seja, o objeto que se quer conceituar; na parte inferior do canto esquerdo, um conjunto de predicados ou as características que formam aquele conceito, e, na parte inferior do canto direito, a forma verbal, que simboliza aquele conceito, representado por um termo.

Dahlberg (1978a) define conceito como “[...] uma unidade de conhecimento, compreendendo afirmativas verificáveis sobre um item selecionado de referência, representado por uma forma verbal”. A autora apresenta um modelo para construção de conceitos (Figura 2), onde estão representadas as relações entre o objeto, o conceito e o termo. O processo mental da formação do conceito se dá por meio de uma linha de pensamento que leva à elaboração do conhecimento, passando por um processo de assimilação da informação pelo cérebro, transformando-a. Após essa elaboração mental, baseada no conhecimento prévio do indivíduo, a unidade de informação se transforma em uma unidade conceitual, que é representada por um termo, o qual possui um único significado, geralmente expresso por símbolos e palavras, com o objetivo de comunicação.

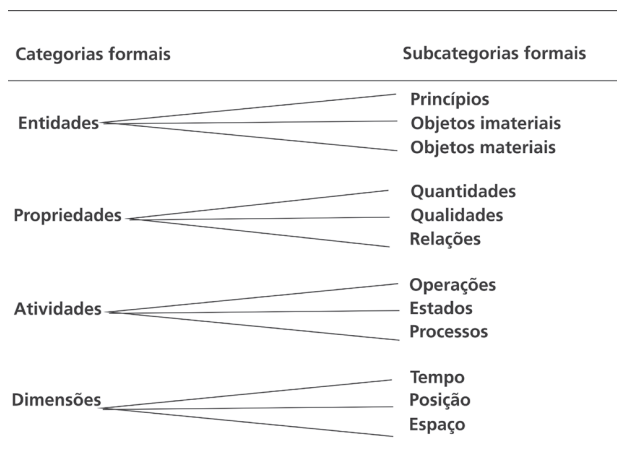
Figura 2: Modelo de construção do conceito.



Fonte: Dahlberg, 1978a, p. 143.

Além disso, Dahlberg propõe quatro categorias formais que se desdobram em 12 subcategorias (Figura 3)³, tendo por base as categorias aristotélicas, que permitem combinação entre elas.

Figura 3: Referentes de conceitos.



Fonte: Dahlberg, 1978a, p. 145.

Outra proposta de categorização foi criada por Shiyali Ramamrita Ranganathan (1892-1972), matemático indiano que se tornou bibliotecário e classificacionista. É um sistema quase matemático, que estrutura o conhecimento de maneira que os assuntos compostos sinteticamente derivam de conceitos “elementares”.

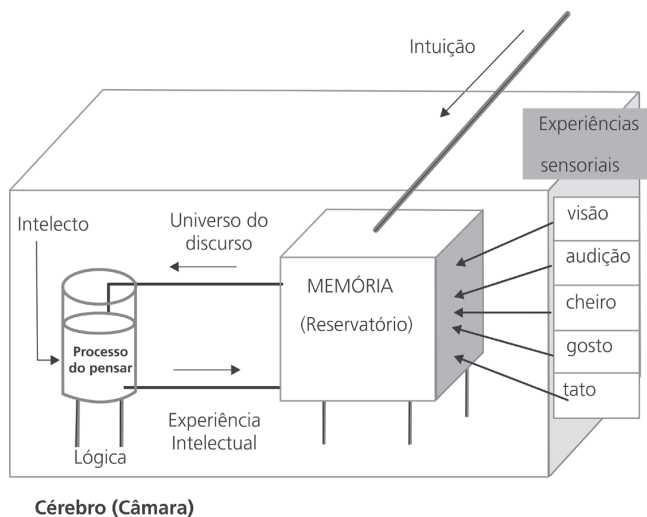
Para Ranganathan (1967), o conceito é produto do pensamento; o conhecimento é representado pela totalidade das ideias armazenadas pela memória humana, e a informação existe quando os conceitos são comunicados. Todos esses processos são sensoriais, permeados pelas experiências cognitivas de cada indivíduo.

³ Não se pretende neste capítulo trazer definições das categorias. Para aprofundar nos estudos, deve-se consultar a literatura citada.

Ranganathan, no *Prolegomena to Library Classification* (1967), define Conceito como “unidade de pensamento de qualquer nível de complexidade”. Considerou o conhecimento como um elemento multidimensional: as interligações de cada conceito espalham-se em muitas direções e, frequentemente, cada assunto constitui-se de uma síntese de vários conceitos múltiplos ligados, o que caracteriza essa abordagem como analítico-sintética.

Nessa abordagem, Ranganathan apresenta uma alternativa à Árvore de Porfírio, que é um método dicotômico composto de dez divisões, organizadas em um eixo único. Propõe, então, um padrão para mapear o conhecimento com a policotomia ilimitada (número ilimitado de divisões das áreas do conhecimento), representado pela Árvore Baniana (*Banyan Tree*) que, em adição a um tronco original, vários troncos secundários podem ser formados de tempos em tempos, ideia que associou a uma árvore de assuntos (Ranganathan, 1967, p. 368). Para ilustrar a formação das ideias, Ranganathan nos apresenta um diagrama esquemático (Figura 4).

Figura 4: Compartimento do cérebro: formação das ideias.



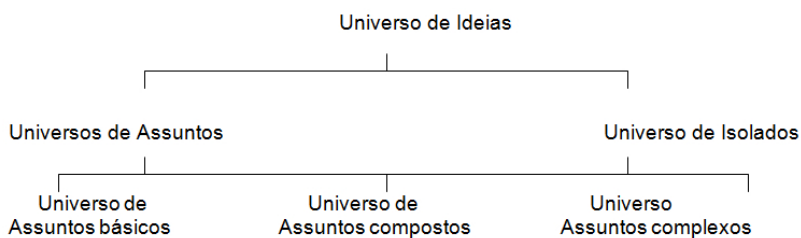
Fonte: Ranganathan, 1967, p. 8.

Assim, Ranganathan (1967, p. 82) descreve o assunto como: "Um corpo de ideias organizadas ou sistematizadas, cuja extensão e intensão devem ser coerentemente com o domínio de interesse e confortavelmente ajustadas à competência intelectual e campo especializado de uma pessoa qualquer."

Todo assunto é oriundo de um assunto básico e seu componente vem de um ou mais conceitos isolados, formando assim o assunto composto. O assunto básico é um assunto sem nenhuma ideia isolada como componente. A ideia isolada (isolado) é alguma ideia ou complexo de ideias, moldada para formar um componente de assunto. Mas, sem seu contexto, ela não é considerada um assunto. Ao estabelecer um Conceito, o modelizador/classificacionista descreve o objeto a partir de um entre inúmeros cenários contextuais possíveis.

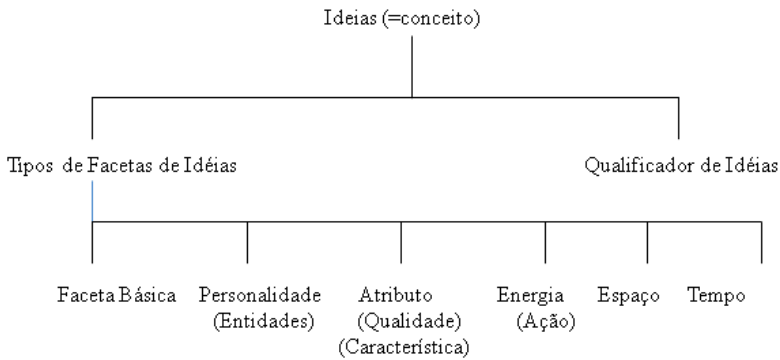
Os esquemas representativos a seguir (Figura 5 e Figura 6) mostram (1) o universo original das ideias em suas relações, seus subuniversos de assuntos e isolados propostos por Ranganathan e (2) os tipos de facetas de ideias propostos por Raghavan.

Figura 5: Universo original das ideias e suas relações segundo Ranganathan.



Fonte: Ranganathan, 1967, p. 85.

Figura 6: Exemplo da categorização de isolados dentro das facetas.



Fonte: Raghavan, 1985, p. 27.

Uma das maiores contribuições de Ranganathan foi a criação das categorias fundamentais, expressas por meio do acrônimo PMEST⁴ (*personality, matter, energy, space, time* – em português: personalidade, matéria, energia, espaço e tempo), que é o princípio normativo adotado para organizar um Universo de Assuntos, ou seja, um “corpo” de conhecimento organizado e sistematizado.

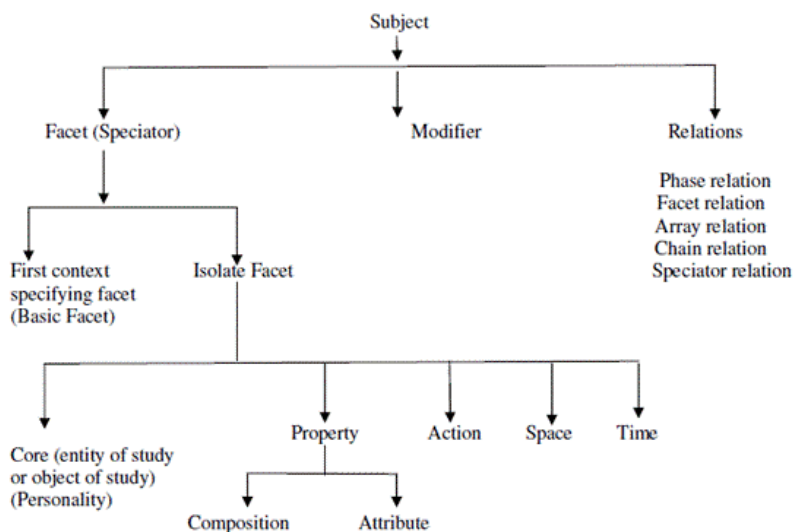
Os fundamentos da Linguística e da Terminologia também contribuem para a modelagem conceitual, porque é preciso considerar o mapeamento de um determinado domínio como um sistema de conceitos, cujos elementos são definidos a partir de diferentes tipos de relações: equivalência (sinonímia), lógicas (gênero-espécie, analítica e de oposição), ontológicas (partitiva, sucessão e material-produto), de efeito (causalidade, instrumental e descendência), entre outras, que formam a “estrutura” de uma linguagem comunicativa. Sendo assim, a Terminologia contribui, segundo Campos

⁴ Não se pretende, neste estudo, detalhar as categorias de Ranganathan. Caso queira aprofundar nos estudos, deve-se consultar a literatura citada.

(2010), com os princípios teóricos para a criação de sistemas de conceitos bem estruturados. Isso acontece seja com a Teoria Geral da Terminologia, iniciada por Wüster, que se baseia nas dicotomias “termo e palavra”, seja com os princípios da Teoria Comunicativa da Terminologia (TCT), proposta por Cabré. Esta última traz novos subsídios teóricos para identificar as funções (pragmáticas, semânticas e sintáticas) dos termos e os aspectos de forma (variação linguística) e conteúdo, além da relação existente entre o termo (como unidade de conhecimento) e o conceito que representa. A determinação do termo para representar o conceito tem papel essencial na Recuperação da Informação (RI), pois sempre podem haver diferentes termos para representar um conceito e, em certa medida, a escolha por um deles será prescritiva.

Neelameghan (1971) desenvolveu um modelo para representar o conhecimento a partir de uma estrutura de facetas generalizadas de assuntos, conforme Figura 7, no qual o processo de formação de conceitos (assuntos) se realiza a partir das formas linguísticas universais conforme expostas por linguistas como Chomsky, Fodor e Katz. Essa estrutura pode ter modelos específicos para diferentes campos de assunto, com a formação das palavras e a formação de assuntos que ocorrem de forma instantânea. Nesta proposta, o autor baseou-se nos princípios da Teoria Geral da Classificação para criar uma sequência útil, utilizando a proposta de Ranganathan, para a formação de assuntos a partir da criação de facetas sintáticas. O modelo proposto por Neelameghan (1971) pode auxiliar no mapeamento de um domínio do conhecimento, a partir da modelagem contextual, por meio de análise sintática.

Figura 7: Estrutura de facetas generalizadas de assuntos desenvolvida por Neelameghan.



Fonte: Dutta & Dutta, 2014, p. 60.

Dutta e Dutta (2014) descrevem os modos de formação de palavras polilexicais na linguística e os comparam com os diferentes modos de formação de assuntos compostos e complexos, como retratado por Ranganathan (1967), Neelameghan (1971, 1973) e Sen (2009). Os autores concluem que há semelhanças entre os modos de formação de palavras e de assuntos.

Esses elementos lógicos e sistematizadores de domínios apresentados por teóricos do âmbito da Biblioteconomia e da Ciência da Informação evidenciam toda uma linha de discussão travada neste campo com vista à representação de conhecimento para a ordenação de documentos e informação.

Nesse sentido, essas teorias dão conta de um espaço onde a interpretação é realizada por humanos: são modelos complexos, mas que não foram construídos para serem processados automática-

mente. Assim, como produtos dos contextos onde estamos inseridos, o ensino também, até o advento das propostas da Web Semântica, vinha se conformando a estratégias teóricas e metodológicas da Web Sintática para dar conta de tais contextos.

3. Modelos de representação de domínios para no novo contexto da Web Semântica

Até o surgimento da Web Semântica, a separação entre documento e instrumento de tratamento estava bem definida. Tesouros e Taxonomias são exemplos de instrumentos que explicitam os conceitos de um domínio e permitem que se tenha um quadro de conceitos e relações a partir de um dado contexto de uso. Esses instrumentos são utilizados para a representação de conhecimentos de um dado domínio, porém, não possuem mecanismos para a derivação de conhecimentos, como é o caso das Ontologias, que se inserem no âmbito da Web Semântica. Nas ontologias, mais precisamente nas ontologias formais, o conhecimento pode ser obtido por meio de inferências sobre o conhecimento imediato disponível. Uma ontologia formal, ao mesmo tempo em que pode ser considerada uma ferramenta de representação, explicita e deriva também o conhecimento sobre um dado domínio (Campos, 2010).

Este novo cenário vai requerer, do profissional da informação, a aquisição e articulação de novos saberes, assim como novas estratégias para o ensino e a pesquisa na área.

A definição mais clássica de Web Semântica é a de Berners-Lee, Lassila e Hendler (2001, p.1), que a apresenta “como uma extensão da Web atual, na qual é dada à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação”. Dessa forma, a proposta da Web Semântica é incorporar ligações semanticamente mais ricas, capazes de dar suporte às

aplicações em negociações. Para isso, a Web Semântica requer que as pessoas façam um esforço extra na codificação de informações em representações passíveis de processamento automático. Com esse esforço, os computadores terão condições de melhor processar, interpretar e concatenar conceitos.

Por outro lado, apesar de os mecanismos de busca terem avançado significativamente, o que ocorre é que a seleção dos dados recuperados, na maioria das vezes, não sofre nenhuma seleção. Recuperamos a informação, mas, por vezes, sem nenhuma precisão. A Semântica que se propõe a esta nova geração da Web, a Web Semântica, pode auxiliar a conseguirmos recuperação com maior consistência. No entanto, é necessário que se tenha claro que a “Web Semântica não é uma Web separada, e sim uma extensão da Web (sintática) atual. Nessa nova Web, a informação terá significado bem definido através de linguagens de marcação semântica” (Breitman, 2005, p. 9). E ainda:

A Web semântica não é Inteligência Artificial. O conceito de documentos compreensíveis por máquinas não implica uma inteligência artificial mágica que faz com que os computadores passem a entender o que os seres humanos falam. Esse conceito apenas indica que computadores passarão a ter a habilidade de resolver problemas bem definidos através do processamento de operações que se utilizam de dados. Em vez de querer que computadores “entendam” a linguagem das pessoas, pedimos que as pessoas façam um esforço extra (codificando a informação em representações passíveis de processamento automático, ontologias, por exemplo). (Breitman, 2005, p. 9).

Para que isso venha a acontecer, será necessário utilizar formalismos. Dito de outra forma, que possa existir uma semântica formal

para os conceitos definidos em linguagem natural. Nesse sentido, as ontologias de fundamentação exercem um papel preponderante.

Segundo Chandrasekaran, Josephson e Benjamins (1999, p. 20), na área de Inteligência Artificial, uma ontologia:

é um vocabulário de representação, freqüentemente especializado em um domínio ou assunto. Mais precisamente não é o vocabulário, como tal, que se qualifica como uma ontologia, mas as conceitualizações que os termos do vocabulário destinam-se a capturar.

Segundo Guarino (1997, p. 16), conceitualização pode ser definida como “uma estrutura semântica intensional que codifica as regras implícitas, restringindo a estrutura de um pedaço da realidade”.

As Ontologias, diferentemente de outras ferramentas semânticas, apoiam-se em mecanismos de representação baseados em lógica, que permitem definir, por meio de axiomas, os conceitos nelas contidos e suas relações. As relações também possuem atributos que facilitam a descoberta de conhecimentos. Esses atributos indicam, por exemplo, se a relação é inversa (ex. *filho de e pai de*), ou simétrica (ex. *irmão de*), ou ainda transitiva (ex. *pré-requisito de*), entre outros. Porém, o uso de tais recursos depende da forma de modelagem dos conceitos adotada na ontologia.

Para que o seu potencial seja plenamente realizado, as ontologias na Web Semântica devem lançar mão de recursos formais (lógicos), de modo que permitam explicitar conceitos de diferentes graus de expressividade e dar conta de requisitos computacionais complexos (Ding et al., 2006). Nesse sentido, espera-se que seja possível expressar não só a identidade dos conceitos e seus atributos, mas também as suas relações com outros, que dependem do contexto e de seu compromisso ontológico. Algumas dessas relações expressam aspectos dinâmicos, como, por exemplo, o papel que uma pessoa

exerce e que pode mudar com o tempo, enquanto outras podem ser influenciadas pelo contexto social. Estas últimas colocam em destaque a interação das pessoas com os recursos descritos pelas ontologias, podendo, de acordo com o perfil e contexto do usuário, sugerir associações com outros recursos potencialmente de interesse (Mendes et al., 2010).

Na prática, a Ontologia Formal de Fundamentação, um tipo de ontologia, pode ser entendida como a teoria das distinções *a priori* sobre as entidades do mundo (objetos físicos, eventos, regiões, quantidades de matéria) e as categorias de metanível para modelar o mundo (conceitos, propriedades, qualidades, estados, papéis e partes) (Guarino, 1998).

Mais precisamente, as ontologias de fundamentação: têm um âmbito de aplicação extenso, podem ser reutilizadas para modelar diferentes situações, são bem fundamentadas do ponto de vista filosófico-conceitual, são axiomatizadas e, por isso, semanticamente transparentes. Elas se concentram em conceitos muito gerais e basilares, como os conceitos de objeto, evento, qualidade, papel, e em relações como constituição, participação, dependência, parte-todo. Tais conceitos e relações não são especificados para um domínio restrito, mas podem ser aperfeiçoados para atender a várias exigências aplicativas. A transparência semântica é oferecida por uma seleção cuidadosa dos termos primitivos e por uma caracterização precisa de seu significado.

Podemos observar que, no espaço de uma Web Semântica, as ferramentas utilizadas para o controle terminológico e recuperação de informação requerem a expansão de seus mecanismos de atuação, ou seja, se os vocabulários controlados e tesauros visavam ao controle terminológico que permitia a tradução da linguagem apresentada em documentos em uma linguagem artificial, a ontologia objetiva algo mais: ser um artefato tecnológico que contém um conjunto de

regras que delimitam o significado intensional⁵ de um vocabulário formal, permitindo que, a partir de um compromisso ontológico⁶, o conhecimento existente em um domínio possa ser representado e, assim, compartilhado e inferido.

Pode-se observar que, neste novo espaço, novas estratégias intelectuais serão requeridas para representar o conhecimento de um domínio e codificá-lo em uma linguagem que permita a leitura por máquinas. As representações devem ser feitas por meio de proposições lógicas, ou seja, os conteúdos das informações devem ser declarativos. Essas proposições são apresentadas na forma de axiomas, que permitam a realização de inferências com base nas proposições lógicas estabelecidas. Além da estrutura taxonômica, a qual é responsável pelo mapa estrutural do domínio coberto pela ontologia, os axiomas são uma chave essencial à construção de ontologias. O detalhamento dos axiomas é necessário para que as respostas oferecidas pelo sistema sejam confiáveis e baseadas no compromisso ontológico estabelecido pela própria ontologia (Medeiros, 2009).

Consideramos importante apontar que, para a elaboração de modelos de representação de domínios para o novo espaço da Web Semântica, o profissional da informação deverá se apropriar de conhecimentos que vão além das teorias terminológicas e das lógicas de classificação. São importantes as noções filosóficas e epistemológicas contidas em ontologias formais. O entendimento dessas noções e as notações adotadas para a sua modelagem são complexas, mas oferecem um campo vasto de aplicação, do qual o profissional da informação não pode se omitir.

⁵ O conceito de “intensional”, neste contexto, pode ser considerado como as características e atributos de um conceito que permitem determinar relações entre os conceitos.

⁶ Compromisso ontológico – acordo estabelecido sobre o que se compreende sobre um conceito, no interior de um domínio ou contexto de atividade.

4. Considerações Finais

A questão que se coloca é a necessidade de nos empenharmos na busca de estratégias de ensino e aprendizagem que considerem novas formas de representação de informação e de conhecimento, para que a máquina “entenda” o conteúdo a representado.

No evento CIFORM/UFBA de 2015 (Campos, 2015), foram discutidas algumas questões relacionadas à formação do profissional de informação para este novo ambiente. Em evento da comunidade ontologista, no domínio da Ciência da Computação, o “Ontology Summit” já está traçando um núcleo básico para a formação do ontologista do futuro ou mesmo, pode-se dizer, do “classificacionista do futuro”, ou seja, aquele que constrói modelos conceituais de domínios. Tal discussão é necessária no espaço das pesquisas e ensino em Ciência da Informação.

Esta formação envolve conteúdos como: 1. **Fundamentos Teóricos:** Lógica (primeira ordem, de segunda ordem, lógica descritiva, lógica de definições); Teoria dos conjuntos, Ontologia filosófica (universais e particulares, tempo); Filosofia da linguagem (a confusão entre sentido e referência, a teoria dos atos de fala,...); Representação do conhecimento, modelagem conceitual, modelagem de dados, metadados; 2. **Linguagens:** RDF, OWL,...; 3. **Construção e edição de ontologias:** aspectos humanos (aplicação de princípios de classificação), ferramentas de software (Protégé,...), problemas de interoperabilidade semântica; 4. **Estratégias e teorias de avaliação de ontologias;** 5. **Teorias e Metodologias de elaboração de ontologias:** ontologias de fundamentação, ontologias de domínio. Além disso, registram-se também as disciplinas afins, a saber: **Lógica avançada** (lógica modal, lógica temporal, lógica padrão,...); **Ontologia Filosófica Avançada** (mereotopology, tropos...); **Ciência da Computação:** linguagens formais, máquinas formais, computabilidade, raciocínio automatizado, teoria de banco de dados; **Linguística / Ciências**

Cognitivas: distinção entre sintaxe, semântica, pragmática, processamento de linguagem natural, geração de linguagem natural, teorias cognitivas de categorização.

Assim, no âmbito deste capítulo, nossa intenção foi lançar algumas discussões sobre os conteúdos até hoje ministrados e aqueles que deverão ser revistos e acrescentados, visando à formação de profissionais de informação aptos a atuar efetivamente no desenvolvimento de modelos conceituais de domínios para uso na Web Semântica.

Referências Bibliográficas

- Aristóteles. (2010). *Órganon*. São Paulo: Edipro.
- Barros, L. A. (2004). *Curso básico de terminologia*. São Paulo: EdUSP.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43. Recuperado de <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. [Acessado em 05/11/18]
- Bliss, H. E. (1929). *The organization of knowledge and the system of the sciences*. New York, NY: Holt and Company.
- Breitman, K. (2005). *Web Semântica: a internet do futuro*. Rio de Janeiro: LTC.
- Campos, M. L. A. (2010). O documento e as ferramentas de tratamento e recuperação de informações na Web Semântica: um novo espaço de identidade. In L. S. Freitas, C. H. Marcondes, & A. C. Rodrigues (Orgs.), *Documento: Gênese e contextos de uso – Estudos da informação* (Vol. 1, pp. 223-234). Niterói: EdUFF.
- Campos, M. L. A. (2012). Ontologia de Fundamentação como uma teoria de representação para a Web semântica: estratégias interdisciplinares na pesquisa e no ensino. In E. I. Murguia & M. E. F. Rodrigues (Orgs.), *Arquivologia, Biblioteconomia e Ciência da Informação: identidades, contrastes e perspectivas de interlocução* (Vol. 2, pp. 101-116). Niterói: EdUFF.
- Campos, M. L. A. (2015). Representação de Domínios na Web Semântica: Desafios para a formação de profissionais de informação. In *Anais do XII CINFORM – Encontro Nacional de Ensino e Pesquisa em Informação* (pp. 73-93). Salvador, BA, Brasil: ICI/ EdUFBA. Recuperado de <https://repositorio.ufba.br/ri/bitstream/ri/18051/1/ANAIS%20DO%20XII%20CINFORM%20ANO%202015.pdf>
- Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, 14(1), 20-26. doi:10.1109/5254.747902
- Dahlberg, I. (1978a). A referent-oriented analytical concept theory of interconcept. *International Classification*, 5(3), 142-150. doi:10.5771/0943-7444-1978-3-142

- Dahlberg, I. (1978b). Teoria do conceito. *Ciência da Informação*, 7(2), 101-107. Retirado de <http://revista.ibict.br/ciinf/article/view/115/115>. [Acessado em 05/11/18].
- Ding, L., Kolari, P., Ding, Z., & Avancha, S. (2007). Using ontologies on the semantic web: a survey. In R. Sharman, R. Kishore, & R. Ramesh (Eds.), *Ontologies: A Handbook of Principles, Concepts, and Applications in Information Systems* (pp. 79-113). Boston, MA: Springer.
- Dutta, B., & Dutta, C. (2014). A linguistic view of subject formation process as described by Ranganathan and Others. *Annals of Library and Information Studies*, 61(1), 56-64. Recuperado de <http://hdl.handle.net/10760/23304> [Acessado em 05/11/18]
- Feibleman, J. K. (1954). Theory of Integrative Levels. *British Journal for the Philosophy of Science*, 5(17), 59-66. doi:10.1093/bjps/V.17.59
- Green, R. (2008). Relationships in Knowledge Organization. *Knowledge Organization*, 35(2/3), 150-159. doi:10.5771/0943-7444-2008-2-3-150
- Guarino, N. (1997). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In Pazienza, M. T. (Ed.), *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. Proceedings of SCIE '97 International Summer School on Information Extraction, Frascati, Italy* (pp. 139-170). Heidelberg: Springer. Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.7515>. [Acessado em 05/11/18]
- Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), *Formal Ontology in Information Systems: Proceedings of FOIS'98, Trento, Italy* (pp. 3-15). Amsterdam: IOS Press. Recuperado de <https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>
- Lima, G. A. (2018). Aspectos cognitivos na categorização: buscando o contexto. In G. A. Lima (Org.), *Bibliotecas digitais: novas tendências na navegação em contexto* (Vol. 1, pp. 1-12). Rio de Janeiro: Interciência.
- Medeiros, J. S. (2011). *Tesouros conceituais e ontologias de fundamentação: análise comparativa entre as bases teórico-metodológicas utilizadas em seus modelos de representação de domínios* (Dissertação de mestrado em Ciência da Informação, Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, Niterói, RJ, Brasil). Recuperado de http://www.ci.uff.br/ppgci/arquivos/Dissert/Diss_JacksonMedeiros.pdf
- Mendes, P. N., Kapanipathi P., Cameron, D. H., & Sheth, A. P. (2010). *Dynamic associative relationships on the linked open data web*. Paper presented at the Web Sci10: Extending the Frontiers of Society On-Line, Raleigh, NC, USA. Retirado de <https://corescholar.libraries.wright.edu/knoesis/760>
- Morujão, C. (2006). A lógica modernorum: lógica e filosofia da linguagem na escolástica dos séculos XIII e XIV. *Revista Filosofia de Coimbra*, 14(28), 301-322. Recuperado de http://www.uc.pt/fluc/dfci/publicacoes/a_logica_modernorum. [Acessado em 05/11/18].
- Neelameghan, A. (1971). Sequence of Component Ideas in a Subject. *Library Science with a Slant to Documentation*, 8, 323-334.
- Neelameghan, A. (1973). Basic Subject. *Library Science with a Slant to Documentation*, 10. Papers F to N.

- Piedade, M. A. (1983). *Introdução à teoria da classificação* (2ª ed.). Rio de Janeiro: Interciência.
- Raghavan, K. S. (1985). The general theory of classification as the basis for structuring of subject headings. In *Proceedings of the Regional Conference of International Federation for Documentation Committee On Classification Research (FID-CR), New Delhi, India* (pp. 24-48). New Delhi: Delhi Library Associations.
- Ranganathan, S. R. (1967). *Prolegomena to library classification* (3rd ed.). London: Asia Publishing House.
- Sen, B. K. (2009). Universe of knowledge from a new angle. *Annals of Library and Information Studies*, 56(1), 7-12.
- Vickery, B. C. (1997). Ontologies. *Journal of Information Science*, 23(4), 227-286. doi:10.1177/016555159702300402
- Wilson, T. D. (1972). The Work of the British Classification Research Group. In H. Wellisch, & T. D. Wilson (Eds.), *Subject retrieval in the seventies* (pp. 62-71). Westport, CT: Greenwood Publishing Co.

CAPÍTULO 5
A VARIAÇÃO TERMINOLÓGICA
EM ORGANIZAÇÃO DA INFORMAÇÃO: ANÁLISE
DE ASSUNTO E ANÁLISE DOCUMENTÁRIA
TERMINOLOGICAL VARIATION IN
ORGANIZATION OF INFORMATION: SUBJECT
ANALYSIS AND DOCUMENTARY ANALYSIS

Cristina Dotta Ortega

Universidade Federal de Minas Gerais (UFMG)

ortega@eci.ufmg.br

ORCID: 0000-0002-9735-7676

Mariângela Spotti Lopes Fujita

Universidade Estadual Paulista (UNESP)

mariangela.fujita@unesp.br

ORCID: 0000-0002-8239-7114

Maria da Graça Simões

Universidade de Coimbra (UC)

Faculdade de Letras/CEIS20

gsimoes@fl.uc.pt

ORCID: 0000-0001-5575-0261

Resumo: Uma área do conhecimento fundamenta-se por sistema conceitual consistente e terminologia específica com relativo nível de consenso e estabilidade. Os termos Análise Documentária (*Analyse Documentaire*) e Análise de Assunto (*Subject Analysis*), enquanto denominações distintas para a etapa de análise do conteúdo temático de documentos a ser representado em sistemas, são usados ora com o mesmo significado, ora reconhecendo-se suas origens históricas e culturais próprias. Objetiva-se analisar os traços semelhantes e distintos entre as abordagens que envolvem os termos citados. Para tanto, analisa-se o uso dos termos em textos científicos representativos de sua conceituação e caracterização (décadas de 1960-1990), com o recurso de segmentação e registro dos conceitos identificados para posterior sistematização, comparação e discussão, segundo as seguintes categorias: 1) definições; 2) objetivos; 3) objetos; e 4) princípios teóricos e metodológicos, incluindo aportes de outras áreas. Em seguida, analisa-se o uso destes termos, segundo as mesmas categorias, nos textos apresentados nas reuniões da ISKO Internacional e nos capítulos ISKO Brasil e ISKO Espanha e Portugal dos últimos cinco anos (2013-2018). Discute-se alguns problemas e avanços de cada abordagem quanto à sua coerência explicativa e aplicabilidade, ressaltando, ao final, a improdutividade da prática recorrente de pesquisa, realizada na ausência de reconhecimento da ampla e diversa produção da área.

Palavras-Chave: Análise documentária. Análise de assunto. Organização da informação. Questões terminológicas em Ciência da Informação.

Abstract: Any area of knowledge should be based on a consistent conceptual system and a specific terminology with a certain level of consensus and stability. The Portuguese translations of the French term “Analyse Documentaire” and the English term “Subject Analysis”, as different denominations for the stage of subject analysis of the documents to be represented in systems, are used with the same meaning although recognizing their historical and cultural origins. The objective of this paper is to analyze the similarities and differences between the approaches that involve these terms. We analyzed the use of these terms in scientific texts that are representative of their

conceptualization and characterization (for the period 1960-1990), grouping and recording the concepts for later systematization, comparison, and discussion, according to the following categories: 1) definitions; 2) objectives; 3) objects; and 4) theoretical and methodological principles, including contributions from other areas. Next, we analyzed the use of these terms, according to the same categories, in the papers presented at ISKO International, ISKO Brazil, ISKO Spain and Portugal in the last five years (2013-2018). We discuss some problems and advances of each approach in terms of their explanatory coherence and applicability, highlighting, in the end, the unproductiveness of the common research practice conducted in the absence of the recognition of the wide and diverse production of the area.

Keywords: Documentary analysis. Subject analysis. Knowledge organization. Terminological issues in Information Science.

1. Introdução

Uma área do conhecimento ganha *status* científico após longo percurso de pesquisas e debates que conduz à construção de conceitos, modelos e teorias e de terminologia com relativo nível de consenso e estabilidade. A terminologia de uma área responde por um conjunto de termos com conceituações precisas que se relacionam entre si, de forma a compor o desenho da área em seus núcleos centrais e circundantes, permitindo assim a comunicação entre especialistas e a consolidação de sua identidade.

Ao tratar da terminologia adotada em Ciência da Informação, Kobashi, Smit e Tálamo (2001, p. 6) afirmam que, como os conceitos atribuídos aos termos devem resultar de relações entre suas características constitutivas, passíveis de serem objetivadas e confirmadas, o reconhecimento de uma denominação e de seu conceito

é tarefa que exige análise da pertinência dessas características em relação ao domínio considerado. Observam que o uso da palavra em estado natural, em geral decorrente de convenções arbitrárias ou de preferências individuais, é sempre um risco.

Os termos Análise Documentária e Análise de Assunto (originados, respectivamente, do francês *Analyse Documentaire*, e do inglês *Subject Analysis*) são denominações adotadas para a etapa de análise do conteúdo temático de documentos a ser representado em sistemas, visando sua recuperação segundo o interesse de usuários. A despeito desta definição, os termos são usados em geral, de modo mais abrangente, referindo-se ao processo de análise e de representação temática, assim como aos instrumentos adotados para tal.

Na literatura de especialidade, os termos Análise Documentária e Análise de Assunto ora representam o mesmo conceito, sobrepondo-se conceptualmente, ora representam realidades conceptuais distintas, sendo comum, ainda no mesmo texto, encontrarem-se os dois termos indistintamente para traduzirem o mesmo conceito. A maior parte destes usos demonstra a ausência de qualquer princípio epistemológico. Assim, parte-se do pressuposto que, subjacente a esta variação terminológica, encontram-se construções culturais, sociais e técnicas historicamente construídas que resultaram em abordagens próprias sobre o mesmo processo.

Com o propósito de contribuir para a estabilidade terminológica na área da Ciência da Informação, em particular no domínio da Organização da Informação, este trabalho busca realizar uma investigação sobre o uso destes termos. Nesta perspectiva, constituiu-se como objetivo analisar similitudes e diferenças entre as abordagens que envolvem os termos Análise Documentária e Análise de Assunto. O cumprimento do propósito enunciado dá-se recorrendo a uma metodologia qualitativa, na qual se privilegiam a revisão de literatura e a análise comparativa. O trabalho se justifica pela percepção

de que, respeitadas as condições históricas de desenvolvimento de cada uma das abordagens indicadas, a construção científica de uma área deve contemplar estudos comparativos e contínua revisão crítica.

Os textos selecionados inicialmente são aqueles considerados emblemáticos para a compreensão destes termos, percorrendo, assim, as décadas de 1960 a 1990. Em seguida, analisamos a variação terminológica em questão nos textos apresentados nas reuniões da ISKO Internacional e nos capítulos ISKO Brasil e ISKO Espanha e Portugal dos últimos seis anos (2013-2018), observando o significado destes usos, passados pouco mais de dez anos da produção fundamental sobre os termos, considerada anteriormente. Os dois grupos de textos foram considerados a partir das seguintes categorias de análise: 1) definições; 2) objetivos; 3) objetos; e 4) princípios teóricos e metodológicos, incluindo aportes de outras áreas.

Como produto da análise, tratamos das possíveis origens e de alguns problemas e avanços de cada abordagem quanto à sua coerência explicativa e aplicabilidade.

2. Literatura fundacional sobre os termos Análise Documentária e Análise de Assunto (décadas de 1960-1990)

2.1. Análise Documentária

Para analisar o uso do termo Análise Documentária, optou-se por adotar centralmente a produção de literatura realizada pelo Grupo Temma, iniciado no Departamento de Biblioteconomia e Documentação, da Escola de Comunicações e Artes, da Universidade de São Paulo (ECA/USP), na década de 1980, e extinto formalmente em 2017. Com a instauração deste Grupo, criou-se na ECA a linha de pesquisa denominada Análise Documentária.

Considerando a extensão das pesquisas e experimentos sobre Análise Documentária realizados pelo Grupo Temma, julgou-se que correspondem adequadamente às pretensões deste trabalho. Trata-se de sistematizações reflexivas acerca do tema, como as críticas e revisões realizadas sobre as obras dos autores Jean-Claude Gardin e Antonio García Gutiérrez, que foram as referências do Grupo. É na obra de Gardin e seus contemporâneos que se encontra a origem do termo Análise Documentária, como se pode constatar no artigo “Analyse documentaire et analyse structurale en archéologie” (Gardin, 1974a), publicado originalmente em 1965, e traduzido e publicado para a língua portuguesa poucos anos depois (Gardin, 1968).

Buscando explorar o termo Análise Documentária, segue revisão bibliográfica da produção de participantes do Grupo Temma, com foco nos dois autores mencionados acima, entre outros.

Cunha, Kobashi e Amaro (1987, pp. 127-128) realizaram uma revisão bibliográfica sobre textos de Linguística Geral, Lógica e Filosofia da Linguagem e de Linguística/Documentação, avaliando o nível de contribuição que eles poderiam realizar à área. Neste sentido, citam a obra de Gardin (1974b), que propõe que a Análise Documentária, dado o conjunto de necessidades crescentes de informatização, proceda a uma formalização dos seus procedimentos. Neste contexto, as referidas autoras entendem que o estudo de Gardin interessa à Biblioteconomia/ Documentação por propor métodos e procedimentos para a análise de texto/discurso que possibilitam análises “objetivas” e “universais” e por realizar a sistematização de métodos e procedimentos lógico-linguísticos de forma a permitir a operacionalização da recuperação e disseminação da informação.

Outro trabalho deste autor, considerado pelas autoras (Gardin, Lagrange, Martin & Natali-Smit, 1981), explora a Análise Documentária, não apenas como a representação do texto, mas como uma reconstrução do percurso epistemológico do discurso, que visa tanto

a identificação do corpus em análise e de sua validação, como a possibilidade de simular construções análogas. Ela é voltada à explicitação e/ou proposição de modelos metodológicos de funcionamento dos processos de identificação, descrição e representação de conteúdos de textos.

Por sua vez, Cunha (1989b, p. 17) apresenta a definição de Gardin, Lagrange, Martin e Natali-Smit (1981, p. 29), na qual a Análise Documentária seria a expressão usada para designar o conjunto de procedimentos efetuados com a finalidade de expressar o conteúdo dos documentos científicos sob formas destinadas a facilitar a recuperação da informação.

Lara (1999, p. 54) afirma que Gardin (1970, 1973a, 1973b) identifica a atividade de representação documentária como um gênero de representação probabilística para a expressão do conteúdo de textos científicos: significados do texto são retirados e substituídos por outros que não são, necessariamente, encontrados no texto, caracterizando operações de construção de metalinguagens, com o objetivo de permitir buscas retrospectivas de informação.

Kobashi (1994, pp. 33-34), ao apresentar a Análise Documentária, trata da articulação da Documentação com a Linguística, fato que levou Gardin a estudos que ultrapassassem os problemas colocados pelas metodologias baseadas em modelos meramente estatísticos. A autora atribui a Gardin a criação das bases para a reflexão teórica e a terminologia adotadas atualmente pela Documentação de orientação francesa e de certas vertentes anglo-saxônicas. Para Gardin, o produto da Análise Documentária é submetido a métodos de avaliação empírica, quanto à sua efetiva capacidade de funcionar como filtro de identificação e localização de documentos, os quais se distinguem de outras modalidades de análise de textos. Outro aspecto de interesse, para Gardin, são os sistemas de representação utilizados via linguagens documentárias. Para ele (Gardin, 1987, p. 49), o produto da análise é uma representação do texto formu-

lada em uma linguagem que não se confunde com a linguagem do texto, mesmo que os termos tenham aparentemente a mesma forma.

Segundo Kobashi (1994, pp. 32-33), uma das contribuições mais notáveis de Gardin dá-se precisamente na proposição de um modelo formal de linguagem documentária que dê conta de sua organização, tanto do ponto de vista do léxico e das estruturas semânticas (eixo paradigmático) quanto de uma sintaxe que articule os encadeamentos entre os termos da linguagem (eixo sintagmático). As reflexões do Grupo Temma sobre esta contribuição, por sua vez, indicam que as relações sintagmáticas não compõem propriamente a linguagem documentária, sendo adotadas no momento de seu uso no sistema, seja no processo de indexação, seja no de busca.

Kobashi (1989, pp. 48-49) cita que Gardin (1974b) considera a Análise Documentária como um tipo de análise semântica que se aproxima da análise de conteúdo do ponto de vista do objeto: textos a serem analisados, devendo-se chegar ainda à sua representação pela mediação de um sistema simbólico, ou seja, uma metalinguagem. Ele destaca dois momentos na Análise Documentária: a interpretação e a descrição. A interpretação é realizada pela identificação da organização lógico-semântica do discurso, enquanto a descrição utiliza uma metalinguagem. Verifica-se aproximação à Semiótica pela preocupação com a formulação de simbolização da linguagem.

Neste contexto dos estudos franceses sobre Análise Documentária, sua relação com os estudos de Linguística é demonstrada pelo verbete ‘documentário’, do Dicionário de Linguística, de Dubois et al. (1998, p. 203): “chama-se *análise documentária* a representação, por meio de termos e de processos sintáticos convencionais, de um certo conteúdo dos documentos (artigos, publicações) científicos com fim de classificação, de pesquisa de informação”.

Pautadas pelas referências francesas, e para além delas, as pesquisadoras do Grupo Temma desenvolveram a linha de pesquisa Análise Documentária, como segue.

Segundo Kobashi (1994, p. 15):

a Análise Documentária é uma disciplina de natureza teórica e prática que integra o domínio conhecido por Biblioteconomia e Documentação, mais modernamente denominado Ciência da Informação. Enquanto parte constitutiva de uma área de conhecimentos, a Análise Documentária assimila, necessariamente, os objetivos globais do tronco do qual é membro, qual seja o de determinar e criar princípios e mecanismos que promovam a circulação de informações e de documentos.

Kobashi afirma que o processo global da Análise Documentária defronta-se com dois objetos: o texto, como a sua unidade de análise; e a linguagem documentária como instrumento comutador. Ao operar com textos, a Análise Documentária lida, portanto, com a linguagem. Esse fato impôs a discussão dos problemas de representação e de significação com os quais se depara a Análise Documentária, desafio que foi enfrentado no trabalho pela apropriação de certos conceitos da Semiologia e da Filosofia da Linguagem (Kobashi, 1994, p. 45, 79).

Para Kobashi (1994, p. 22), a Análise Documentária é disciplina de natureza metodológica que, para avançar teórica e praticamente, deve criticar continuamente seus pressupostos, procedimentos e instrumentos; deve, ao mesmo tempo, com base na reflexão permanente, elaborar novas hipóteses de trabalho que contribuam para aperfeiçoar os processos que lhe dizem respeito.

Para Cunha (1989a, p. 122), a Análise Documentária não é apenas um processo pelo qual se sintetiza o conteúdo de um documento após a leitura e a análise, através de palavras-chave, descritores, termos ou paráfrases, mas um processo mais completo que envolve o próprio processo de leitura, análise de texto e mecanismos mentais e psicológicos de atribuição de conceitos que possam simbolizar

o conjunto do texto. Cunha considera o processo mental do indexador, além das próprias características do texto, aspecto pouco explorado, mas tratado especialmente por Anna Maria Cintra (1987) em seu texto sobre leitura documentária.

Para Tálamo, Lara, Kobashi e Amaro (1992, pp. 1-2), a Análise Documentária compreende uma série de procedimentos metodológicos, cujo objetivo é o de viabilizar a análise de textos técnico-científicos para seleção de conteúdos informativos, que serão representados, recuperados e disseminados. Quanto ao objetivo, trata-se de propor princípios para a elaboração de informações documentárias, definidas, por sua vez, como múltiplas representações de textos para fins de recuperação face às diferentes demandas de informação. Este fato implica a proposição de procedimentos que sejam simultaneamente flexíveis e rigorosos.

As autoras afirmam ainda que a Análise Documentária integra o campo da Ciência da Informação mas, de outro modo, insere-se no processo geral de Comunicação, na medida em que elabora produtos que promovem a socialização de bens culturais. Por ser regulada por parâmetros institucionais, as atividades pragmáticas da Análise Documentária encontram-se sobremodalizadas por variáveis situacionais, decorrentes de sua inserção do processo global de produção e circulação de bens simbólicos. Assim, os procedimentos de Análise Documentária, embora objetiváveis, não são neutros (Tálamo, Lara, Kobashi & Amaro, 1992, p. 2).

Em perspectiva que foi motivadora das pesquisas iniciais do Grupo Temma, Cunha (1987, p. 38) afirma que, até então, os métodos da análise de textos foram considerados operações empíricas de “bom senso”. Do mesmo modo que Cunha, Kobashi (1989, p. 45) fala que a forma de realizar a Análise Documentária é tradicionalmente vista como dependente da experiência, intuição e familiaridade com o assunto. Sob este ponto de vista, os passos executados nesse processo são operações estritamente técnicas, fundadas nos mitos

da objetividade e da neutralidade e na crença da possibilidade de representação fiel de conteúdos de textos. As “receitas” apresentadas nos manuais de Biblioteconomia e Documentação, tanto nacionais quanto internacionais, reduzem o processo documentário às etapas de determinação do assunto e tradução dos conceitos selecionados para uma linguagem documentária, sem oferecer, porém, uma explicitação muito clara dos critérios a serem seguidos na execução desses passos. Posteriormente, Kobashi (1994), afirma que, apoiada fortemente no conhecimento empírico, na intuição e no hábito, a área, por um longo período, rejeitou a teoria, considerando-a supérflua. Essa mesma visão refletiu-se no ensino dessas técnicas, cujo aprendizado dependia mais da capacidade de memorização do que da compreensão, já que ocorria de forma descontextualizada, sem que se revelassem os princípios que sustentam as operações praticadas.

Kobashi questiona a consistência das regras de indexação e de elaboração de resumos, em função de dois aspectos: propõem operações de análise e de seleção de informações que apelam de forma dominante para os conhecimentos intuitivos; e, são formuladas dentro de padrões terminológicos ambíguos. Via de regra, as metodologias disponíveis não definem, de fato, a matéria-prima com a qual trabalham, estabelecendo, por exemplo, equivalências entre os termos “discurso”, “texto” e “documento”. Semelhantes equivalências costumam ser, ainda, combinadas com noções intuitivas como “informação principal”, “conceitos mais importantes” ou mesmo “informação nova” dos textos.

A consideração da autora é a de que a precariedade das regras compromete a qualidade das informações documentárias e a sua recuperação, sem que se possa estabelecer parâmetros confiáveis de avaliação. Pode-se, nessas circunstâncias, avaliar apenas a pobreza ou riqueza da recuperação, mas não o porquê ou o grau dos resultados atingidos. Uma metodologia possibilitaria que a avaliação desses dois aspectos corrigisse esses procedimentos e superasse deficiências.

As reflexões realizadas pelo Grupo Temma demonstraram a necessidade de estudos e da proposição de modelos fundamentados, passíveis de validação.

Em sua tese de doutorado, Kobashi (1994) propõe metodologia de elaboração de informações documentárias (em especial, de indexação e de elaboração de resumos). A proposta realizada na tese decorre, entre outros, da avaliação de modelos propostos e testados anteriormente no contexto do Grupo Temma, como aqueles em que a Análise de Discurso foi adotada. Para ela (Kobashi, 1994, pp. 64-65), as diversas tentativas de incorporar seus procedimentos na Análise Documentária não foram bem-sucedidas. Apresenta duas razões para tal: em Análise Documentária, não há preocupação com as condições socio-históricas da produção discursiva, pois seus objetivos restringem-se à identificação da base temática do texto; e a noção de informação relevante de um texto varia de acordo com os objetivos que se persegue, de tal modo que os mecanismos utilizados no âmbito da Análise do Discurso não podem ser transferidos mecanicamente para a Análise Documentária.

Deste modo, Kobashi contextualiza seu tema de tese, tratando das seguintes operações básicas: coleta de documentos, tratamento de documentos e difusão dos documentos. Para a autora, o fluxo citado supõe diversas atividades coordenadas no espaço e no tempo, necessariamente contextualizadas no âmbito de instituições informacionais (como bibliotecas, centros de documentação, etc.). A análise rigorosa do contexto no qual se insere o sistema precede as operações documentárias propriamente ditas e são dadas por uma política global de tratamento e recuperação de informações, pautada pelas seguintes categorias: necessidade do usuário, domínio tratado, recursos disponíveis, produtos e serviços e relação custo/desempenho (Kobashi, 1994, p. 15, 17 e 18).

Kobashi (1994, p. 21) apresenta o fluxo documentário, por meio de duas operações básicas e complementares: a fabricação e a

recuperação da informação, ambas mediadas pela representação. Na fabricação de informações documentárias, os processos levam à transformação de um objeto (documento) em um ou outro objeto (informação documentária), por meio de operações de análise e de síntese. A função da Análise Documentária é, então, a de permitir selecionar, de um universo de objetos, aqueles que poderão responder a uma necessidade de informação. A recuperação da informação é a identificação das informações pertinentes, correlacionando a questão proposta pelo usuário e as informações documentárias armazenadas. Portanto, entre a produção de informação documentária e a recuperação existe uma relação de pressuposição não recíproca: a recuperação não é factível sem uma ação anterior, denominada Análise Documentária (Kobashi, 1994, p. 22).

Nas etapas de análise e de síntese, os textos são desestruturados, quando ocorre a distinção entre informação essencial e acessória. Em seguida, as informações selecionadas são estruturadas, isto é, são submetidas a um processo de combinação, de modo a convertê-las em novos textos, ditos resumos, ou em símbolos de uma linguagem documentária (no caso da indexação). Assim, há a etapa de análise e síntese como operações de desestruturação e a etapa de representação como aquela em que se realiza a estruturação (Kobashi, 1994, p. 24).

Kobashi (1994, p. 99) afirma que as propostas que procuram construir modelos sobre o processo de indexação podem, grosso modo, ser classificadas em dois grupos: de um lado, aquelas que se propõem a sistematizar o bom senso, explicitando as regras utilizadas inconscientemente pelo indexador, como Beghtol (1986) e, de outro, aquelas que se propõem a elaborar regras explícitas, considerando a especificidade da atividade, como Frohmann (1990) e Cooper (1978, 1979). Embora entenda que os estudos do primeiro grupo sejam válidos em algumas situações, a autora diz que é preciso privilegiar pesquisas voltadas à construção de princípios que

regem o processo, propondo métodos rigorosos para sua realização. Informa, desse modo, que pretende seguir o segundo grupo e discorre sobre a questão, como segue.

Para Kobashi, Beghtol (1986) descreve a classificação bibliográfica e a elaboração de resumos como operações de determinação do *aboutness* (do que fala) um texto. Assim, a seleção de informações para fins de indexação seria o resultado do uso de regras mentais que guiam automaticamente os mecanismos de compreensão do texto.

Frohmann (1990), por outro lado, considera que a inconsistência das representações documentárias não poderá ser resolvida pela descoberta de regras implícitas seguidas pelos indexadores. Trata-se, muito mais, de trocá-las por regras suficientemente precisas que sirvam simultaneamente para definir padrões e treinar as atividades documentárias. Nessa medida, caberia à teoria da indexação enfrentar o desafio, propondo regras explícitas e bem formuladas que possam ser efetivamente utilizadas para analisar documentos e transformá-los em representações (Kobashi, 1994, pp. 99-100).

Cooper (1979, p. 107), anteriormente, afirmou que alguns estudos procuram investigar como o indexador profissional indexa, mais do que como ele poderia indexar. Ele conclui não haver ainda respostas consensuais entre especialistas sobre o que indexador deve fazer ou como o seu desempenho poderia ser avaliado (Kobashi, 1994, p. 100).

De certo modo, o modelo de Van Dijk e Kintsch (1983) também relativiza o papel das regras inconscientes, ao constatar que as propriedades estruturais do texto são relevantes para a sua compreensão e, em decorrência, para a elaboração de condensações. Afirmam que certos tipos de discursos exibem estruturas convencionais que, enquanto padrão organizacional adicional, podem ser relevantes para o discurso como um todo, e os chamam de superestruturas (Kobashi, 1994, pp. 100-101).

Com isto, Kobashi apresenta uma metodologia de indexação e uma metodologia de elaboração de resumos que, em linhas gerais, parte do princípio de que a seleção de dados deve ser antecedida pela operação de identificação e hierarquização de informações textuais, sendo necessário um paradigma específico de coleta de dados, para cada tipo de produto documentário. Os paradigmas foram elaborados a partir da apropriação dos conceitos de tema, textualidade, superestrutura textual e estrutura temática (Kobashi, 1994, p. 175).

Para contemplar adequadamente as metodologias propostas, seria necessário discorrer mais profundamente sobre superestruturas textuais e estruturas temáticas no contexto das diversas tipologias textuais, como desenvolvido pela autora. No entanto, apresentamos aqui apenas alguns elementos destes conceitos. A superestrutura textual, segundo Van Dijk (1992), é considerada um elemento fundamental para a compreensão dos textos porque configura-se como um esquema abstrato que estabelece a ordem global de um texto, compondo-se de uma série de categorias, cujas possibilidades de combinação se baseiam em regras convencionais, além de ser sempre conhecida de uma dada comunidade linguística (Kobashi, 1994, p. 122). A estrutura temática, por sua vez, constituiu-se por cinco categorias, quais sejam: o quê, quem, quando, onde, como. A categoria “quem” é essencial no texto narrativo e na notícia jornalística, mas não foi adotada na operação de identificação do tema por Kobashi por não ser pertinente a textos de natureza técnico-científica, os quais constituíram o corpus central de análise da pesquisa de Kobashi (1994, p. 121). Kobashi (1994, pp. 111-112) opera com a compreensão de Tálamo (1987, pp. 3-4) de que o tema central de um texto é, na realidade, o seu objetivo principal. O tema é definido como o elemento central em torno do qual se estrutura a mensagem, constituindo o núcleo informativo do texto.

Kobashi apresenta um esquema da metodologia, por meio da segmentação de cada uma delas em princípios e operações. O princípio refere-se à natureza do processo, enquanto as operações referem-se ao conjunto de procedimentos realizados para obter um produto final (Kobashi, 1994, pp. 119-124):

Figura 1: Esquema das metodologias, por meio da segmentação de cada uma delas em princípios e operações.

Metodologia	Princípio	Operações		
		ação	descrição	parâmetro
De Indexação	Caracterização do conteúdo de um documento por meio de uma linguagem documentária	1. Identificação do tema	Leitura para identificação do objetivo do texto	Categorias da estrutura temática (categoria essencial – o quê – e categorias acessórias – quando, onde, como)
		2. Elaboração do enunciado temático	Seleção e combinação das categorias temáticas pertinentes	Política do sistema de indexação
		3. Representação do enunciado temático	Conversão do enunciado temático para a linguagem do sistema	Linguagem de indexação e política do sistema
De Elaboração de Resumos	Produção de representações condensadas de um texto	1. Identificação do tema	Leitura para identificação do objetivo do texto	Categorias da estrutura temática (categoria essencial – o quê – e categorias acessórias – quando, onde, como)
		2. Identificação das informações do texto	Leitura do texto para identificar e hierarquizar informação	Superestruturas textuais
		3. Seleção das informações mais importantes	Determinação das informações textuais pertinentes	Tipo de resumos e categoriais textuais pertinentes
		4. Representação da estrutura informacional	Redação do resumo	Regras formais de elaboração de resumos

Fonte: Kobashi, 1994, p. 120.

Em toda a tese, em especial na explicitação da metodologia proposta, há menção ao termo “tema”, mas não ao termo “assunto”, salvo nas citações a outros textos, cujos autores adotam o termo, e em equivalência à ideia de tópico e à ideia de produto da análise do texto. Para Kobashi (1994, p. 24), uma cadeia de termos de indexação é uma representação de nível inferior ao do texto, em que subsiste apenas o tópico (assunto) do texto, representado por meio de uma palavra, um sintagma ou uma unidade lexical mais complexa. Neste sentido, tema é adotado em contraponto a assunto, o qual é foco da abordagem constituída sob o termo Análise de Assunto, como tratamos à frente.

Em trabalho de conclusão de curso de graduação do curso de Biblioteconomia, da ECA/USP, sob orientação de Regina Keiko Obata Amaro, Virgínia Castilho (1995) realiza revisão dos métodos de análise de textos em sua relação com a Análise Documentária, visando subsidiar uma discussão sobre o uso destes métodos para a indexação automática. O trabalho explora textos do segundo autor citado como referência do Grupo Temma – García Gutiérrez –, e textos das pesquisadoras deste Grupo.

Neste sentido, cita García Gutiérrez (1992, p. 21) que define a Análise Documentária como um conjunto de procedimentos utilizados tanto pelo documentalista na captação das proposições do texto, sua segmentação e representação formal, como pelo usuário no processo interrogador e interativo com a memória documental. Segundo Castilho, esta definição enfatiza o papel do usuário como elemento da Análise Documentária, mas não se distancia daquela proposta por Gardin, pois ambas se referem à recuperação da informação como objetivo pragmático da Análise Documentária, diferenciando-se de outros tipos de análise semântica.

Castilho (1995, pp. 8-9) menciona novamente García Gutiérrez (1992, p. 31), para quem a Análise Documentária que tem por objetivo a recuperação do documento mediante a representação de

traços descritivos de seu conteúdo por meio de termos, sintagmas ou outras unidades discretas, denomina-se indexação.

À semelhança de Cunha e de Kobashi, García Gutiérrez constata a falta de um método científico na Análise Documentária que conduza a procedimentos específicos. Este autor observa que se encontra em geral, tanto no ensino como na aplicação, apenas uma normalização e, mesmo esta, enfatiza a codificação ou a representação, não abrangendo o processo como um todo (García Gutiérrez, 1992, p. 24).

Castilho realiza uma sistematização sobre as etapas da Análise Documentária, discutindo algumas propostas e fazendo contraponto entre elas. Neste contexto, trata da proposta de García Gutiérrez (1992, pp. 26-28), denominada por ele de paradigma do processo documentário (P/PD), que afirma ser baseado nas operações que afetam o conteúdo do documento:

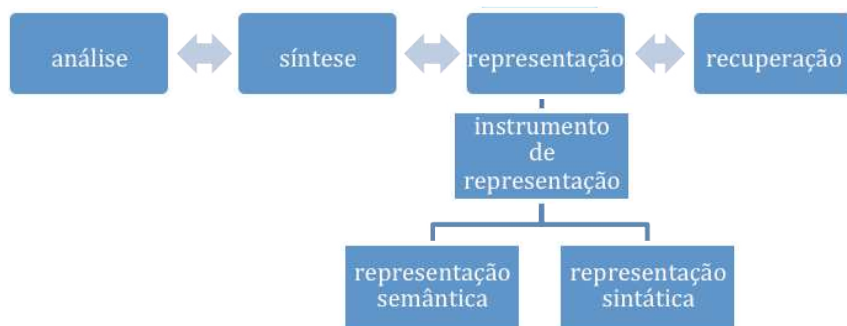


Figura 2: Esquema de operações que estão na base do paradigma do processo documentário de García Gutiérrez. Fonte: García Gutiérrez, 1992, p. 28.

A etapa de análise, para este autor (García Gutiérrez, 1992, pp. 28-29), é a operação inicial equivalente a uma leitura empírica que consiste em dois momentos: reconhecimento e desestruturação do texto de acordo com certos parâmetros, por um lado, e seleção

das proposições do produtor, por outro. A síntese é a recomposição indutiva do discurso fragmentado previamente por análise dedutiva. Segundo Castilho observa, esta definição de síntese não coincide com o que Kobashi apresenta em seu esquema. Para Kobashi (1994), como tratamos anteriormente, síntese é parte do processo de desestruturação do texto, enquanto para García Gutiérrez, trata-se de um processo de recomposição.

Também é importante destacar, segundo Castilho, que no esquema proposto por García Gutiérrez, as flechas que interligam as operações indicam para ambos os sentidos. Portanto, para ele, as etapas da Análise Documentária não apresentam uma rigidez em sua ordem de realização e pode-se voltar a qualquer uma no momento em que se realiza a outra. Isso quer dizer que, mesmo o usuário no momento da recuperação da informação pode contribuir para a análise do texto. Castilho considera o P/PD de García Gutiérrez o mais completo para representar as operações envolvidas na Análise Documentária.

Como observamos, o termo Análise Documentária traz em seu cerne a relevância da etapa de análise para uma representação mais propícia a uma recuperação da informação satisfatória. Cunha (1989b, p. 18) afirma que a Análise Documentária apresenta três momentos determinantes: a leitura, a segmentação e a representação em linguagens documentárias do texto analisado, sendo que os dois primeiros momentos constituem a análise do texto da qual depende intrinsecamente a representação. Kobashi (1989), de modo semelhante, centra-se na análise, em sua proposta de modelo, por considerar como problema maior e anterior ao do controle do vocabulário a necessidade de estabelecer metodologias que permitam reconhecer estruturas discursivas e identificar as informações substantivas dos textos do ponto de vista da Análise Documentária.

Quanto às relações com outras áreas, após discorrer sobre a interface da Documentação com a Linguística, inaugurada por

Gardin, Kobashi (1994) trata, entre outros, da Inteligência Artificial. Para ela (1994, p. 35), essa aproximação se explica pelo fato de a Inteligência Artificial ter desenvolvido procedimentos que se mostraram ser aplicáveis aos processos documentários, como é o caso dos processos de catalogação, de indexação e de produção automática de resumos. A despeito do interesse pelos processos documentários, Kobashi (1994, p. 79) chama a atenção para a impossibilidade de se transpor procedimentos de forma mecânica, já que a especificidade da Análise Documentária requer a elaboração de modelos próprios. Afirma que muitos sistemas são protótipos laboratoriais, sem possibilidade de aplicação em escala real.

Das relações com a Linguística, o Grupo Temma passou a adotar elementos terminológicos para os estudos sobre Análise Documentária, visando contemplar concretamente aspectos pragmáticos, por meio da incorporação da terminologia dos usuários à linguagem documentária. Ainda que tratemos aqui de Análise Documentária em sentido estrito, apresentamos o texto que segue, uma vez que faz-se necessário distinguir processo e instrumento, tanto quanto estabelecer as relações devidas entre eles, e suas implicações.

Tálamo, Lara, Kobashi e Amaro (1992) afirmam que as reflexões mais significativas entre a Linguística e a Ciência da Informação tiveram início na década de 1970 com os trabalhos de Sparck Jones e Kay, Montgomery e Gardin. Neste sentido, as autoras consideram que a Ciência da Informação apresenta-se como um campo privilegiado para o estudo e validação de certos modelos da Linguística. No entanto, as relações com a Terminologia vêm sendo exploradas por oferecerem novas alternativas sobre questões para as quais ainda não se havia obtido respostas efetivas.

As autoras tratam da Linguística Documentária, desenvolvida por García Gutiérrez (1990), para quem trata-se de uma área que se ocupa da formalização do processo comunicativo-documental,

o qual é composto de elementos cujo comportamento é fundamentalmente linguístico. As autoras entendem que, mesmo considerando o avanço promovido pela proposta deste pesquisador, para que as linguagens documentárias possam produzir representações que não sejam apenas virtuais, é necessária a incorporação de elementos teórico-metodológicos da Terminologia. Afirmam que pesquisas que pretendem resgatar o aspecto propriamente linguístico, entendendo a linguagem documentária como mera nomenclatura, promovem processo de indexação restrito à substituição de palavras por outras palavras (Tálamo, Lara, Kobashi & Amaro, 1992, pp. 5-6).

2.2. Análise de Assunto

Uma vez a abordagem construída sobre o termo Análise de Assunto ter origem no idioma inglês, selecionamos algumas obras neste idioma significativas para compreensão do uso do termo.

Deste modo, cabe tratar do livro “*The subject approach to information*”, de Antony Charles Foskett, traduzido no Brasil como “A abordagem temática da informação” publicado em 1973. Como é recorrente na literatura, no capítulo “Análise de Assuntos” que estamos tratando, Foskett (1973, pp. 34-46) não discorre sobre o processo de análise, mas sobre os instrumentos adotados para tal e as características de suas unidades. Neste capítulo o autor inicia falando brevemente das noções de paradigma e sintagma, citando sua introdução por Gardin na década de 1960, no contexto das linguagens de indexação. Assim, afirma que uma linguagem de indexação, como qualquer outra linguagem, é composta por vocabulário e sintaxe, questão reformulada posteriormente pelo Grupo Temma, como dissemos, em função da compreensão de que o eixo sintagmático ocorre efetivamente no momento da indexação e da busca, e não na linguagem de indexação. Quanto ao uso de

assunto, assim como, assunto simples e assunto composto, por Foskett, sobre a proposta de Gardin, o estranhamento se faz uma vez que ‘assunto’ não compõe o quadro explicativo adotado por este autor. Os desenvolvimentos posteriores da área permitiram esclarecer a relação termo-conceito, contribuindo para a revisão da ideia de assunto composto como unidade constituída por mais de uma palavra que, por este motivo, faria referência a mais de um conceito. Essas questões merecem maior atenção, ainda hoje, dado o fato de Foskett ser adotado como referência na área, dada a relevância de sua obra.

Em capítulo anterior, “O documento como evidência”, Foskett (1973, p. 23) pergunta sobre como determinar o assunto de um documento, propondo o que ele chama de atalhos: ler o sumário, o prefácio ou a introdução, o comentário do editor na orelha da obra, o resumo, no caso de um artigo de periódico ou relatório técnico, e as reivindicações quando se tratar de uma patente. Chama a atenção, ainda, para os casos em que o título não contribuirá para a identificação do assunto abordado no documento. Apresenta, então, algumas técnicas como a indexação de palavras do título que dispensem o trabalho intelectual, entendendo haver poucas pessoas qualificadas para executá-lo.

Em um dos artigos sobre Análise de Assunto do ARIST (Lancaster, Elliker & Connell, 1989), o tema é introduzido como a presença, identificação e expressão de assunto (*subject matter*) em documentos textuais, bases de dados, vocabulários controlados e linguagens naturais, pedidos de informações e estratégias de pesquisa. A definição adotada para o tema no artigo é ampla, não se restringindo ao processo de Análise do Assunto. Observamos também que a literatura selecionada é aquela publicada em língua inglesa, a qual não fornece visão abrangente sobre outras abordagens, apesar de pretender ser uma revisão anual.

Outra obra de título pertinente a este trabalho é o livro *Subject analysis*, de Langridge (1989). Para este autor, Análise de Assunto é a parte inicial, a mais importante e a mais difícil de toda classificação e indexação, pois todo sistema de recuperação depende da análise em que foi baseado. O autor afirma que este tema é, em geral, indiretamente abordado na literatura e esta lacuna tem deixado deficiências na prática (Langridge, 1989, p. 1).

Para Langridge, a aplicação particular da Análise de Assunto se dá na seleção de materiais, na interpretação de perguntas e na indexação. A interpretação de perguntas compreende uma série de processos paralelos àqueles da indexação, como mostra o diagrama:



Figura 3: Esquema de processos paralelos entre a prática da Análise de Assunto e a interpretação de perguntas, de acordo com Langridge. Fonte: Langridge, 1989, p. 5.

Tratando das etapas da indexação, Langridge (1989, p. 6) diz que primeiro devem ser determinadas as características significantes de um documento, e então é que se procede à análise propriamente. O segundo estágio da indexação é o de traduzir o assunto percebido de um documento em termos de uma linguagem de índice, o que significa representar o assunto na notação apropriada de um esquema de classificação ou no vocabulário selecionado de uma lista de cabeçalho de assunto. O terceiro estágio da indexação é o de proceder a uma entrada apropriada em um índice para o documento.

Com relação à autonomia do processo de Análise de Assunto frente aos instrumentos de representação, o autor argumenta que

os conteúdos de um livro seriam aquilo que são se não houvesse esquemas de classificação ou de cabeçalhos de assunto. A Análise de Assunto é sempre a mesma porque retrata o documento e não o sistema (Langridge, 1989, p. 7).

Como vemos, Langridge ressalta várias questões relevantes quanto ao processo de análise, como sua não subordinação aos instrumentos de representação e seu papel primordial no resultado dessa representação.

Quanto a tratar do processo de indexação e o de classificação no contexto da Análise de Assunto, comentamos o que segue, visando precisar os processos citados. A classificação bibliográfica não se refere exclusivamente aos aspectos temáticos do documento, como já havia constatado Sayers (1915), ao analisar a Classificação Decimal de Dewey (CDD): segundo ele, há obras compostas de vários assuntos, como enciclopédias e periódicos; obras nas quais a forma predomina sobre o assunto, como as obras literárias; e obras em que assuntos específicos são tratados segundo pontos de vista particulares, como em uma abordagem histórica, teórica e outros.

O autor trata da Análise de Assunto de dois modos. Um deles é a sumarização, que é aplicada a um trabalho, seja um documento extenso ou breve, com o objetivo de o recuperar em uma coleção em resposta a uma pergunta em particular. O outro modo, conhecido como indexação profunda, apresenta escopo menor, mas sua profundidade é muito maior, e é realizada por meio da indexação de cada tópico mencionado em um documento, objetivando a recuperação da informação. No processo de sumarização, têm-se como produtos notações de uma classificação ou cabeçalhos de assunto, adotados para arranjo de livros em bibliotecas, catálogos de bibliotecas gerais ou especializadas, e bibliografias e índices de artigos de periódicos. Na indexação profunda, os produtos são índices individuais de documentos, como os índices de livros.

O que se conclui sobre a abordagem proposta por Langridge é que a sumarização serve para representar um grande assunto do documento, enquanto a indexação profunda é que representaria o conteúdo do documento de forma específica. Assim, para realizar uma busca por determinados conteúdos em uma biblioteca, o usuário primeiro buscaria por grandes assuntos e, ao localizar os documentos correspondentes a eles, o usuário faria uma busca em cada um deles pelo seu índice interno de assuntos.

O autor afirma desconhecer sistemas que realizem a indexação profunda em todos os livros do sistema, compondo um único índice. Dado que a obra de Langridge foi publicada em 1989, é de se estranhar que ele não faça referência às grandes bases de dados temáticas e às bases de dados dos centros de documentação, uma vez que estas realizam a indexação de todos os documentos, em especial, artigos de periódicos, mas também livros e outros.

Outro ponto é que a diferença de objetivos implica análises próprias. O índice de um livro objetiva contribuir para a sua leitura, subsidiando a busca por tópicos de interesses no corpo do texto. Em um sistema, o índice representa todos os documentos que compõem a coleção e atende ao objetivo de acesso às informações identificadas nestes documentos.

Quanto a isso, o artigo de Chu e O'Brien (1993) apresenta o instigante título: "Subject analysis: the critical first stage in indexing". As autoras afirmam que muitos estudos sobre indexação são, de fato, sobre índices, o que conduziu a que uma parte do processo de indexação, ou seja, a Análise de Assunto fosse negligenciada.

As autoras consideram como estágios do processo de indexação:

- análise do assunto do texto
- expressão do conteúdo do assunto nas palavras do indexador
- tradução em um vocabulário de indexação
- expressão do assunto em termos do índice

Chu e O'Brien (1993) entendem que, para catalogadores de assunto e indexadores experientes, os primeiros três passos podem acontecer quase simultaneamente, mas isto não implica que se trate de uma atividade simples. Reforçando a percepção de Langridge, se o indexador realiza a Análise de Assunto com base em um vocabulário, e não distingue devidamente as etapas do processo, ele pode vir a desconsiderar algumas das nuances que poderiam melhorar os termos do índice.

Já no estágio do processo que se chamou de determinação do *aboutness* no texto (ou, do que trata o texto), as autoras mencionam a existência de muitas abordagens. A abordagem teórica desenvolve uma interpretação comportamental e/ou linguística e a abordagem prática é aquela presente em livros-texto ou manuais sobre indexação e guias de indexação com ênfase sobre sistemas individuais.

Quanto aos estudos que tratam da consistência no processo de indexação, as autoras afirmam que a escolha dos termos apropriados é fortemente dependente do resultado do estágio da Análise de Assunto e não deve ser totalmente atribuída, como em muitos estudos de indexação, à experiência do indexador e a particularidades do sistema de indexação.

As autoras, no entanto, não discorrem sobre um método para a Análise de Assunto; elas observam, na prática, os modos de realização deste processo, buscando entender os fatores de influência que predominam na análise, sem a limitação de algum sistema particular de indexação.

Partindo para o contexto brasileiro, a norma NBR 12676 (*Associação Brasileira de Normas Técnicas [ABNT], 1992*), intitulada "*Métodos para a análise de documentos – Determinação de seus assuntos e seleção de termos de indexação*", define indexação como o ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação. Como estágios da indexação, apresenta: exame do docu-

mento e estabelecimento do assunto de seu conteúdo; identificação dos conceitos presentes no assunto; e tradução desses conceitos nos termos de uma linguagem de indexação. A norma trata, em seguida, da análise do documento, que depende de sua forma física, sendo os impressos considerados os mais comuns em bibliotecas e centros de informação, para os quais recomenda-se a leitura dos seguintes elementos: título e subtítulo; resumos, se houver; sumário; introdução; ilustrações, diagramas, tabelas e seus títulos explicativos; palavras ou grupos de palavras em destaque e referências bibliográficas. Já para os documentos não impressos, como multimeios e realia, a norma informa que a indexação é geralmente realizada pelo título e/ou sinopse. Deste modo, na norma NBR 12676 são indicados elementos do documento a considerar na análise. Assim, o objeto em questão é o documento, não o texto, e os procedimentos que permitem a análise e a síntese não são apresentados, o que indica o não reconhecimento da especificidade da análise como etapa e, portanto, de sua função.

Considerando a especificidade da tese de Naves (2000), intitulada “Fatores interferentes no processo de análise de assunto: estudo de caso de indexadores”, para este trabalho, segue uma breve síntese dos trechos considerados mais importantes.

A autora afirma que há ambiguidade no uso do termo “assunto”, definindo-o em seu estudo como “o tema, o conteúdo informativo de um documento” (Naves, 2000, p. 249). Ela estabelece, portanto, equivalência entre os termos “assunto” e “tema”. Quanto ao termo Análise de Assunto, em glossário elaborado ao final da tese, afirma tratar-se de operação base da indexação de assuntos, compreendendo o processo pelo qual passa o indexador para extrair o conteúdo de um documento.

Naves cita Cesarino e Pinto (1980), que definem Análise de Assunto como a operação de base para todo o procedimento de recuperação de informações, que ocorre em duas situações. Uma

delas se dá quando profissionais recebem um documento e devem dar entrada do mesmo no sistema. Nesta situação, farão uma análise com o objetivo de determinar o conteúdo informativo do documento em questão, tendo em vista o objetivo do sistema e as necessidades do usuário. A outra ocorre quando estes profissionais recebem um pedido de informação e devem fazer uma análise deste com o objetivo de compreender a necessidade de informação transmitida pelo usuário, identificar os conceitos existentes no pedido e traduzi-los para a linguagem adotada no sistema. Observamos a relevância da fala das autoras, uma vez tratar-se de abordagem sistêmica do processo, incluindo o percurso do usuário.

Albrechtsen (1993), também citado por Naves, pergunta de que se está tratando quando se fala sobre assuntos de livros e outros documentos, apresentando um modelo alternativo que permita colocar a indexação em um contexto social mais amplo, além dos métodos de evolução mecânica. Este autor afirma que a indexação deveria ter a função principal de busca por conhecimento e recomenda que o indexador não focalize exclusivamente o conteúdo de documentos, mas tente antecipar o impacto e o valor de um documento para seu uso potencial (Naves, 2000, p. 61). Para Albrechtsen, há três conceitos ou pontos de vista diferentes sobre Análise de Assunto: a) simplista; b) orientado para o conteúdo; e c) orientado para a demanda. Neste último, os documentos são criados e tratados como instrumentos para transmissão de informações às pessoas interessadas, de modo que o indexador não deve se limitar a representar ou resumir apenas a informação explícita no documento, mas perguntar-se como poderia tornar o conteúdo, ou parte dele, visível para o usuário potencial, e que termos utilizar para levar esse conhecimento até o leitor interessado. Podemos dizer que é neste ponto de vista em que o autor busca contemplar o caráter comunicativo, a despeito de falar em transmissão. Outro ponto é o da ‘tentativa’ do indexador

em promover o uso do documento, afirmação que supõe não haver busca de método para realizar este intento.

Naves cita também Todd (1992), para quem há uma considerável confusão terminológica sobre assunto, que pode ser observada por alguns autores clássicos: Cutter define assunto como tema ou tópico, podendo ou não estar no título do documento; Kaiser toma “assuntos” como “coisas em geral”, reais ou imaginárias, e as condições para designá-las, que são chamadas “concretos e processo”; Ranganathan fala sobre o pensamento contido no documento; Coates identifica assunto como abstração da ideia global corporificada no conteúdo de uma unidade literária dada; e Vickery refere-se ao tema a partir do qual livros, partes de livros, artigos ou partes de artigos são escritos. Para Todd, na literatura mais recente, *aboutness* é usado como sinônimo de *subject* de um documento e, segundo Metcalfe, assunto não é um termo satisfatório no âmbito do processo de recuperação da informação por causa de sua ambiguidade (Naves, 2000, p. 59).

Uma definição de assunto que merece destaque, segundo Naves, é a de Hjørland (1992). Sintetizando, assunto seriam as potencialidades de documentos para o avanço do conhecimento (Naves, 2000, pp. 59-60). Este autor afirma que se deve evitar dizer “a” descrição “do” assunto, por não haver um assunto definido para um documento. Entendemos que a fala demonstra a impropriedade da ideia de que os assuntos estariam contidos nos documentos, como algo previamente existente neles.

Naves (2000, p. 60) trata também de Giasson (1993) que busca distinguir assunto e ideia principal. Afirma que o assunto pode ser descoberto quando se pergunta de que trata um artigo, e a ideia principal quando se pergunta sobre qual é a coisa mais importante que o autor nos quer dizer no texto. Assim, diferente de Hjørland, Giasson fala de um conteúdo que está no texto e precisa ser “descoberto”.

Quanto às fases que compõem a Análise do Assunto, Naves (2000, p. 40) discorre do seguinte modo. Do ponto de vista do indexador, o processo é iniciado pela leitura do texto (para o que se necessita do conhecimento de tipos e estruturas de textos), passa-se à fase da extração de conceitos que possam representar o conteúdo temático do texto, para se chegar ao momento da representação da atinência, em que são definidos os termos em linguagem natural que são, depois, traduzidos para uma linguagem de indexação. Explica, ainda, que atinência é a tradução de *aboutness*, adotada no livro de Lancaster (1993), a qual seria a mais utilizada na literatura nacional.

Após realizar uma revisão de literatura sobre a questão da leitura e depois da extração de conceitos, Naves parte para a discussão sobre o processo de atinência, definido como a seleção dos conceitos que sintetizam o assunto do documento. Uma das autoras citadas é Beghtol (1986), para quem um texto tem uma atinência relativamente permanente, mas um número variado de significados. Ou seja, o mesmo documento pode ter diferentes significados para o mesmo leitor em diferentes épocas, mas o documento, este, imutável, possui uma atinência fundamental (Naves, 2000, p. 67).

Citando texto mais antigo, Naves (2000, p. 58) apresenta a questão de outro modo: para Bernier (1965), é difícil ensinar e aplicar a identificação de assunto, especialmente quando são complexos ou implícitos e, para a pessoa sem conhecimento no campo indexado, é impossível a identificação consciente de assuntos. Observamos que estas afirmações são relativas, pois os métodos de indexação permitem interpretação e representação do texto, embora o conhecimento do tema seja necessário.

Em seguida, Naves (2000, p. 71) aborda a interdisciplinaridade em Análise de Assunto, justificando que o indexador sofre influência de vários campos, principalmente oriundos da Linguística, Ciência Cognitiva e Lógica.

Ao tratar dos fatores linguísticos, a autora apresenta algumas questões dos autores brasileiros, que tratamos, que trabalham com a abordagem da Análise Documentária. Em outro momento, cita Svenonius (1981) que aborda a noção de relevância em aspectos de indexação e considera que o termo pode ser considerado como um relacionamento entre o signo (termo do índice) e o significado (conteúdo do assunto) (Naves, 2000, p. 71).

Quanto aos fatores lógicos e cognitivos, Naves discorre, entre outros, sobre Cunha (1987), autora já citada sobre a abordagem da Análise Documentária que, junto com suas colegas, desenvolveu o aporte da Lógica para fins de Documentação, decorrendo depois em disciplina do curso de graduação de Biblioteconomia. Naves cita também a 3ª Conferência Anual da *German Society of Classification*, na qual, em 1979, discutiu-se o tema classificação e cognição.

Naves entende que se o processo de análise é desenvolvido por um indivíduo, significa que ele pensa e, por consequência, segue um raciocínio lógico, em uma determinada língua, “tudo isso regado de uma grande dose de subjetividade” (Naves, 2000, resumo da tese). Para ela, considerando o fator subjetividade, tudo o que for dito pode ser correto ou não, dependendo da interpretação realizada pelas pessoas: ainda que existam várias tentativas de estabelecer alguns critérios e sistematizar o processo, não há um consenso quanto à forma mais adequada de se fazer essa leitura, visando à extração e ao posterior tratamento das informações contidas no texto (Naves, 2000, pp. 53-54).

A revisão bibliográfica realizada por Naves, segundo ela mesma avalia, aponta para textos que adotam o nome Análise de Assunto, mas tratam de estudos sobre linguagens de indexação, formulação das entradas dos índices, problemas relativos à padronização e escolha do melhor termo para representar o conceito, além da indexação como um todo, em detrimento da análise de conteúdo dos documentos propriamente dita.

Naves (2000, p. 40) conclui que há muita confusão conceitual na literatura: diferentes termos atribuídos ao mesmo processo e o mesmo termo atribuído a diferentes processos. Pela leitura da tese, constatamos que o primeiro caso ocorre quando os termos Análise de Assunto e Análise Documentária são usados em referência ao mesmo processo e o segundo caso ocorre quando se trata do uso do termo Análise de Assunto em diferentes acepções. A autora (Naves, 2000, p. 62) aponta o problema da falta de entrosamento entre grupos de pesquisa que sofrem influências diversas, ora da Documentação, da escola francesa, ora da Ciência da Informação, da escola americana, referindo-se claramente à realidade brasileira.

Posteriormente, Eduardo Wense Dias, orientador da tese de Naves, publicou, em coautoria com ela, o livro *Análise de Assunto: teoria e prática*, em duas edições (de 2007 e de 2013). Neste livro, bastante adotado no Brasil, os autores são zelosos em alertar os leitores sobre as distinções de significado intrínsecos aos termos Análise de Assunto e Análise Documentária. Optando usar o primeiro termo, desenvolvem o texto contemplando textos pertinentes, relativos a ambas abordagens. Na introdução, apresentam as definições iniciais de Análise Documentária (que tratamos anteriormente) e afirmam tratar-se – assim como consta na tese de Naves (2000, p. 40) – de “operação semântica, mesmo que não obedecendo a nenhuma regra precisa e varie em função de cada organismo e do analista. Este seleciona as palavras-chave, normalmente de forma intuitiva, em função de sua ocorrência e do seu interesse para a instituição” (Dias & Naves, 2013, p. xi). No entanto, como vimos, a prática intuitiva é motivadora do desenvolvimento da Análise Documentária e não sua característica.

Quanto à comparação dos termos, para Rivier (1992, p. 80 e 84)

as tendências mais recentes parecem girar em torno da idéia do que os anglo-saxões denominam *aboutness*, que, traduzido,

literalmente, é “do que” trata um texto, isto é, seu assunto. [...] Este estudo mostra a vitalidade das pesquisas teóricas sobre linguagens de indexação, levadas a efeito por mais de quarenta anos nos países anglo-saxões, no leste europeu e na Índia, em número bem mais expressivo do que nos países de língua francesa onde se nota modesta contribuição do ponto de vista quantitativo.

Segundo este autor, no início da década de 1990, a literatura sobre o enfoque da Análise de Assunto era significativamente maior, haja vista a predominância da orientação anglo-americana em nossa área.

3. Comparação entre os termos Análise de Assunto e Análise Documentária na literatura fundacional (décadas de 1960-1990)

Nos quadros seguintes, apresenta-se uma seleção das ideias apresentadas sobre cada termo de acordo com as categorias: 1) definições; 2) objetivos; 3) objetos; e 4) princípios teóricos e metodológicos, incluindo aportes de outras áreas.

1) Definições: Análise Documentária	Autoria
Representação, por meio de termos e de processos sintáticos convencionais, de um certo conteúdo dos documentos científicos com fim de classificação e de pesquisa de informação.	Dubois et al., 1998, p. 203
Conjunto de procedimentos utilizados tanto pelo documentalista na captação das proposições do texto, sua segmentação e representação formal, como pelo usuário no processo interrogador e interativo com a memória documental.	García Gutiérrez, 1992, citado por Castilho, 1995
Compreende uma série de procedimentos metodológicos, cujo objetivo é o de viabilizar a análise de textos técnico-científicos para seleção de conteúdos informativos, que serão representados, recuperados e disseminados.	Tálamo, Lara, Kobashi & Amaro, 1992

Figura 4: Definições: Análise Documentária

1) Definições: Análise de Assunto	Autoria
Operação de base para todo o procedimento de recuperação de informações, que ocorre em duas situações: na entrada do documento no sistema e quando há um pedido de informação.	Cesarino & Pinto, 1980
Presença, identificação e expressão de assunto (<i>subject matter</i>) em documentos textuais, bases de dados, vocabulários controlados e linguagens naturais, pedidos de informações e estratégias de pesquisa.	Lancaster, Elliker & Connell, 1989
Operação base da indexação de assuntos, compreendendo o processo pelo qual passa o indexador para extrair o conteúdo de um documento.	Naves, 2000, p. 249

Figura 5: Definições: Análise de Assunto

2) Objetivos: Análise Documentária	Autoria
Significados do texto são retirados e substituídos por outros que não são, necessariamente, encontrados no texto, caracterizando operações de construção de metalinguagens, com o objetivo de permitir buscas retrospectivas de informação.	Gardin, 1970, 1973a, 1973b, citado por Lara, 1999
Não se trata apenas da representação do texto mas de uma reconstrução do percurso epistemológico do discurso, que visa tanto a identificação do corpus em análise e de sua validação, como a possibilidade de simular construções análogas. Ela é voltada à explicitação e/ou proposição de modelos metodológicos de funcionamento dos processos de identificação, descrição e representação de conteúdos de textos.	Gardin, J.-C., Lagrange, M.-S., Martin, J.-M., & Natali-Smit, J., 1981, citado por Cunha, Kobashi & Amaro, 1987
Aquela que tem por objetivo a recuperação do documento mediante a representação de traços descritivos de seu conteúdo por meio de termos, sintagmas ou outras unidades discretas, denomina-se indexação.	García Gutiérrez, 1992, p. 31
Propõe princípios para a elaboração de informações documentárias, definidas, por sua vez, como múltiplas representações de textos para fins de recuperação face às diferentes demandas de informação.	Tálamo, Lara, Kobashi & Amaro, 1992
Determina e cria princípios e mecanismos que promovam a circulação de informações e de documentos.	Kobashi, 1994
Permite selecionar, de um universo de objetos, aqueles que poderão responder a uma necessidade de informação.	

Figura 6: Objetivos: Análise Documentária

2) Objetivos: Análise de Assunto	Autoria
A entrada de documentos no sistema é feita por uma análise com o objetivo de determinar o conteúdo informativo do documento em questão, tendo em vista o sistema e as necessidades do usuário. Frente a um pedido de informação, é feita uma análise deste com o objetivo de compreender a necessidade de informação, identificar os conceitos existentes no pedido e traduzi-los para a linguagem adotada no sistema.	Cesarino & Pinto, 1980, citadas por Naves, 2000

Tem aplicação particular na seleção de materiais, na interpretação de perguntas e na indexação.	Langridge, 1989
Há dois modos de Análise de Assunto. A sumarização é a aplicação em um trabalho todo, seja um longo livro ou um artigo pequeno, cujo objetivo principal é recuperar documentos de uma coleção em resposta a uma pergunta em particular. A indexação profunda apresenta escopo menor, mas de profundidade maior, sendo realizada por meio da indexação de cada tópico mencionado em um documento, objetivando a recuperação da informação.	
No processo de sumarização, têm-se como produtos números de classificação ou cabeçalhos de assunto. Na indexação profunda, os produtos são índices individuais de documentos.	

Figura 7: Objetivos: Análise de Assunto

3) Objetos: Análise Documentária	Autoria
É um tipo de análise semântica que se aproxima da análise de conteúdo do ponto de vista do objeto: textos a serem analisados.	Gardin, 1974b, citado por Kobashi, 1989
Dois objetos: o texto, como a sua unidade de análise; e a linguagem documentária como instrumento comutador. Ao operar com textos, a Análise Documentária lida, portanto, com a linguagem.	Kobashi, 1994
Do ponto de vista metodológico, define-se, fundamentalmente, como uma operação com textos.	

Figura 8: Objetos: Análise Documentária

3) Objetos: Análise de Assunto	
Cutter define assunto como tema ou tópico, podendo ou não estar no título do documento; Kaiser toma “assuntos” como “coisas em geral”, reais ou imaginárias, e as condições para designá-las, que são chamadas “concretos e processo”; Ranganathan fala sobre o pensamento contido no documento; Coates identifica assunto como abstração da ideia global corporificada no conteúdo de uma unidade literária dada; e Vickery refere-se ao tema a partir do qual livros, parte de livros, artigos ou parte de artigos são escritos; para Todd, na literatura mais recente o termo <i>aboutness</i> é usado como sinônimo do termo <i>subject</i> de um documento.	Todd, 1992, citado por Naves, 2000, p. 59
Há ambiguidade no uso do termo, mas ele é definido no trabalho como o tema, o conteúdo informativo de um documento.	Naves, 2000, p. 249

Figura 9: Objetos: Análise de Assunto

4) Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise Documentária	Autoria
É um tipo de análise semântica que se aproxima da análise de conteúdo do ponto de vista do objeto: textos a serem analisados, devendo-se chegar ainda à sua representação pela mediação de um sistema simbólico, ou seja, uma metalinguagem. Aproxima-se também da Semiótica pela preocupação com a formulação de simbolização da linguagem.	Gardin, 1974b, citado por Kobashi, 1989

Desenvolvimento, junto com colegas, de aporte da Lógica para fins de Documentação, decorrendo em disciplina do curso de graduação de Biblioteconomia.	Cunha, 1987
A Linguística Documentária é uma área que se ocupa da formalização do processo comunicativo-documental, o qual é composto de elementos cujo comportamento é fundamentalmente linguístico.	García Gutiérrez, 1990, citado por Tálamo, Lara, Kobashi & Amaro, 1992
Integra o campo da Ciência da Informação, mas se insere no processo geral de Comunicação, na medida em que elabora produtos que promovem a socialização de bens culturais.	Tálamo, Lara, Kobashi & Amaro, 1992
As reflexões mais significativas entre a Linguística e a Ciência da Informação tiveram início na década de 1970 com os trabalhos de Sparck Jones e Kay, Montgomery e Gardin.	
Posteriormente, relações com a Terminologia passaram a ser exploradas por oferecerem novas alternativas sobre questões para as quais não se havia obtido respostas efetivas apenas por meio da Linguística Documentária.	
A Inteligência Artificial desenvolveu procedimentos que mostram ser, à primeira vista, aplicáveis aos processos documentários, mas é preciso observar a impossibilidade de se transpor procedimentos de forma mecânica, já que a especificidade da Análise Documentária requer a elaboração de modelos próprios, e muitos sistemas são protótipos laboratoriais, sem possibilidade de aplicação em escala real.	Kobashi, 1994
Certos conceitos da Semiologia e da Filosofia da Linguagem relacionados aos problemas de representação e de significação.	

Figura 10: Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise Documentária

4) Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise de Assunto	Autoria
A interdisciplinaridade em Análise de Assunto é abordada frente ao fato de que o indexador sofre influência de vários campos, principalmente oriundos da Linguística, Ciência Cognitiva e Lógica.	Naves, 2000

Figura 11: Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise de Assunto

4. Comparação entre os termos Análise de Assunto e Análise Documentária nos eventos e capítulos da ISKO (2013-2018)

Com a proposta de investigar os avanços acerca da Análise de Assunto e da Análise Documentária, analisamos o uso dos termos

no contexto da literatura publicada nas nove edições dos eventos da ISKO International (2014, 2016, 2018), ISKO Brasil (2013, 2015, 2017) e ISKO Espanha e Portugal (2013, 2015, 2017) ocorridas nos últimos seis anos, no período de 2013 a 2018¹. A análise foi realizada segundo as mesmas categorias da análise do período anterior: 1) definições; 2) objetivos; 3) objetos; e, 4) princípios teóricos e metodológicos, incluindo aportes de outras áreas. A escolha dos eventos da ISKO e dos dois capítulos foi motivada pela certeza da identificação dos termos em trabalhos de pesquisa da área de Organização do Conhecimento – termo que compõe o nome desta Sociedade e bastante adotado hoje – e, também, porque os eventos dos capítulos do Brasil e de Espanha e Portugal são representativos da Ibero-América onde coexistem as respectivas linhas teóricas dos termos investigados.

O levantamento quantitativo nas nove edições dos três eventos registrou um total de 638 trabalhos publicados, dos quais, em 56 foram identificados os termos Análise de Assunto ou *Análisis de Assunto* ou *Subject Analysis* e, em 70 foram identificados os termos Análise Documentária ou Análise Documental ou *Análisis Documental* ou *Documentary Analysis* (a forma Análise Documental é usada em língua portuguesa no Brasil e em Portugal). Com os textos, procedemos à identificação e extração de trechos dos textos publicados que tivessem relação com qualquer uma das categorias de análise, cuja sistematização foi registrada em quadros específicos de cada categoria. Essa sistematização propiciou a análise comparada de cada categoria para ambos os termos – Análise Documentária e Análise de Assunto – após seleção dos trechos mais significativos.

¹ O acesso às referências bibliográficas deve ser realizado pelo link: <https://drive.google.com/drive/folders/1m-HajKlH0wbcRFpO7uEUbOGJCAw792-F?usp=sharing>

1) Definições: Análise Documentária	Autoria
As definições são de autores cujos estudos aportam fundamentos da Linguística com destaque para as definições de Coyaud (1966) e de Gardin (1973a), este último considerado o autor que fundou as bases da Análise Documentária de linha francesa. Coyaud (1966): operação de reconhecimento das unidades lexicais que representam as noções importantes de um documento. Gardin (1973a): um conjunto de procedimentos efetuados com a finalidade de expressar o conteúdo dos documentos científicos, sob formas destinadas a facilitar a recuperação da informação.	ISKO Espanha e Portugal, 2013 a) Gomes, Guedes & Santos
Na ISKO Brasil de 2015, a programação incluiu uma mesa de pesquisadores do Grupo Temma para apresentarem seus estudos sobre Análise Documentária cujas definições seguem a linha de Gardin. Nas citações destes autores e dos outros que participaram do evento, é possível observar as definições voltadas para a análise de conteúdo de textos, bem como a função de facilitar a recuperação.	ISKO Brasil, 2015 a) Smit b) Lima c) Alves & Moraes d) Cândido, Moraes & Sabbag
Os trabalhos que incluem definição de Análise Documental seguem a mesma visão de Gardin (1973a), acompanhada de definições com base nos pensamentos de Coyaud (1966) (matriz linguística), estudiosos da linha francesa de Análise Documental.	ISKO Espanha e Portugal, 2015 a) Guimarães, Dalessandro & Martínez-Ávila
A decomposição da estrutura textual é incluída como resultado da Análise Documentária.	ISKO Brasil, 2017 a) Bezerra & Santos

Figura 12: Definições: Análise Documentária

1) Definições: Análise de Assunto	Autoria
A Análise de Assunto é considerada etapa do tratamento temático com a finalidade de extração de conceitos.	ISKO Espanha e Portugal, 2015 a) Araújo, Paula, Oliveira & Torres
Além da extração de conceitos, a Análise de Assunto é relacionada à teoria da indexação com enfoque na determinação da tematicidade pela identificação e seleção de conceitos que a distingue da etapa de tradução.	ISKO Brasil, 2017 a) Tartarotti, Dal'Evedove & Fujita b) Santos & Corrêa
Com denominação equivalente às análises temática, documentária, conceitual e de conteúdo, a Análise de Assunto segue três concepções durante a indexação, segundo Albrechtsen (1993): simplista, orientada para o conteúdo e orientada para a demanda.	ISKO Espanha e Portugal, 2017 a) Tognoli, Milani & Guimarães b) Tartarotti, Dal'Evedove & Fujita
A indexação é vista pela abordagem semiótica e registra mudança na Análise de Assunto pelo contexto de produção de significados implícitos.	ISKO International, 2018 a) Kleineberg

Figura 13: Definições: Análise de Assunto

2) Objetivos: Análise Documentária	Autoria
Aplica-se a diferentes tipos de estruturas textuais para realizar a identificação e representação do conteúdo mais relevante para posterior representação.	ISKO Espanha e Portugal, 2013 a) Cândido & Moraes b) Guimarães, Matos, Fernández-Molina & Lopes
A análise documental engloba a indexação cujo objetivo é o tratamento de conteúdos documentais e sua recuperação pelo usuário.	ISKO Espanha e Portugal, 2015 a) Terra, Fujita & Agustín-Lacruz
A Análise Documentária é colocada como uma das três perspectivas junto à indexação e análise temática com o objetivo de representar a informação pela análise de conteúdo dos documentos.	ISKO Brasil, 2017 a) Felipe & Pinho
A Análise Documentária no contexto de arquivos é realizada com aporte da diplomática para a identificação, representação e organização dos conteúdos com objetivo de troca, difusão e uso.	ISKO International, 2018 a) Tognoli & Rodrigues

Figura 14: Objetivos: Análise Documentária

2) Objetivos: Análise de Assunto	Autoria
Influência das concepções de Análise de Assunto orientada para a demanda no contexto da unidade ou sistema de informação em que o documento se insere com o objetivo de atender as demandas de usuários.	ISKO Espanha e Portugal, 2015 a) Guimarães & Evangelista
A identificação e representação do conteúdo para a recuperação por assuntos são evidenciados como objetivos.	ISKO International, 2016 a) Tartarotti & Fujita
Ressaltam-se como objetivos a identificação e seleção de conceitos, a representação do conteúdo e a recuperação por assuntos.	ISKO Espanha e Portugal, 2017 a) Neves & Fujita b) Souza
O enfoque é a qualidade da representação por assuntos para o alcance das necessidades dos usuários mediante recuperação da informação.	ISKO Brasil, 2017 a) Tartarotti, Dal'Evedove & Fujita

Figura 15: Objetivos: Análise de Assunto

3) Objeto: Análise Documentária	Autoria
A análise documental tem como objeto o conteúdo documental.	ISKO Espanha e Portugal, 2015 a) Guimarães, Dalessandro & Martínez-Ávila
A imagem como conteúdo não textual, considerado como objeto da Análise Documentária; Gardin (1973a), ao definir a informação documentária como produto da Análise Documentária, esclarece que é resultado da transformação do texto original.	ISKO Brasil, 2015 a) Souza b) Lima
O texto narrativo é visto como objeto da Análise Documentária, além do texto científico citado na definição de Gardin (1973a) para Análise Documentária.	ISKO Espanha e Portugal, 2017 a) Caprioli, Lima & Moreira

Figura 16: Objeto: Análise Documentária

3) Objetos: Análise de Assunto	Autoria
A Análise de Assunto, sob enfoque do processamento cognitivo do indexador, é realizada mediante estratégias de identificação e seleção de conceitos do seu objeto, o documento e seu assunto.	ISKO Brasil, 2015 a) Tartarotti, Dal'Evedove & Fujita
Considerados como elementos, os objetos de Análise de Assunto são o documento e seu assunto e os outros dois elementos, descrição do assunto e entrada de assunto, são produtos.	ISKO International, 2016 a) Tartarotti & Fujita
O bibliotecário, ao realizar a Análise de Assunto, toma como objeto de sua compreensão o conteúdo do documento.	ISKO Espanha e Portugal, 2017 a) Tognoli, Milani & Guimarães b) Souza

Figura 17: Objeto: Análise de Assunto

4) Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise Documentária	Autoria
O Grupo Temma no Brasil é responsável por trabalhos na linha de Análise Documentária com aportes teóricos da Linguística, Lógica e Psicologia Cognitiva.	ISKO Brasil, 2013 a) Fujita
A Análise Documentária desenvolveu referenciais teórico-metodológicos para o processo de organização da informação fundamentada em matriz linguística francesa de Gardin (1973a) e Coyaud (1973) que produziu influência na Espanha e no Brasil. (citação a) O Grupo Temma, criado na década de 1980, introduziu os estudos de Análise Documental a partir dos estudos de Gardin (1973a) com enfoque em aspectos linguísticos, lógicos e terminológicos. Na Espanha, a influência francesa de Gardin se desenvolveu a partir da década de 1990. (citação b)	ISKO Espanha e Portugal, 2013 a) Gomes, Guedes & Santos b) Guimarães, Matos, Fernández-Molina & Lopes
A Análise Documentária é parte da Semiologia informacional porque é uma forma de manipulação sógnica de natureza representacional. A influência de Gardin demonstra a importância de outras experiências relacionadas à análise do discurso, às proposições lógicas, redes semânticas, à noção de organização semântica ou paradigmática, às categorias usadas na formulação de proposições.	ISKO Brasil, 2015 a) Tálamo & Maimone b) Lara
A Análise Documentária tem aportes interdisciplinares, principalmente, da Linguística e da Lógica, demonstrados em estudos do Grupo Temma. A Análise Documentária de matriz francesa é uma das vertentes teóricas, além da catalogação de assunto e da indexação, do Tratamento Temático da Informação (TTI).	ISKO Brasil, 2017 a) Castanha & Grácio b) Almeida, Cabrini & Castanha

Figura 18: Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise Documentária

4) Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise de Assunto	Autoria
A Análise de Assunto é parte do processo de indexação com aporte teórico-metodológico de linha inglesa.	ISKO Brasil, 2013 a) Fujita
A Análise de Assunto é vista como um processo de decomposição de assuntos por meio de conceitos que são manifestações de facetas e categorias vistas em trabalhos de Kaiser (1911) e Ranganathan (1967).	ISKO Espanha e Portugal, 2013 a) Sales & Guimarães
A Análise de Assunto automatizada é aplicada a todos os objetos culturais a partir do uso da estrutura de camadas de interpretação, conforme desenvolvida por Panofsky (1962): assunto primário, assunto secundário, e significado ou conteúdo intrínseco.	ISKO International, 2014 a) Zeng, Gracy & Žumer
A leitura realizada para Análise de Assunto é vista pela perspectiva interacionista da leitura enquanto processo de comunicação e interação entre três variáveis interligadas: texto-leitor-contexto.	ISKO Brasil, 2015 a) Tartarotti, Dal'Evedove & Fujita
A Análise de Assunto, conforme Hjørland (2016), é considerada como processo de organização do conhecimento ao lado da catalogação, indexação e classificação. As concepções de Análise de Assunto, simplista, orientada para o conteúdo e orientada para a demanda, formuladas por Albrechtsen (1993), devem ser orientadas pela perspectiva sociocognitiva que prevejam o contexto de análise. O ato leitor na Análise de Assunto é vinculado ao entendimento dos aportes teóricos e metodológicos da leitura compreensiva.	ISKO Espanha e Portugal, 2017 a) Mendes b) Tognoli, Milani & Guimarães c) Terra
Pesquisa em Organização do Conhecimento demonstra que o uso de facetas predefinidas na análise e representação do assunto, bem como na interface de pesquisa e navegação, leva a resultados de pesquisa quantitativos e qualitativamente melhores.	ISKO International, 2018 a) Hudon & Fortier

Figura 19: Princípios teóricos e metodológicos, incluindo aportes de outras áreas: Análise de Assunto

5. Considerações finais

A literatura de base selecionada para este estudo, relativa às décadas de 1960 a 1990, permitiu constatar que os termos Análise Documentária (*Analyse Documentaire*) e Análise de Assunto (*Subject Analysis*) nem sempre são abordados quanto à etapa de análise, como as duas denominações indicam. É comum que a indexação (processo como um todo) e a análise (etapa do processo de indexação) não sejam claramente discriminadas. O fato de a análise

não ser explicitada indica sua secundarização, comprometendo a compreensão desta etapa.

Quanto ao objeto, na abordagem da Análise Documentária, o texto é indicado como tal, sendo que, inicialmente, o texto científico foi priorizado. Os termos texto, documento e discurso são questionados quando adotados à falta de rigor terminológico. Assunto – significante central na outra abordagem – não é adotado, mas reconhecido como tópico do texto e como produto da indexação; opta-se por tema, no sentido de objetivo do texto, a partir do qual o método de análise é construído. Na abordagem da Análise de Assunto, toma-se o assunto como objeto, mas essa indicação explícita não foi localizada na literatura estudada; ao levantarem definições de assunto, seus autores afirmam que há imprecisão e que isso dificulta a construção de definições, objetivos e procedimentos. Tanto ‘assunto’ e ‘tema’, quanto ‘texto’ e ‘documento’, são, em geral, usados um pelo outro.

Na literatura sobre Análise Documentária, mostrou-se evidente a busca por métodos rigorosos que fundamentem a etapa da análise, segundo aspectos que caracterizam um texto. A questão, colocada no âmbito da Análise Documentária desde seu início, na década de 1960 por Gardin e seus colegas, é a de que faz-se necessário objetivar o processo de indexação, buscando generalizações. Na literatura sobre Análise de Assunto, recorre-se à identificação dos modos de realização da indexação por profissionais, visando alcançar seu aprimoramento. Neste caso, o caráter subjetivo do processo e sua observação ganha destaque. Como orientação para realização do processo, recomenda-se a leitura de elementos componentes do documento, como título e resumo, ou sugere-se ao indexador que considere as prerrogativas do usuário.

Na Análise Documentária, podemos identificar conceitos e métodos voltados à representação dos diversos documentos, incluindo suas partes, com ênfase ao conteúdo temático, apresentados pelo belga Paul Otlet, em seu Tratado de Documentação, de 1934. Mas,

enquanto linha de pensamento e metodologia de aplicação, a Análise Documentária surgiu a partir das décadas de 1960, em um contexto de ascensão do trabalho com a linguagem sobre textos científicos. Daí, os estudos voltados à fundamentação da linguagem documentária, notadamente o tesouro, que pautados por elementos da Linguística e da Terminologia, poderiam funcionar como instrumentos de comunicação da informação. Já o pensamento e as práticas construídas em torno do termo Análise de Assunto são bastante anteriores, sendo identificados em práticas seculares e em estudos europeus, depois sistematizados no século XIX, por Charles Ammi Cutter, nos Estados Unidos, ganhando visibilidade internacional. O termo Análise de Assunto decorre do contexto de produção dos chamados cabeçalhos de assunto, utilizados na representação dos assuntos dos documentos em sistemas de fichas catalográficas; novos assuntos eram propostos a cada vez que fossem identificados como necessários no momento da indexação dos documentos. A literatura fundacional, no entanto, ocorreu entre as décadas de 1960 a 1990, como tratamos. Desse modo, podemos dizer que as duas abordagens estudadas foram discutidas e fundamentadas no mesmo período.

A produção sobre Análise Documentária, desenvolvida pelo Grupo Temma, partindo de Gardin, pauta-se pela perspectiva comunicacional via estudos de linguagem. Essa produção demonstra as mudanças de enfoques decorrentes da superação dos anteriores, incluindo os contrapontos entre os autores, dinâmica típica da produção científica. No caso da produção sobre Análise de Assunto, encontramos falas sobre a necessidade de mudança quanto ao pensamento predominante, o qual não contemplaria aspectos comunicacionais.

Por sua vez, nos estudos produzidos nos eventos internacionais e capítulos Brasil e Espanha/Portugal da ISKO, realizados de 2013 a 2018, há menor distinção entre as abordagens apresentadas pelos dois termos, embora a diferença ainda seja bastante clara em alguns

textos. É relevante, nestes textos mais recentes, a busca por definições, explicações, etc., sobre os termos adotados e os processos em causa.

Mudança observada no Brasil é a alteração da denominação Análise Documentária, segundo tradução do francês adotada pelo Grupo Temma, para Análise Documental, forma bastante disseminada nos últimos anos em função desta escolha por alguns professores da Universidade Estadual Paulista (UNESP), câmpus Marília.

A despeito da fala de Rivier, que apresentamos anteriormente, sobre a modesta contribuição francesa, já que esta seria menor em termos quantitativos, nosso estudo aponta para algumas particularidades regionais. O levantamento realizado sobre os textos apresentados nos eventos e capítulos da ISKO indicam, como dissemos, que 70 deles pautam-se pela Análise Documentária, frente a 56 que explicitaram a orientação da Análise de Assunto. Mas esses números referem-se principalmente ao Brasil, e secundariamente, a Portugal e Espanha. Além disso, é preciso considerar que o objetivo deste estudo implicou selecionar textos em que os termos em questão foram definidos, seus objetos explicitados ou seus métodos descritos, excluindo-se, desse modo, todos os textos desenvolvidos sem esses elementos. Esse recorte permite explicar o fato de que, contrariamente ao identificado nesta pesquisa, a vivência acadêmica no Brasil hoje evidencia a predominância da abordagem anglo-americana.

Por fim, o estudo nos leva a ressaltar que, dadas as distinções de abordagem construídas entre os termos Análise Documentária e Análise de Assunto, não é viável usá-los indiscriminadamente para indicar o mesmo processo. Afinal, não se pode dizer que se referem a uma mesma forma de pensar e realizar o processo, uma vez que apresentam suas próprias motivações, orientações e resultados.

Referências Bibliográficas

- Associação Brasileira de Normas Técnicas. (1992). *Métodos para a análise de documentos: Determinação de seus assuntos e seleção de termos de indexação: NBR 12676*. Rio de Janeiro: Autor.
- Albrechtsen, H. (1993). Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, 18(4), 219-224. Recuperado de https://www.theindexer.org/files/18-4/18-4_219.pdf
- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113. doi:10.1108/eb026788
- Bernier, C. L. (1965). Indexing process evaluation. *American Documentation*, 16(4), 323-328.
- Castilho, V. (1995). *Para uma indexação automática: métodos de análise de textos em sistemas informatizados aplicados à indexação* (Trabalho de Conclusão de Curso não publicado). Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, Brasil.
- Cesarino, M. A. da N., & Pinto, M. C. M. F. (1980). Análise de assunto. *Revista de Biblioteconomia de Brasília*, 8(1), 32-43. Recuperado de <https://www.brapci.inf.br/index.php/article/view/0000001772>
- Chu, C. M., & O'Brien, A. (1993). Subject analysis: the critical first stage in indexing. *Journal of Information Science*, 19(6), 439-454. doi:10.1177/016555159301900603
- Cintra, A. M. M. (1987). Estratégias de leitura em Documentação. In J. W. Smit (Coord.), *Análise documentária: a análise da síntese* (pp. 27-36). Brasília: IBICT.
- Cooper, W. S. (1978). Indexing documents by Gedanken experimentation. *Journal of the American Society for Information Science*, 29(3), 107-119.
- Cooper, W. S. (1979). Utility theoretic indexing: a note on Wilson's note. *Journal of the American Society for Information Science*, 30(3), 170-172.
- Cunha, I. M. R. F. (1987). Análise documentária. In J. W. Smit (Coord.), *Análise documentária: a análise da síntese* (pp. 37-60). Brasília: IBICT.
- Cunha, I. M. R. F. (1989a). Análise documentária: do texto à representação. *Revista Brasileira de Biblioteconomia e Documentação*, 22(3/4), 121-127. Recuperado de <https://rbbd.febab.org.br/rbbd/article/view/391/365>
- Cunha, I. M. R. F. (1989b). Contribuição para a formulação de um quadro conceitual em Análise Documentária. In I. M. R. F. Cunha, N. Y. Kobashi, & R. K. Amaro (Coords.), *Análise documentária: considerações teóricas e experimentações* (pp. 16-30). São Paulo: FEBAB.
- Cunha, I. M. R. F., Kobashi, N. Y., & Amaro, R. K. (1987). Revisão bibliográfica. In J. W. Smit (Coord.), *Análise documentária: a análise da síntese* (pp. 113-133). Brasília: IBICT.
- Dias, E. W., & Naves, M. M. L. (2007). *Análise de assunto: teoria e prática*. Brasília: Thesaurus.
- Dias, E. W., & Naves, M. M. L. (2013). *Análise de assunto: teoria e prática* (2ª ed. rev.). Brasília: Briquet de Lemos.

- Dubois, J. et al. (1998). *Dicionário de Linguística* (10a ed). São Paulo: Cultrix.
- Foskett, A. C. (1973). *A abordagem temática da informação*. São Paulo: Polígono; Brasília: Ed.UnB.
- Frohmann, B. (1990). Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 46(2), 81-101. doi:10.1108/eb026855
- García Gutiérrez, A. (1990). *Estructura lingüística de la documentación: teoría y método*. Murcia: Universidad de Murcia.
- García Gutiérrez, A. (1992). *Análisis documental del discurso periodístico*. Madrid: CTD-Centro de Tratamiento de la Documentación.
- Gardin, J.-C. (1968). Análise documentária e análise estrutural em arqueologia (J. T. Coelho Netto, trad.). In S. Cordier (Ed.), *Lévi-Strauss*. São Paulo: Documentos.
- Gardin, J.-C. (1970). Procédures d'analyse sémantique dans les sciences humaines. In J. Pouillon & P. Maranda (Orgs.), *Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire* (pp. 628-657). Paris: Mouton.
- Gardin, J.-C. (1973a). Document analysis and linguist theory. *Journal of Documentation*, 29(2), 137-168. doi:10.1108/eb026553
- Gardin, J.-C. (1973b). Linguistique et Documentation. *Bollettino d'Informazione*, 13(2/3), 67-85.
- Gardin, J.-C. (1974a). Analyse documentaire et analyse structurale en archéologie. *Langages* (Nº. temático: Problèmes et méthodes de la sémiologie), 35, 82-86. Recuperado de https://www.persee.fr/doc/lgge_0458-726x_1974_num_8_35_2267
- Gardin, J.-C. (1974b). *Les analyses de discours*. Neuchâtel: Delachaux et Niestlé.
- Gardin, J.-C. (1987). L'analyse logiciste. In J.-C. Gardin, *Systèmes experts et sciences humaines: le cas de l'archéologie* (pp. 17-26). Paris: Eyrolles.
- Gardin, J.-C., Lagrange, M.-S., Martin, J.-M., & Natali-Smit, J. (1981). *La logique du plausible: essais d'épistémologie pratique*. Paris: Maison de Sciences de l'Homme.
- Giasson, J. (1993). *A compreensão na leitura*. Lisboa: Asa.
- Hjørland, B. (1992). The concept of "subject" in Information Science. *Journal of Documentation*, 48(2), 172-200. doi:10.1108/eb026895
- Kobashi, N. Y. (1989). Análise documentária: considerações sobre um modelo lógico-semântico. In I. M. R. F. Cunha, N. Y. Kobashi, & R. K. Amaro (Coords.), *Análise documentária: considerações teóricas e experimentações* (pp. 45-57). São Paulo: FEBAB.
- Kobashi, N. Y. (1994). *A elaboração de informações documentárias: em busca de uma metodologia* (Tese de doutorado em Ciências da Comunicação). Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, Brasil.
- Kobashi, N. Y., Smit, J. W., & Tálamo, M. de F. G. M. (abr. 2001). A função da terminologia na construção do objeto da Ciência da Informação. *DataGramaZero: Revista de Ciência da Informação*, 2(2). Recuperado de <http://www.brapci.inf.br/index.php/article/view/0000001233/a427cd934576c6cfd5d315d13a9598pdf>
- Lancaster, F. W. (1993). *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos.

- Lancaster, F. W., Elliker, C., & Connell, T. H. (1989). Subject analysis. *Annual Review of Information Science and Technology (ARIST)*, 24, 35-84.
- Langridge, D. W. (1989). *Subject analysis: principles and procedures*. London: Bowker-Saur.
- Lara, M. L. G. de (1999). *Representação e linguagens documentárias: bases teórico-metodológicas* (Tese de doutorado em Ciências da Comunicação, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, Brasil). Recuperado de <https://teses.usp.br/teses/disponiveis/27/27143/tde-02122019-153131/pt-br.php>
- Naves, M. M. L. (2000). *Fatores interferentes no processo de Análise de Assunto: estudo de caso de indexadores* (Tese de doutorado em Ciência da Informação). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- Rivier, A. (1992). Construção de linguagens de indexação. *Revista da Escola de Biblioteconomia da UFMG*, 21(1), 56-99. Recuperado de <https://www.brapci.inf.br/index.php/res/v/71744>
- Sayers, W. C. B. (1915). *Canons of classification applied to "The subject", "The expansive", "The decimal" and "The Library of Congress" classifications: a study in bibliographical classification method*. London: Grafton.
- Svenonius, E. (1981). Directions for research in indexing, classification and cataloging. *Library Resources & Technical Services*, 25(1), 88-103.
- Tálamo, M. de F. G. M. (1987). *Elaboração de resumos*. São Paulo: ECA/USP.
- Tálamo, M. de F. G. M., Lara, M. L. G. de, Kobashi, N. Y., & Amaro, R. K. (1992). Interface Análise Documentária, Linguística Documental e Terminologia. In *Resumo: Simposio Iberoamericano de Terminología, Madrid, Espanha*. Madrid: Red Iberoamericana de Terminología.
- Todd, R. T. (1992). Academic indexing: what's it all about? *The Indexer*, 18(2), 101-104. Recuperado de https://www.theindexer.org/files/18-2/18-2_101.pdf
- Van Dijk, T. (1992). *Cognição, discurso e interação* (4ª ed.). São Paulo: Contexto.
- Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Orlando, FL: Academic Press.

CAPÍTULO 6
SUBSÍDIOS TEÓRICO-METODOLÓGICOS PARA A
CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS
THEORETICAL-METHODOLOGICAL SUBSIDIES
FOR THE CONSTRUCTION OF CONTROLLED
VOCABULARIES

Benildes Coura Moreira dos Santos Maculan
Universidade Federal de Minas Gerais (UFMG)
benildesmaculan@ufmg.br
ORCID: 0000-0003-4303-9071

Célia da Consolação Dias
Universidade Federal de Minas Gerais (UFMG)
celiadias@eci.ufmg.br
ORCID: 0000-0003-0891-6454

Marilda Lopes Ginez de Lara
Universidade de São Paulo (USP)
larama@usp.br
ORCID: 0000-0003-1906-8606

Resumo: Os vocabulários controlados proveem um meio eficaz para organizar a informação a ser disponibilizada em sistemas de recuperação de informação. Sua construção é orientada por normas internacio-

nais que compilam as melhores práticas e as convenções sobre os elementos que devem ser considerados para o desenvolvimento de instrumentos mais consistentes. Neste capítulo apresentamos os princípios de construção e uso de diferentes tipos de vocabulários controlados (tesauros, sistemas de classificações, taxonomias, listas de cabeçalhos de assunto e ontologias), tendo como parâmetro a norma ISO 25964 (Parte 1, 2011; Parte 2, 2013). Também são discutidos alguns aportes da Semiótica, da Linguística, da Lógica e da Terminologia para a construção desses instrumentos, destacando a função sógnica do vocabulário controlado e os critérios de organização que dão origem aos vínculos de significação com os públicos da informação.

Palavras-Chave: Vocabulários controlados. Linguagens documentárias. Construção de vocabulários controlados. Organização do Conhecimento e da Informação.

Abstract: Controlled vocabularies provide an effective way to organize information to be made available in information retrieval systems. Its construction is guided by international standards that compile best practices and conventions on the elements that should be considered for the development of more consistent tools. In this chapter we present the principles of construction and use of different types of controlled vocabularies (thesauri, classification systems, taxonomies, lists of subject headings and ontologies), having as parameter ISO 25964 (Part 1, 2011; Part 2, 2013). Also discussed are some contributions of Semiotics, Linguistics, Logic and Terminology for the construction of these tools, emphasizing the process of semiosis, or sign functioning in the controlled vocabulary and the organizational criteria that give rise to the bonds of meaning with the information audiences.

Keywords: Controlled vocabularies. Documentary languages. Construction of controlled vocabularies. Organization of knowledge and information.

1. Introdução

No âmbito da Ciência da Informação, as linguagens especiais, “relacionadas a um determinado conjunto de conhecimentos” (Fujita, Santos & Alves, 2018, p. 23) são denominadas por termos tais como linguagem de indexação (Melton, 1964), linguagem documentária (Gardin, 1969) e vocabulário controlado (Lancaster, 1993). A “variação terminológica, em uma área do conhecimento, mostra suas influências [teóricas] e tendências sócio-históricas e culturais” (Fujita et al., 2018, p. 24). Isso demonstra a convivência de “correntes teóricas como outro componente aliado à questão terminológica (...) como é o caso do termo ‘linguagem documentária’, adotado pela linha de tradição francesa, e do termo ‘linguagem de indexação’, utilizado pela linha anglo-saxã” (Fujita et al., 2018, p. 24).

Construímos vocabulários controlados nos processos de tratamento de recursos informacionais para a alimentação de sistemas de informação em diferentes unidades de informação. O objetivo primordial dos vocabulários controlados é prover um meio eficaz para organizar a informação que será disponibilizada, a partir da descrição de diferentes tipos de informação. Um vocabulário controlado pode ser definido como: “Conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como no da recuperação. A finalidade principal desse controle é fazer coincidir a linguagem do pesquisador com a do indexador” (Cunha & Cavalcanti, 2008, p. 378).

Esse conjunto de termos tem como objetivo facilitar a comunicação em um domínio. É um “instrumento de comunicação entre a informação, o sistema de informação e o usuário, deve assegurar o acesso a essa informação, possibilitando sua adequada recuperação” (Boccatto & Fujita, 2006, p. 18). Ele é considerado uma linguagem construída para propósitos específicos e, citando Gardin (1969), Cintra, Tálamo, Lara, e Kobashi (2002) acrescentam que é preciso

que o vocabulário controlado possua três elementos básicos: um léxico, uma rede paradigmática de relações, em geral, estáveis, e uma rede sintagmática de relações, que expressam as contingências de um contexto particular.

A construção de vocabulários controlados é orientada por normas internacionais que compilam as melhores práticas e as convenções sobre os elementos que devem ser considerados para que os diferentes tipos de instrumentos sejam mais consistentes. A consistência nos projetos de vocabulários controlados se refere à maneira como os recursos informacionais são descritos, o que os tornam familiares e intuitivos para uso pelos usuários, que precisarão apenas de um mínimo de treinamento de recuperação com o instrumento (Hedden, 2016).

No Brasil, destacam-se duas obras nacionais: um projeto do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), coordenado por Hagar Espanha Gomes e publicado em 1984, que teve por base a BSI 5723 (1987), da *British Standard Institution*, a partir da norma ISO 2788 (International Organization for Standardization [ISO], 1986) (*Guidelines for the establishment and development of monolingual thesauri*); e uma tradução para o português, em 1993, da segunda edição das diretrizes da Unesco, publicadas por Austin e Dale (1981), também realizada pelo IBICT. As duas obras se limitam a apresentar diretrizes para a construção de vocabulários controlados do tipo tesouros.

A segunda e última edição da norma ISO 2788 (ISO, 1986) foi a base para a revisão da norma americana ANSI/NISO Z39.19, publicada em 1993. Essa norma já teve diversas atualizações, e a versão mais nova é a de 2005 (National Information Standards Organization [NISO], R2010), que é a quarta edição. Já em 2007, a *International Organization for Standardization* (ISO), tendo por base a norma britânica BS 8723 (British Standards Institution [BSI], 2008), revisou e unificou as normas ISO 2788 e ISO 5964, dando origem à norma

internacional ISO 25964 (ISO, 2011, 2013), em duas partes, que é a mais recente.

2. Princípios de construção e aplicações

Destacamos que entre as normas internacionais dedicadas a oferecer princípios e diretrizes para a construção de vocabulários controlados, a ANSI/NISO Z39.19-2005 (NISO, R2010) e a ISO 25964, Parte 1 e 2 (ISO, 2011, 2013) são as mais recentes e utilizadas pelos desenvolvedores.

Uma análise comparativa entre a ANSI/NISO Z39.19-2005 (NISO, R2010) e a ISO 25964 (ISO, 2011, 2013) demonstra que elas são muito semelhantes, o que evidencia que houve cooperação entre as equipes de seus desenvolvedores (Hedden, 2016), situação que fica clara na leitura de alguns dos relatórios de revisões disponíveis no site oficial da ISO. A diferença mais importante parece ser em relação ao escopo das duas normas, uma vez que a norma ISO abrange recomendações para a construção e manutenção de vocabulários controlados monolíngues e multilíngues e a ANSI/NISO somente para monolíngues.

Hedden (2016) destaca que trabalhar com a norma internacional ISO na construção de vocabulários controlados permite ter uma visão mais ampla, mesmo trabalhando com um instrumento monolíngue, pois será possível uma noção geral dos usos e possibilidades dos instrumentos multilíngues. Partindo dessa premissa, neste capítulo vamos apresentar os princípios de construção e aplicação de diferentes tipos de vocabulários controlados tendo como parâmetro a ISO 25964-1/2.

Ressaltamos que essa norma, em suas duas partes, trata de vocabulários dos tipos: tesouros, sistemas de classificações, taxonomias, listas de cabeçalhos de assunto, ontologias, terminologias, lista de

nomes de autoridades e anel de sinônimos. Desta lista, apenas os três últimos não serão detalhados nesta seção: as terminologias porque elas têm como finalidade tradicional servir como glossários e dicionários; a lista de nomes de autoridades foi excluída porque consideramos que ela tem como função o apoio nas atividades de catalogação; o anel de sinônimos não foi incluído porque apenas compila uma lista de termos para representar um dado conceito.

As próximas seções irão apresentar os princípios de construção e as aplicações dos seguintes vocabulários controlados: tesouros, sistemas de classificações, taxonomias, listas de cabeçalhos de assunto e ontologias.

2.1. Tesouros

O tesouro é um tipo de vocabulário controlado que também pode ser classificado como um tipo de linguagem de indexação, de linguagem documentária e de sistema de organização do conhecimento (SOC). É um conjunto de vocábulos (termos), representados por descritores, preferidos e não preferidos, que expressam os conceitos relevantes de um domínio, estruturados por meio de uma rede lógico-semântica, que pode ser desenvolvido para áreas mais gerais ou para áreas específicas, tais como medicina, artes, música e outras. Nesse sentido, a finalidade essencial de um tesouro é cobrir a terminologia de um domínio e “(...) guiar o indexador e o usuário na seleção de um mesmo termo preferido ou da mesma combinação de termos preferidos para representar um dado assunto” (ISO, 2011, p. 12).

Para Boccato, Ramalho e Fujita (2008, p. 201), os tesouros são “(...) estruturas combinatórias e pós-coordenadas, constituídas de termos-unidades linguísticas provenientes da linguagem de especialidade e da linguagem natural, denominados de descritores”. As

autoras afirmam que a pós-coordenação é uma característica que permite que os conceitos simples sejam combinados durante o processo de indexação, para especificar um assunto, e também no momento da pesquisa pelo usuário, para determinar a sua expressão de busca (Moreira, 2003).

Um tesauro apresenta termos que representam conceitos que estão relacionados segundo princípios lógicos e semânticos (Tálamo, Lara & Kobashi, 1992). Esse instrumento é um “vocabulário controlado e estruturado, no qual os conceitos são representados por termos organizados de tal forma que os relacionamentos entre conceitos ficam explícitos e os termos preferidos são acompanhados de termos sinônimos ou quase-sinônimos” (ISO, 2011, p. 12).

É possível ligar termos conceitualmente semelhantes (relações de equivalência), mais amplos e mais particulares (relações hierárquicas), relacionados e contrastados (relações associativas). Essas relações estão detalhadas a seguir.

1) Relações de equivalência

Como meio convencional para representar as relações de equivalência, em português, é utilizado o símbolo USE, indicando o descritor preferencial, e “usado por” ou UP, apontando os descritores não preferenciais. A norma 25964 (ISO, 2011) ainda orienta sobre a determinação de outros tipos de relações de equivalência, a saber: (a) Variante Ortográfica ou VO; (b) Erro de Soletração ou ES; (c) Abreviação ou AB; e (d) Forma Completa do Termo ou FCT.

As relações de equivalência ocorrem em um tesauro entre dois ou mais descritores, sendo que um deles será o descritor preferido e os demais serão os descritores não preferidos, considerados como sinônimo total ou parcial de um mesmo conceito dentro do escopo de representação do instrumento. Isso significa dizer que os descritores não preferidos não poderão ser empregados pelo indexador no momento da indexação. Porém, como há esse controle

de equivalência, no momento da recuperação o usuário poderá buscar por quaisquer descritores (preferidos ou não preferidos) que o resultado será o mesmo. Ou seja, o usuário irá recuperar os mesmos documentos.

O controle da equivalência na estrutura do tesauro minimiza as inconsistências e ambiguidades encontradas na linguagem natural, aprimorando a precisão na recuperação de informações. Svenonius (2000) declara que o problema consiste em determinar o grau de liberdade para estabelecer relações de equivalência entre dois termos, pois, muitas vezes, dois termos podem ser próximos em valor significativo, mas, na verdade, se referirem a fenômenos distintos. Sendo assim, é preciso saber, dentro do escopo e propósito de um tesauro, quando dois termos podem ser considerados sinônimos.

Podemos encontrar basicamente quatro tipos diferentes de relações de equivalência na norma ISO 25964-1 (ISO, 2011):

- **Sinonímia:** quando um mesmo conceito pode ser representado por diferentes termos descritores, sendo possível criar sinonímia contingencial ou artificial, que somente será válido dentro daquele contexto específico. Teremos graus distintos de equivalência entre termos: (a) o total, que é uma situação muito rara, em que dois termos são sinônimos absolutos, com mesma denotação e conotação, quando uma pode substituir a outra em qualquer contexto, como, por exemplo, o termo “aspirina” que é sinônimo total de “ácido acetilsalicílico”; e (b) parcial, quando dois termos são sinônimos incompletos ou limitados, sendo possível ocorrer somente em contextos específicos.
- **Oposição:** quando dois conceitos apresentam traços semânticos antagônicos, tal como acontece entre os termos “nascimento” e “morte”. Não é consenso na literatura considerar a oposição como um tipo de relação de equivalência, e *Svenonius (2000) a classifica como uma relação associativa, enquanto Tristão,*

Fachin e Alarcon (2004), como uma forma de comparação entre conceitos (ou seja, de equivalência).

- Correspondência em outro idioma: ocorre entre descritores em tesouros multilíngues, e podem ter diferentes graus de correlação entre o idioma de origem do tesouro e o idioma da tradução, sendo necessário representar a complexidade dos idiomas abrangidos pelo tesouro, respeitando-se a diversidade encontrada.
- Variações linguísticas: podem acontecer entre dois ou mais termos que representam diferentes signos variantes para um mesmo conceito, por diferenças temporais (épocas diferentes), geográficas (diferentes lugares ou etnias: mandioca ~ macaxeira), por grupo social (contexto de uso) ou por diferentes situações de comunicação (formal ou informal).

Quando estamos construindo um tesouro multilíngue, conforme esclarecem as orientações da norma ISO 25964 (ISO, 2011), algumas diferenças devem ser observadas no estabelecimento de relações de equivalência. Em tesouros monolíngues, as relações de equivalência só podem ocorrer entre termos-descritores preferenciais e não preferenciais. Assim, nesse tipo de tesouro, uma relação de equivalência não pode ocorrer entre um termo e um conceito. Já nos tesouros multilíngues, uma relação de equivalência pode ocorrer entre conceitos das diferentes línguas representadas no tesouro. Existem dois aspectos a serem considerados no estabelecimento de relações de equivalência nos tesouros multilíngues: o semântico (significado) e o cultural (correspondência interlinguística, exata ou não).

2) Relações hierárquicas

Como símbolos convencionais para representar as relações hierárquicas, em português, os símbolos convencionalmente utilizados são Termo Genérico ou TG, para indicar o termo superordenado,

e Termo Específico ou TE, para indicar o termo subordinado, para representar qualquer tipo de relação hierárquica.

As relações hierárquicas se caracterizam por princípios de superordenação e de subordinação, quando um assunto (ou objeto ou fenômeno) é subdividido em classes de nível mais gerais e mais específicos. Essas relações hierárquicas são estabelecidas com base em diferentes graus de superioridade e de subordinação (ISO, 2011). Em geral, todo conceito/descritor subordinado carrega como herança as propriedades (predicados, ação ou adjetivo) do conceito/descritor superordenado, se referindo a essa mesma natureza de conceito.

Por exemplo, Material de Construção (material) e Janelas, Portas (material), podem estabelecer relação hierárquica; por sua vez, Material de Construção (material) e Decoração de Casas (ação) não podem estar em relação hierárquica, uma vez que são conceitos de naturezas diferentes. Obedecendo a esse princípio lógico, a estrutura que compõe as relações hierárquicas também se caracteriza por princípios de coordenação, determinando um conjunto de classes e de subclasses de mesmo nível. Por exemplo, a classe de Material de Construção (material) poderia ser o superordenado que teria como subordinados um conjunto de conceitos coordenados: Janelas, Portas, Telhas, Calhas, Material Elétrico, Fechaduras, Cimentos, Argamassas, Blocos e Tijolos.

A norma ISO 25964-1 (ISO, 2011) sugere a diferenciação entre três tipos de relações hierárquicas básicas – genérica, partitiva e de instância.

- relação genérica: determina o gênero (grupo de seres ou objetos não individualizados) e suas diversas espécies, que possuem características básicas semelhantes, formando um conjunto. Confirmando as posições de Dahlberg (1979) e Svenonius (2000), Dodebei (2002, p. 92) afirma que “(...) se dois conceitos diferentes possuem características idênticas e um deles possui uma característica a mais do que o outro,

então entre eles se estabelece a relação hierárquica ou relação de gênero-espécie”. Dessa maneira, as relações hierárquicas genéricas indicam uma ligação entre um referente, que representa um conjunto, e suas espécies semelhantes.

- relação partitiva: indica a relação entre o todo (superordenado) e suas partes (subordinados). Para Dahlberg (1979), nos tesouros as relações hierárquicas partitivas comumente ocorrem na representação de objetos que podem ser segmentados, mas podem incidir também em casos especiais de divisão de uma área do conhecimento. Conforme recomendação da norma ISO 25964-1 (ISO, 2011, p. 60), essas relações “cobrem um limitado conjunto de situações nas quais uma parte de uma entidade ou sistema pertence unicamente a um todo em particular”. Muitos outros casos de relação todo-parte não podem ser considerados como ligações hierárquicas, uma vez que a parte em questão pode pertencer a mais de um todo. Esse é o caso, por exemplo, da ligação que em geral não deve ser estabelecida entre a parte roda e o todo bicicleta, tendo em vista que a roda pode também pertencer a outros tipos de meios de transporte, tais como carrinhos de mão e automóveis (ISO, 2011).
- relação de instância: indica uma ligação entre uma classe mais geral de objetos, entidades, eventos, propriedades ou fenômenos, que representa um conjunto desses elementos. Essa relação é também denominada “relação de exemplificação” ou relação “*isA*” (Lorenzon, 2011). A relação de instância se caracteriza pela exemplificação de uma classe de coisas e/ou eventos que estabelece uma referência não denotacional com os itens que estão elencados nessa classe, ou seja, é uma referência denominada de significado social que constitui diferentes sistemas linguísticos (Silva, 2006).

Na estruturação dos três tipos de relações hierárquicas em tesauros podemos empregar o rótulo de nó que, segundo a ISO 25964-1 (ISO, 2011), tem o papel de indicar uma subdivisão de uma classe de conceitos, sem ser considerado um descritor. Como as classes devem conter somente elementos de mesma natureza, o rótulo de nó funciona como um elemento de subdivisão semântica.

Também nos tesauros pode ocorrer o fenômeno da poli-hierarquia, quando um mesmo conceito (descritor) pode ter mais de um conceito (descritor) genérico, como, por exemplo, “o órgão (instrumento musical) pode ser listado sob a classe dos instrumentos de teclado e sob a classe dos instrumentos de sopro” (ISO, 2011, p. 8). Nesse caso, os atributos e relacionamentos estabelecidos para o descritor “órgão”, sobretudo os seus termos específicos TE e termos relacionados TR serão os mesmos, não importando se ele estiver ligado a instrumento de teclado ou instrumento de sopro. A poli-hierarquia também pode ocorrer em relações partitivas, tal como no exemplo: a disciplina Bioquímica que pode estar ligada tanto à disciplina Biologia e quanto à disciplina Química. Por sua vez, ainda pode ocorrer o fenômeno da poli-hierarquia quando ligamos os conceitos ossos e crânio, com base em uma relação genérica (gênero-espécie), e ainda ligamos o mesmo conceito crânio ao conceito cabeça, que tem por base uma relação partitiva.

Por sua vez, a ISO 25964-1 (ISO, 2011) recomenda utilizar símbolos que identificam os diferentes tipos de relações hierárquicas, a saber: (a) para as relações hierárquicas genéricas: Termo Genérico (Genérico) ou TGG (superordenado) e Termo Específico (Genérico) ou TEG (subordinado); (b) para as relações hierárquicas partitivas: Termo Genérico (Partitivo) ou TGP (superordenado) e Termo Específico (Partitivo) ou TEP (subordinado); e (c) para as relações hierárquicas de instância: Termo Genérico (Instância) ou TGI (superordenado) e Termo Específico (Instância) ou TEI (subordinado).

3) Relações associativas

Os símbolos convencionais empregados para representar as relações associativas nos tesouros, no idioma português, em geral, são Termo Associado ou TA, assim como Termo Relacionado ou TR.

As relações associativas ocorrem entre conceitos pertencentes a uma mesma categoria ou pertencentes a categorias diferentes. Em geral, há reciprocidade (bidirecional) nas relações associativas, porém, Soergel (1974) observa que não há essa obrigatoriedade, uma vez que esse tipo de ligação, em algumas situações especiais, pode ser unidirecional. Elas englobam “(...) ligações entre dois conceitos que não estão relacionados hierarquicamente, mas que são semanticamente ou conceitualmente próximos”, indicando ao indexador “(...) termos adicionais ou alternativos para uso no processo de indexação ou de recuperação” (ISO, 2011, p. 63). Para estabelecer a relação associativa, Svenonius (2000) recomenda verificar se um conceito é utilizado para determinar a definição de outro conceito, pois, como afirma Campos (2001, p. 113), a “relação associativa ocorre entre um conceito e uma de suas características, presente na definição”. Campos et al. (2004) afirmam que as relações associativas ligam dois conceitos para que mantenham um elo espacial ou temporal – de causa e efeito, produtor e produto, ou etapas de um processo.

De maneira semelhante, Dahlberg (1978) sugere que as relações associativas têm um papel funcional nos tesouros, do tipo “se-então”, e ocorrem entre conceitos que denotam processo, atividade e operação, em relação a um conceito que indica uma propriedade ou entidade, sendo de natureza sintagmática. Hutchins (1975) afirma que as relações associativas não são unicamente sintagmáticas (válidas em um dado contexto), sendo também paradigmáticas (válidas em quase todos os contextos). Limitar o escopo e abrangência das relações associativas é um desafio, sendo “(...) imprescindível que os critérios norteadores desse processo sejam claros e precisos o suficiente para evitar que esse tipo de relacionamento seja defi-

nido superficialmente, por exclusão ou negação (se não é relação de equivalência, nem relação hierárquica, é associativa)” (Marroni, 2006, p. 110).

No que diz respeito às aplicações dos tesouros, destacamos que quando acoplado em um sistema de recuperação da informação, a estrutura do tesouro pode ser explorada para “expansão da busca, sugestão de busca de termos alternativos, suporte para agrupamentos ou outros meios de refinar uma pesquisa, identificação de erros ortográficos comuns, suporte de indexação automática” (ISO, 2011, p. 15). Na expansão e sugestão de busca por termos alternativos, um tesouro permite acesso a seus termos equivalentes e a termos semanticamente relacionados entre si (por exemplo, o termo “acesso” pode ser expandido para termos tais como “acesso a recursos” (equivalente) ou “acessibilidade” (semanticamente relacionado)). Também o controle de equivalência possibilita mapear os termos com erros ortográficos comuns que são utilizados em um dado domínio (por exemplo, “mortandela” usado para “mortadela”), assim como as variações linguísticas e regionais (por exemplo, *a mandioca* que possui diversas denominações: aipim, macaxeira, maniva, uaipi, macamba, entre outros). Por sua vez, a estrutura hierárquica do tesouro oferece suporte para agrupamentos ou refinamento da busca, permitindo ao usuário identificar o termo mais relevante para a sua necessidade de informação. Já como suporte à indexação automática, o tesouro pode servir como base de conhecimento em sistemas inteligentes de indexação com o objetivo de otimizar o processo para recursos informacionais de grandes bases de dados, por exemplo.

2.2. Sistemas de classificação

Dahlberg (1979, p. 352) considera que a “arte de classificar, tão antiga quanto a humanidade, apenas recentemente adquiriu uma

base teórica adequada – base esta que nos permite presumir que ela progrediu do *status* de arte para o de ciência”. Os sistemas de classificação são esquemas verbais-simbólicos de recursos de informação e/ou de conhecimento que são organizados, em geral, em formato hierárquico.

Segundo a norma ISO 25964-2 (ISO, 2013), desenvolver sistemas de classificação é organizar itens de informação em classes, dividindo o universo de informações em segmentos lógicos e gerenciáveis, agrupando-os em um conjunto de conceitos que têm algum atributo em comum. É essa propriedade compartilhada que fornece à classe a sua identidade particular. Em geral, a partir de Piedade (1983), eles podem ser caracterizados por sete elementos básicos:

- Um esquema de classificação: estrutura de itens de classificação.
- Um princípio de divisão: a classe obedece a um único critério pré-estabelecido de divisão.
- Classes da classificação: item de alto nível dentro de uma classificação expressando um conceito principal.
- Definição de classes: identificação das características essenciais de uma classe, que ilustra uma clara fronteira entre ela e outras classes de classificação, podendo ser: (a) classes simples (única característica, não cabem mais subdivisões); e (b) classes compostas (mais de uma característica, e podem caber mais subdivisões).
- Item de classificação: único conceito definido, dentro de uma classe da classificação.
- Termo de classificação: designação de uma classe de classificação ou item de classificação por meio de uma expressão linguística.
- Notação: um identificador alfanumérico, e é um sistema de códigos expressando o arranjo de uma classificação.

Dentre as classificações mais elaboradas, que buscam explicitar os princípios que foram aplicados, Svenonius (2000) destaca as classificações de fundo filosófico, elaboradas por filósofos, enciclopedistas e epistemólogos, que querem dar conta da representação do conhecimento acumulado no mundo, e as do tipo bibliográfico, que estão preocupadas na organização de documentos, no arranjo para sua disposição física e na sua função de possibilitar a recuperação.

Sobre os sistemas de classificação filosófica, “(...) a primeira classificação sistemática das ciências de que temos notícia foi a de Aristóteles” (Chauí, 2003, p. 226), que aplicou três critérios para classificar os saberes: da ausência ou presença do homem nos seres investigados, da imutabilidade e da modalidade prática.

Ranganathan (1967) indica cinco características para os sistemas de classificação bibliográfica: (a) enumerativos, com tabela única (sistema de classificação da Biblioteca do Congresso); (b) quase enumerativos, com tabelas de assunto e subdivisões comuns (Classificação Decimal de Dewey – CDD); (c) quase facetados, com tabelas enumerativas, tabelas de subdivisões comuns e de subdivisões especiais (Classificação Decimal Universal – CDU; de Bliss); (d) rigidamente facetados, com assuntos básicos, tabelas de subdivisões comuns, tabelas auxiliares especiais, e determinações rígidas de sequência de combinação entre os conceitos (*Colon Classification* até 1950); e (e) analítico-sintéticos: diferem dos sistemas rigidamente facetados somente em relação às determinações rígidas de sequência de combinação entre os conceitos, que passam a ser regidas pelos princípios a ser deliberados (*Colon Classification* após 1950).

Os sistemas de classificação bibliográfica enumerativos são esquemas nos quais todas as classes possíveis são enumeradas de acordo com certas características. Existe uma abordagem *top-down*, em que uma série de classes subordinadas é produzida e onde assuntos simples e complexos são listados. A vantagem deste esquema é

que a sua estrutura é mostrada pela notação e os usuários podem encontrar facilmente as classes coordenadas e subordinadas, criando um mapa do assunto. A desvantagem é que é difícil acomodar novos assuntos e revisões frequentes podem ser necessárias.

Os sistemas de classificação bibliográfica analítico-sintéticos resolvem alguns dos problemas dos esquemas de classificação enumerativa, pois o conceito por trás deste esquema é que o assunto de um documento será dividido em seus elementos constituintes, sendo o esquema usado para encontrar as notações para cada um desses elementos, que serão então combinadas (síntese) de acordo com as regras prescritas para preparar a notação final (Piedade, 1983). Para a autora, esse esquema supera os dois principais problemas dos esquemas de classificação enumerativa, pois fornece várias tabelas, símbolos e regras notacionais específicas, que evitam a necessidade de uma longa lista de classes; por outro lado, eles também fornecem flexibilidade aos classificadores, pois números específicos podem ser construídos e o profissional não fica restrito pela disponibilidade de uma notação já pronta.

Os sistemas de classificação bibliográfica facetados listam as várias dimensões de cada assunto e fornecem um conjunto de regras para a construção de números de classes por meio do método da análise de facetas (Ranganathan, 1967). O autor esclarece que a ideia básica é que o componente de um assunto pode se encaixar em categorias pré-estabelecidas, que, no *Colon Classification* são cinco categorias fundamentais: Personalidade, Matéria, Energia, Espaço e Tempo.

Destacamos também que a norma ISO 25964-2 (ISO, 2013) traz recomendações para a construção dos sistemas de classificação de negócios, que é um modelo conceitual para o gerenciamento de atividades de negócios – que especifica o que a organização faz e como faz. É um esquema que identifica as funções de negócios, assim como as atividades e transações associadas.

Nas recomendações da ISO 25964-2 (ISO, 2013) fica estabelecido que na prática de construção de sistemas hierárquicos de classificação, os diferentes itens de um domínio podem ser definidos em diversificados grupos (conforme o propósito), a partir da escolha de determinadas características importantes e compartilhadas. Desse modo, os membros de cada grupo são semelhantes entre si, e distintos dos de outros grupos. Criam-se classes, subdivididas em subclasses, e assim sucessivamente. Por fim, chega-se a um conjunto cada vez mais particular da propriedade da classe anterior: essas são as classes subordinadas.

Sobre as aplicações de sistemas de classificação, destacamos que eles podem ser decididos localmente e representam um consenso, pois quanto maior a quantidade ou a complexidade dos itens, mais importante é empregar uma classificação que representa um consenso. Eles podem ser usados para: (a) agrupar itens fisicamente; (b) arranjar livros em uma biblioteca; (c) organizar bens de varejo em um supermercado; (d) organizar papéis em um arquivo; (e) organizar referências a objetos de informação; (f) criar diretório em um computador; (g) criar diretório da Internet; e (h) organizar sistema de páginas amarelas, entre outros.

Os sistemas de classificação facetados estão sendo amplamente utilizados na contemporaneidade, e, muitas vezes, reinterpretados em distintos contextos. Eles têm sido aplicados na classificação de objetos, em páginas estruturadas na Web, pelo e-comércio (busca e recuperação), na estruturação de assuntos de livros, na modelagem de objetos em aplicações computacionais, entre outros. Em ambiente digital eles são bem mais fáceis de aplicar, pois há facilidade de organizar a informações a partir de diferentes formas. Também são muito úteis nos catálogos, pois permitem que as distintas dimensões possam ser buscadas independentemente, facilitando a recuperação pelo usuário.

2.3. Taxonomias

Uma taxonomia é um tipo de vocabulário controlado constituído por um conjunto de termos estruturados de forma hierárquica e que representam as áreas nas quais são aplicadas. Ela surgiu na área da Biologia, com Linnaeus (Taxonomia de Lineu), no século XVIII, que criou um sistema hierárquico dos seres vivos.

A estrutura formal de uma taxonomia é representada por classes ou tipos de objetos dentro de um domínio, organizando o conhecimento ou um recurso de informações a partir de um vocabulário controlado, para facilitar o compartilhamento e a localização de informações relacionadas. Uma taxonomia deve seguir um formato hierárquico e fornecer nomes ou rótulos para cada objeto, capturando as propriedades de associação de cada objeto em relação aos outros objetos representados em sua estrutura. A construção de uma taxonomia exige rigor na especificação, garantindo que qualquer objeto recém-descoberto irá se encaixar em uma e somente uma categoria ou objeto, que irá herdar as propriedades da classe superior, mas também poderá ter propriedades adicionais.

Segundo Campos e Gomes (2008), há três tipos básicos de taxonomias: (a) canônicas (dicotômicas), (b) de domínio (policotômicas), e (c) de processos e tarefas (policotômicas), sendo que as duas últimas são determinadas a partir do contexto institucional e ao propósito do serviço, e representam a tarefa ou o domínio trabalhado em classes de conceitos com um princípio de divisão pré-definido (categorias conceituais).

Conforme descreve a norma ISO 25964-2 (ISO, 2013, p. 59), “uma taxonomia típica é apresentada como um vocabulário hierárquico, usado para classificação ou categorização, organização, consulta, navegação, busca e/ou filtragem de qualquer tipo de conteúdo em ambiente digital”. Sendo assim, na contemporaneidade ela tem sentido mais abrangente, podendo ser vista como gráficos direcionados,

com classes como nós e relacionamentos entre classes como arestas. Se o relacionamento entre tópicos e classes for ignorado, esses nós funcionarão como pontos e não terão extensão. Mas se a relação entre tópicos e classes é levada em consideração – como deve ser – as taxonomias podem ser vistas como um espaço semântico definido por vizinhanças de tópicos associados. Embora as classes sejam nós dentro de um gráfico geral do sistema, cada classe também é estruturada internamente como um gráfico.

De modo geral, Campos e Gomes (2008) descrevem como características de uma taxonomia: (a) lista estruturada de termos; (b) termos organizados hierarquicamente; (c) permite a navegação através de seus termos estruturados; (d) aceita agregação de dados; (e) explicita o modelo conceitual do domínio que representa; e (f) mecanismo de busca e recuperação de informações.

Como mapas conceituais de um domínio, as taxonomias dão orientação que é permitida pela estrutura de termos que dão acesso aos tópicos, para exploração (navegação) em um serviço de recuperação. Nesses serviços, por vezes encontramos uma taxonomia facetada, cujo conteúdo não está restrito a uma única dimensão, oferecendo diferentes opções de busca ao usuário, na qual cada termo está em uma ou mais relação tipo pai/filho (geral/específico) em relação a outro termo, isto é, estão conectados em uma hierarquia ou poli-hierarquias.

A norma ISO 25964-2 (ISO, 2013, p. 59) determina que “algumas taxonomias são customizadas para refletir a terminologia de uma organização em particular, e usadas como base para compartilhamento de conhecimento na organização”. Elas podem apoiar as atividades de negócio servindo como instrumento de comunicação e dispositivos de aprendizagem, integrando os conhecimentos e as informações para as tomadas de decisão mais assertivas.

Uma taxonomia possui aplicabilidade em: esquemas taxonômicos para atividades filosófica e científica; auxiliar na recuperação de

informações; mapear áreas emergentes de conhecimento ou atividade, ou quando os relacionamentos entre os termos ainda não podem ser completamente definidos; mapear áreas interdisciplinares; ambiente de serviço de informação digital; avaliação e administração de informações; atividades de indexação; automação parcial no processo de armazenagem de informações; navegação na Web, para orientar o usuário; entre outros. As taxonomias vêm sendo aplicadas para a gestão de informações corporativas, no e-comércio, entre outros, e o seu uso está estreitamente ligado às formas automatizadas de organização da informação (Plosker, 2005). No ambiente corporativo, as organizações usam as taxonomias, sobretudo para obter melhor qualidade de dados disponibilizados, organizar os metadados em um formato de fácil compreensão (mapa do site), gerenciar ativos de dados por meio do *data governance*, facilitar a curadoria de informações, auxiliar a aprendizagem de máquina e identificar tendências e padrões de dados.

2.4. Listas de cabeçalhos de assunto

Os cabeçalhos podem ser formados por uma ou mais de uma palavra para representar o conteúdo dos livros. A origem dos cabeçalhos de assunto pode ser associada à necessidade de se contar com outras formas para representar o assunto das publicações, visto que o título nem sempre as representava adequadamente, bem como a necessidade de subdividir os assuntos dos livros com o objetivo de torná-los mais específicos para ser usado para a indexação. Outros fatores que podem ser associados à origem dos cabeçalhos é a potencialidade dos livros para tratar de mais de um assunto, esses assuntos muitas vezes eram retratados em tempos e lugares diferentes e tais aspectos precisavam ser representados para possibilitar a sua recuperação pelo usuário. O cabeçalho de

assunto é parte de uma lista de cabeçalho de assunto, usado no momento da indexação para representar os assuntos presentes nos recursos de informação.

As listas de cabeçalho de assunto pertencem a uma modalidade de linguagem de indexação ou documentária do tipo alfabética e estruturada. São formadas por palavras simples ou compostas e que abrangem a totalidade das noções ou dos conceitos de um determinado objeto e que estão dispostas em uma ordem alfabética – que são os cabeçalhos autorizados para uso – ou também daqueles cabeçalhos não autorizados para uso na indexação. Esta situação de cabeçalhos autorizados e não autorizados corresponde ao que nos tesouros se identificam como sendo os termos preferidos ou termos não preferidos.

Guinchat e Menou (1994) caracterizaram este tipo de linguagem como combinatórias, ou seja, são pré-coordenadas e, por esse motivo oferecem a possibilidade de combinação entre os termos. São formadas por descritores independentes e contam com a presença de remissivas que são usadas para guiar o usuário para o uso de um determinado cabeçalho. De maneira geral, e seguindo um dos princípios de Cutter, os cabeçalhos de assunto estão ligados a outros através de referências cruzadas expressas como termos mais amplos e termos mais específicos, conforme estabelecido na LCSH.

Além das remissivas observou-se, também, o uso das referências, recurso importante e útil para indicar uma relação entre os assuntos e para informar ao usuário sobre a possibilidade de ampliação de sua busca para termos relacionados ao assunto procurado. Observou-se que, tal como ocorre com outro tipo de linguagem de indexação do tipo alfabética como os tesouros, as listas de cabeçalhos de assunto possuem relação do tipo hierárquica, de equivalência e de associação entre os termos. Posteriormente, como ressalta Martins (2014, p. 49), “houve uma progressiva ‘tesaurização’ das listas de

cabeçalhos de assuntos com o aparecimento das bases de dados, embora muitas bases de ciências sociais continuem a utilizar cabeçalhos de assuntos”.

Vale lembrar que o campo da organização da informação obteve ganhos significativos com os estudos de Cutter e com o impacto das suas regras para a catalogação e para a construção dos catálogos das bibliotecas. Cutter, em 1876, definiu os princípios para elaboração dos cabeçalhos de assunto e as regras para a construção dos catálogos alfabéticos que são utilizadas até os tempos atuais. De acordo com Cesarino e Pinto (1978) tais princípios podem ser resumidos em três: específico, de uso e sindético. O primeiro está relacionado à adoção de entrada direta e mais específica do cabeçalho em detrimento do uso de uma classe mais geral. Este princípio da especificidade permitiu o uso de termos mais específicos e, conseqüentemente, mais representativos de um assunto. O segundo princípio considera a importância do usuário na escolha dos cabeçalhos, isto é, a recomendação pela adoção dos termos ou de cabeçalhos reconhecidos pela maioria dos usuários, ou seja, validados pela garantia do usuário. Gomes e Marinho (1984, p. 18) ainda ressaltam que “a gramática e a sintaxe do cabeçalho de assunto foram estabelecidas, portanto, a partir da gramática (da língua inglesa)”, e o vocabulário ditado pelo uso do público. E, por último, pensando nos problemas que a ordem alfabética poderia trazer para a reunião ou separação dos assuntos relacionados, Cutter (1904) propôs a criação de uma rede de remissivas e de referências cruzadas.

Ainda sobre as características desta linguagem, as listas de cabeçalhos de assunto possuem um arranjo alfabético e os cabeçalhos são escolhidos de um dicionário organizado alfabeticamente como palavras. Outro aspecto descrito por Cesarino e Pinto (1978) diz respeito aos relacionamentos. Segundo as autoras, são linguagens não hierárquicas e o princípio da hierarquia ocorre apenas em alguns casos. Outro ponto é que este tipo de linguagem oferece pouca

possibilidade de síntese, que só é possível de ocorrer através do uso dos subcabeçalhos acrescentados aos cabeçalhos já existentes.

Os cabeçalhos de assunto constituem a linguagem alfabética mais antiga. Como exemplos desse tipo de linguagem podem ser citados: a Lista de Cabeçalho de Assunto da Biblioteca do Congresso – *Library of Congress Subject Headings* (LCSH) – criada em 1897, desenvolvida para representar os assuntos do acervo e, com isso, possibilitar o acesso à coleção; o *Medical Subject Headings* (MeSH), da área da Saúde, publicado e mantido pela National Library of Medicine (NLM), EUA.

A primeira edição da lista da Biblioteca do Congresso Americano, chamada de “Cabeçalhos de Assunto Usados nos Catálogos Dicionários da Biblioteca do Congresso” foi impressa em partes entre 1909 e 1914. O título deste instrumento foi alterado para “Cabeçalhos de Assunto da Biblioteca do Congresso” quando a oitava edição foi publicada, em 1975 (Library of Congress, n.d.). Em relação à sua estrutura, as listas de cabeçalho de assunto apresentam os cabeçalhos grafados em negrito e possuem os seguintes elementos: Componentes da entrada; Número de classe; Notas de escopo e as relações entre os assuntos, conforme descrito a seguir:

- Os componentes da entrada: os Cabeçalhos são formados por uma palavra ou mais. Os conceitos estão na forma singular e os objetos no plural. Alguns cabeçalhos contêm duas palavras geralmente formadas por um substantivo e um adjetivo que podem aparecer na ordem normal ou na ordem invertida. A política atual da Biblioteca do Congresso Americano em relação à sua Lista de Cabeçalhos de Assunto é usar a ordem normal de palavras para os assuntos, exceto para aqueles que são formados pelos seguintes elementos: idiomas, nacionalidade ou adjetivos étnicos e cabeçalhos qualificados por período de tempo, como por exemplo “Arte, medieval”. Outra característica dos cabeçalhos é o uso dos nomes geográficos associados a um assunto.

- Número de classe: é registrado para cada cabeçalho e representa o aspecto principal de cada assunto. Entretanto, para aquelas situações em que o assunto pode ser tratado de vários aspectos e estão registrados em classes diferentes, observa-se que a Lista de Cabeçalhos contém a indicação da referência do contexto de uso do cabeçalho, informação útil para o usuário compreender as situações em que o cabeçalho deve ser usado para representar os assuntos.
- Notas de escopo: são usadas para assegurar a consistência de uso de determinado cabeçalho de assunto, apontando as devidas distinções entre assuntos relacionados ou indicando a aplicação de determinado cabeçalho em um contexto específico.
- As relações entre os assuntos: as listas de cabeçalho estão estruturadas com a indicação dos relacionamentos do tipo hierárquicos, de equivalência e associativo, com o uso de operadores para explicitar tais relações e fazer o controle da ambiguidade. Assim, a relação de equivalência nas listas de cabeçalho de assunto é feita pela adoção das referências de “use” (USE, em inglês) para evidenciar a existência de cabeçalhos preferidos ou autorizados, para indicar os sinônimos, termos variantes, termos alternativos e formas anteriores. O “used for” (usado por) precede os cabeçalhos ou palavras que não devem ser usadas. Já a relação de hierarquia nas listas de cabeçalhos de assuntos é explicitada pelos termos mais gerais BT e termos mais específicos NT por meio do uso de referências cruzadas. Isso significa que em ambas as situações deve haver uma reciprocidade, ou seja, um termo que é indicado como um mais geral BT deve ser correspondido com o relacionamento invertido, um termo específico NT. Desta forma, o vértice da relação hierárquica representa a classe que um cabeçalho pertence, denotando uma supe-

rordenação enquanto um termo que faz parte desta classe é precedido pela indicação NT, termo específico, representando uma subordinação. A relação do tipo associativa é indicada pelo uso do RT ou TR, termo relacionado, conectando dois termos que denotam uma associação mental e compartilham de forte relação semântica (ISO, 2013), isto é, aquele termo que em certa medida pode ser associado a outro, é relação não hierárquica com indicação de reciprocidade entre os termos.

A combinação de vários conceitos em um único cabeçalho de assunto é facilitada pelo uso das subdivisões que são usadas ostensivamente. Como exemplos destas subdivisões podem ser citadas as de forma, que compreendem a indicação da forma como o material foi apresentado. Nestes casos, a informação sobre a forma deve ser adicionada ao último elemento de cada cabeçalho. A subdivisão geográfica cumpre o papel de indicar que um determinado cabeçalho requer a inclusão do nome de um lugar representado por uma subdivisão geográfica. Assim, como as listas de cabeçalho de assunto incluem o uso de divisão geográfica, há também a recomendação de que um determinado assunto não seja dividido geograficamente.

A LCSH já foi traduzida para diversas línguas, adaptada para uso em bibliotecas menores e é usada pela maioria das bibliotecas universitárias e por grandes bibliotecas nos Estados Unidos. Regularmente a Divisão de Políticas e Normas (PSD) recebe propostas de acréscimos e alterações que precisam ser analisadas e aprovadas pela equipe desta Divisão.

De acordo com Agustín-Lacruz, Fujita e Terra (2014), as últimas edições da LCSH apresentaram uma evolução passando a usar “a nomenclatura de relações semânticas adotadas por tesauro, incluindo uma estrutura lógica de relações hierárquicas anteriormente atrofiada na linguagem.” (Agustín-Lacruz et al., 2014, p. 94).

Da mesma forma Yi e Chan (2010) descrevem os dois níveis de estruturas hierárquicas incorporadas pela LCSH, um em nível local e outra em nível global. Essas duas estruturas são chamadas de estrutura relacional local e estrutura hierárquica global. Segundo os autores, a estrutura hierárquica de nível local de um cabeçalho de assunto refere-se a uma lista de cabeçalhos de assuntos que têm relação hierárquica com o cabeçalho em um nível imediatamente acima ou abaixo em suas posições hierárquicas (mais gerais ou mais específicas). Já a estrutura hierárquica em nível global de um cabeçalho de assunto é semelhante ao nível local, exceto que os cabeçalhos dos assuntos estão relacionados hierarquicamente em um ou mais níveis. Já uma estrutura relacional local (LRS) da LCSH compreende a manifestação de um cabeçalho principal e seu termo imediatamente relacionado mostrando relações de equivalência, relações hierárquicas e relações associativas. Por outro lado, a estrutura hierárquica global (GHS) da LCSH apresenta a exibição de um grupo de cabeçalhos de assuntos preferenciais que são hierarquicamente conectados uns aos outros, direta ou indiretamente (Yi & Chan, 2010).

A norma ISO 25964 (ISO, 2013) fornece orientações para a interoperabilidade entre linguagens. A interoperabilidade é a habilidade de dois ou mais componentes para trocar informação e usar a informação trocada. De acordo com a norma, os vocabulários podem promover a interoperabilidade pela inclusão das relações para outros vocabulários através da apresentação de dados em formato padrão e pelo uso de sistemas que suportam os protocolos comuns das máquinas, incluindo mapeamentos para outros vocabulários, estabelecendo as relações entre os conceitos de dois ou mais SOCs, apresentando dados em formatos padrão e usando sistemas como suporte.

Nesta seção discutiu-se acerca das listas de cabeçalhos de assunto e, em alguns casos, a evolução deste tipo de SOC para um tesouro

ou a tesauroização das listas de cabeçalho de assunto. A norma ISO 25964 (ISO, 2013) preserva e até amplia a padronização dos modelos estruturais de mapeamentos entre vocabulários.

As ontologias são outro tipo de sistema de organização do conhecimento abordado neste capítulo e que será apresentado a seguir.

2.5. Ontologias

Na filosofia a ontologia está relacionada com a teoria acerca dos objetos que existem no mundo e, também, com todas as sentenças que tornam verdadeiras a existência de um objeto no mundo. A ontologia pode denotar uma teoria dos objetos ou a teoria do ser. Assim, pode-se pensar que, no sentido filosófico, uma ontologia é vista como um sistema particular de categorias representando “uma certa visão do mundo” (Guarino, 1998, p. 4) ou, como afirma Smith (2003, p. 1), “é a ciência do que existe, os tipos e as estruturas dos objetos, propriedades, eventos, processos e relações em toda área da realidade”.

De acordo com Noy e McGuinness (2005), uma ontologia define um vocabulário para investigadores que necessitam compartilhar informações em um domínio. Neste sentido, os autores citam alguns motivos que levam a criação de ontologias, tais como analisar o conhecimento de um domínio, permitir a reutilização de conhecimento de um domínio e compartilhar o entendimento comum da estrutura de informação entre humanos e agentes de software.

Na Ciência da Informação as ontologias são estudadas do ponto de vista dos sistemas de informação. Neste sentido, algumas definições e características das ontologias são apresentadas a seguir. Soergel (1997) aborda as ontologias como um tipo de SOC que possibilita a representação do conhecimento. Da mesma forma, Hodge (2000) discute as ontologias como pertencentes à tipologia de modelos con-

ceituais específicos de SOCs. Ainda segundo a autora, as ontologias “podem representar relações complexas entre objetos, e incluem as regras e axiomas ausentes de redes semânticas” (Hodge, 2000, p. 7).

Para Noy e McGuinness (2005), uma ontologia é uma descrição explícita e formal de conceitos em um domínio de discurso, com a apresentação das propriedades de cada conceito, a descrição das características e dos atributos do conceito bem como as restrições sobre as propriedades.

Já para Almeida e Bax (2003) as ontologias são classes organizadas em uma taxionomia e são constituídas pelos seguintes componentes: as relações que representam tipos de interações entre conceitos, os axiomas que são utilizados para modelar sentenças que são sempre verdadeiras e as instâncias que representam elementos específicos, ou seja, representar os próprios dados.

Entretanto, além desses elementos, Silva, Souza e Almeida (2008) identificaram em Gruber (1993) e em Gómez-Pérez, Fernández e Vicente (1996) outros componentes que fazem parte das ontologias. São eles:

- (a) Classes conceituais, que organizam os conceitos de um domínio em uma taxonomia;
- (b) Atributos de classes, que são propriedades relevantes do conceito;
- (c) Instâncias, que são utilizadas para representar objetos específicos de um conceito;
- (d) Atributos de instâncias, que são propriedades relevantes que descrevem as instâncias de um conceito;
- (e) Relações entre classes, que representam o tipo de interação entre os conceitos de um domínio;
- (f) Constantes, que têm sempre o mesmo valor e, geralmente, são usadas em regras ou fórmulas para inferir conhecimento na ontologia;
- (g) Termos, que designam os conceitos de um domínio;
- (h) Axiomas formais, que restringem a interpretação e o uso dos conceitos envolvidos na ontologia;
- (i) Regras, que determinam

condições ao domínio, além de inferir valores para atributos. (Silva et al., 2008, p. 61)

A norma ISO 25.964 (ISO, 2013) apresenta os seguintes componentes semânticos presentes nas ontologias: classes, propriedades, axiomas, classes hierárquicas, indivíduos ou instâncias e axiomas. As classes apresentam um conjunto de propriedades que estabelecem os critérios para a associação a estas classes. As linguagens formais nas ontologias são usadas para expressar as propriedades que servem para definir as classes de forma explícita. Cada classe pode ser descrita por propriedades como atributos e relações entre membros de uma classe ou desta com outras classes. Os axiomas representam as declarações de uma classe. A estrutura hierárquica de uma classe é criada através de axiomas de subclasses entre classes. Desta forma a relação é explicitada através de todas as declarações axiomáticas de uma classe pai que também se aplicam à sua classe filho. Os indivíduos são referidos como objetos do discurso em um domínio. E as declarações são afirmações feitas sobre os objetos de um domínio.

O estudo de Almeida e Bax (2003, p. 9) apresenta uma possível tipologia de ontologias

(...) relacionando-as à sua função (Mizoguchi, Vanwelkenhuysen & Ikeda, 1995), ao grau de formalismo de seu vocabulário (Uschold & Gruninger, 1996), à sua aplicação (Jasper & Uschold, 1999) e à estrutura e conteúdo da conceitualização (Van-Heijst, Schreiber & Wielinga, 1997; Haav & Lubi, 2001).

Para as aplicações de ontologias, a Web Semântica se compromete em facilitar a compreensão dos agentes de software acerca dos significados dos recursos de informação. Além disso, a Web Semântica possibilita a realização de inferências dos dados a partir

da aplicação dos formalismos das representações do conhecimento nas ontologias, evitando ou reduzindo a ambiguidade e contribuindo para melhorar a recuperação da informação. Silva et al. (2008) citam alguns exemplos de linguagens para a representação de ontologias usadas para representar a conceitualização: o *Resource Description Framework* (RDF), o *Resource Description Framework Schema* (RDFS) e a *Web Ontology Language* (OWL). A linguagem RDF é usada para a escrita das declarações compostas pela tripla formadas por sujeito, predicado e objeto. De acordo com o W3C (2012), a OWL é desenhada para representar conhecimento rico e complexo sobre coisas, grupos de coisas e relações entre coisas.

As ontologias podem, ainda, ser aplicadas a muitos projetos. Almeida e Bax (2003) revelam algumas possibilidades de aplicação e afirmam que as ontologias “são utilizadas em projetos de domínios como gestão do conhecimento, comércio eletrônico, processamento de linguagens naturais, recuperação da informação na Web, de cunho educacional, entre outros” (p. 9).

Do ponto de vista da interoperabilidade das ontologias com os tesouros, a norma ISO 25964 (ISO, 2013) recomenda três abordagens: (a) a elaboração de uma reengenharia de um tesouro como uma ontologia; (b) o uso complementar de um tesouro e uma ontologia; e (c) uma abordagem híbrida das duas mencionadas anteriormente.

A primeira abordagem deve ser implementada a partir da análise e reformulação dos conceitos e relacionamentos de um tesouro para eliminar a ambiguidade prejudicial à realização das inferências necessárias pelas ontologias. O uso complementar de tesouros e ontologias pode ser mediado pelo uso de um conjunto de metadados. Segundo a norma, com esta abordagem, as definições e os atributos dos metadados deveriam ser estudados e deveria ser desenvolvido um modelo de aplicação do domínio. E, por último, em uma abordagem híbrida, os elementos de um esquema de metadados poderiam ser integrados a uma ontologia.

Tendo finalizado a descrição de diferentes tipos de vocabulários controlados, a seguir apresentamos alguns aportes vindos da Semiótica, Linguística, Lógica e Terminologia para a construção desses instrumentos.

3. Contribuições da Semiótica, da Linguística, da Lógica e da Terminologia

Várias são as perspectivas pelas quais se podem falar das disciplinas relacionadas à linguagem quanto às contribuições para a construção de vocabulários controlados. Neste texto destacaremos, entre outras, a função sógnica da linguagem documentária e os critérios de organização que dão origem aos vínculos de significação.

3.1. Linguagem documentária como signo

A ponte que liga uma base de dados referenciais a uma comunidade de usuários é a linguagem documentária, a partir da qual são representados os documentos. Por meio de enunciados condensados expressos por descritores ou por cabeçalhos de assunto, a linguagem documentária é o instrumento linguístico-comunicacional próprio dos ambientes documentário-informacionais. Já em 1966, Coyaud enfatizava a função das linguagens documentárias como “ferramentas mediatas entre usuários e autores dos documentos” (p. 127).

Sob diversas denominações – linguagens de indexação, codificações documentárias, linguagens de informação, vocabulários controlados, lista de assuntos autorizados, etc. – as linguagens documentárias são sistemas simbólicos que exercem a mediação sinalizando o conteúdo dos documentos ao mesmo tempo que

expressam seu caráter social enquanto pontes entre a linguagem do sistema de informação e a da comunidade de usuários.

Do ponto de vista semiótico, a linguagem documentária e suas unidades são signos, interpretantes de um sistema de significação que, a partir de referências de comunidades discursivas específicas, estabelece vínculos com seus públicos e permite a semiose ou interpretação. Não há vínculos com os públicos via estruturas universais de organização porque os contextos culturais não são homogêneos. É essa uma das razões pelas quais as linguagens documentárias precisam ter pontos de partida explícitos e apoiados em sistemas de significação aceitos por suas comunidades de discurso. Essa é a garantia para transformar estoques informacionais em fluxos: cada linguagem documentária funciona a partir de uma hipótese de organização. Tálamo (2001), seguindo Pêcheux, mostra que formação discursiva e formação social estão associadas, razão pela qual a transmissão da informação faz parte dos efeitos de sentido previstos no funcionamento da linguagem. A formação discursiva diz respeito ao que se pode dizer no interior de uma dada formação ideológica. A recepção nos ambientes documentário-informacionais, portanto, depende da capacidade da linguagem documentária em intermediar os discursos.

Em resumo, para que a linguagem documentária exerça o papel de instrumento de comunicação em ambientes documentários há que existir referências comuns ou princípios de compartilhamento: só assim os signos documentários podem funcionar como elos entre os públicos e as bases de dados informacionais.

3.2. Critérios de organização

Uma das principais características dos princípios de organização do conhecimento é o uso de redes de relações entre os termos. Tais

redes não são uma invenção recente, nem um produto da Ciência da Informação que estabelecem conexões entre conceitos, estando presentes na Linguística, na Lógica, na Terminologia.

Os elos que unem os conceitos são de natureza semântica e expressos sob a forma de termos, uma vez que os conceitos não existem independentemente da linguagem. Para a organização do conhecimento, o elemento semântico é de fundamental importância, pois o significado das unidades linguísticas é um dos principais meios para conduzir o acesso aos documentos e à informação. Nos sistemas formalizados, quando se busca a interoperabilidade, ou seja, quando se deseja prover a comunicação entre sistemas informatizados, o elemento semântico faz toda a diferença, porque permite conexões formais mais potentes do que as viabilizadas por coincidência de caracteres, tal como ocorre em HTML¹. Eles são fundamentais para permitir que a Web funcione a partir de dados estruturados, dados ligados entre si, combinados sintática e morfológicamente, seguindo a tripla: sujeito, objeto, predicado.

A construção das redes semânticas obedece a critérios para o arranjo ou agrupamento dos termos partindo da pressuposição de que a compreensão e a interpretação são adesivas, isto é, pressupõem vínculos (Baitello, 1994) baseados em conhecimentos anteriores. É a existência de vínculos que garante a comunicação, razão pela qual a definição de critérios para a organização do conhecimento deve levar em conta o conhecimento compartilhado que, combinado aos objetivos do sistema de informação, garantem o fluxo entre a linguagem do sistema e a dos públicos da informação.

¹ *Hypertext Markup Language* (HTML), em português Linguagem de Marcação de Hipertexto.

Os critérios de organização podem ser de natureza lógica, linguístico-comunicacional, terminológica e pragmática ou sua combinação, sendo por isso mesmo denominadas redes lógico-semânticas, modo utilizado por J. C. Gardin e equipe para falar das linguagens documentárias (Cros, Gardin, & Lévy, 1968). A diversidade dos instrumentos de organização do conhecimento advém dos diferentes princípios de organização, temas, objetivos e necessidades de classificação, mas sua arquitetura ou estrutura é consequência dos critérios que respondem pela determinação de seus pontos de partida. Os critérios podem variar, mas a formação dos vértices segue uma dada definição, seja ela implícita ou explícita.

3.2.1. Critérios lógico-semânticos

Do ponto de vista lógico, os elos entre os termos têm como base a semelhança entre os conceitos ou a associação entre eles. Os vínculos baseados na semelhança pressupõem uma definição anterior estável, amplamente aceita, a partir da qual é possível deduzir outros conceitos; de forma inversa, pode-se chegar aos conceitos por indução.

As categorias aristotélicas são baseadas em vínculos estabelecidos por semelhança: gozam de estabilidade e reconhecimento. Exemplo clássico é a Árvore de Porfírio, cujas subdivisões seguem o princípio dicotômico que postula a diferenciação baseada na inclusão ou exclusão de uma qualidade: A e não A. A Árvore de Porfírio pressupõe relações de implicação entre gênero/espécie e todo/parte: "(...) se uma substância for racional, será necessariamente animada; se for animada, será necessariamente corpórea e, se corpórea, será substância" (Eco, 1984, p. 46). Tais subordinações são determinadas pela escolha das qualidades (ou a definição) que orientam a hierarquia, implicando as possíveis deduções dela derivadas (Figura 1).

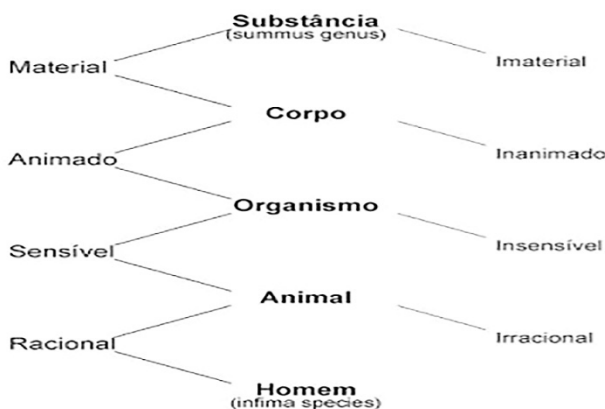


Figura 1: Árvore de Porfírio. Fonte: Burke, 2003, p. 82.

Nas linguagens documentárias, as hierarquias dicotômicas são menos comuns, porque na escolha das categorias são combinados princípios de compartilhamento e objetivos pragmáticos: ao critério de semelhança se associa o de funcionalidade local para contemplar funções previstas para o sistema. As linguagens documentárias constituem, assim, objetos culturais que representam escolhas e, ao mesmo tempo, determinam o modo de organização da informação.

Os conceitos podem também ser reunidos por aproximação no espaço e/ou no tempo, a exemplo das conexões causa-efeito ou das ligações produtor-produto, disciplina-seu objeto, operação-instrumento, ação-resultado da ação, ação-objeto da ação, caracterizadas como conceitos que se relacionam ontologicamente. As relações lógicas se dão por abstração; as relações ontológicas, no âmbito da realidade empírica.

Os critérios lógicos dão origem à superordenação e à subordinação, ou seja, a relações hierárquicas onde há herança de propriedades entre os conceitos: as espécies herdaram características do gênero, sucessivamente. Na perspectiva das relações ontológicas são expressas as relações partitivas (todo/parte) e as relações

sequenciais (contiguidade espaço-temporal): entre o todo e suas partes, embora a apresentação possa ser em árvore, não há herança de propriedades.

Nas hierarquias (lógicas ou ontológicas), os pontos de partida ou vértices são fundamentais para determinar as espécies, os tipos e as partes. Outro modo de dizer que as definições determinam a forma de organização.

3.2.2. Critérios linguístico-comunicacionais

Os critérios linguístico-comunicacionais têm como base o fato de que a significação é contextual, institucional. Tomando como referência a língua como um sistema de relações, conforme a linguística saussuriana, e considerando que sistema equivale à estrutura, verifica-se que a significação se dá a partir de relações de oposição: uma palavra significa tudo aquilo que as outras não significam, ou: uma palavra tem simultaneamente uma definição negativa e uma definição positiva. A significação é consequência das relações instituídas entre unidades linguísticas no interior da linguagem. Decorre daí que um só termo não comporta significação, pois a significação pressupõe necessariamente da existência da relação (cada termo se opõe a outros termos). Por exemplo,

Estrela:

Angelina Jolie é a estrela do filme *Malévola*.

O sol é a estrela central do Sistema Solar.

Banco:

O Banco do Brasil abriu uma nova agência.

Há um banco de pedra em frente àquela casa.

Nos exemplos acima, há duas definições de estrela: atriz do cinema e astro do sistema solar. Do mesmo modo, no segundo exemplo: banco, instituição financeira, e banco, artefato para sentar.

A proposta de sistema de Saussure foi aprofundada por Greimas (1966) na apresentação da estrutura elementar de significação. Para ele, a significação tem de ser focalizada a partir das descontinuidades, no plano da percepção. Perceber é apreender diferenças.

1. Perceber diferenças quer dizer captar ao menos dois termos-objetos como simultaneamente presentes.

2. Perceber diferenças quer dizer captar a relação entre os termos, ligá-los de um ou de outro modo. (Greimas, 1966, p. 28)

Disso decorre que,

1. Para que dois termos-objetos possam ser captados juntos é preciso que tenham algo em comum (é o problema da semelhança e, em suas extensões, o da identidade).

2. Para que dois termos-objetos possam ser distinguidos, é preciso que sejam diferentes, qualquer que seja a forma (é o problema da diferença e da não-identidade).

(...) De fato, a relação manifesta (...) sua dupla natureza: é ao mesmo tempo *disjuntiva* e *conjuntiva*. (Greimas, 1966, p. 29, grifo do autor)

O exemplo abaixo mostra a conjunção e a disjunção.

/veículo/

barco

automóvel

avião

Em /veículo/ se manifesta o aspecto conjuntivo: todos os termos subsumidos são os tipos veículo. Em /barco/, /automóvel/, /avião/... são apresentadas as disjunções: os veículos distinguem-se entre si por características particularizantes.

Nos tesouros, instrumentos que mais desempenham o papel de linguagem controlada no âmbito da Ciência da Informação, a conjunção/disjunção é um recurso estruturante muito utilizado. De certo modo, ele substitui uma definição usando a regra gênero próximo, diferença específica.

Exemplo²:

Arma

TE Arma branca

TE Arma de arremesso

TE Arma de choque

TE Arma de fogo

Arma branca

TG ARMA

TE Arpão

TE Espada

TE Punhal

De forma correspondente, na ótica da Lógica, a definição por gênero próximo e diferença específica, proposta por Aristóteles, é uma definição do tipo intensional, onde:

- O gênero, que é uma definição dada anteriormente, constitui parte de uma nova definição dela derivada.

² Fonte: Tesouro de Folclore e Cultura Popular Brasileira. (n.d.). Rio de Janeiro: Instituto do Patrimônio Histórico e Artístico Nacional. Recuperado de <http://www.cnfcp.gov.br/tesouro/>

- A diferença inclui o *definiendum*, ou o que está sendo definido, dentro do gênero a partir da qualidade que tem em comum com ele, distinguindo-se dele mediante uma qualidade particular.

3.2.3. Critérios pragmáticos

Os critérios pragmáticos são aqueles definidos considerando a função do sistema de informação, os objetivos da linguagem documentária e, principalmente, as necessidades dos usuários. Têm como base o compartilhamento de princípios de ordenação. Podemos afirmar que eles dependem da observação colateral³ (Peirce, 1977), ou seja, das referências culturais ou dos sujeitos que estão envolvidos, do seu nível de conhecimento e, em especial, de seu conhecimento prévio.

No exemplo abaixo⁴, as denominações de áreas de atividade gozam de relativo reconhecimento, permitindo o acesso aos tipos de cursos técnicos oferecidos.

Cursos técnicos SENAC

Cursos por área

Arquitetura e urbanismo

Beleza e estética

Comunicação e artes

Design

³ “Por observação colateral entendo uma prévia familiaridade com aquilo que o signo denota. Assim, se o Signo for a sentença ‘Hamlet era louco’, para compreender o que isto significa deve-se saber que, às vezes, os homens ficam nesse estado estranho: deve-se ter visto homens loucos ou deve-se ter lido sobre eles; e será melhor se se souber especificamente (...) qual era a noção que Shakespeare tinha da insanidade” (Peirce, 1977, p. 161).

⁴ Fonte: Serviço Nacional de Aprendizagem Comercial (n.d.). *Cursos técnicos*. Copyright 2007-2019. Recuperado de <http://www.sp.senac.br/jsp/default.jsp?template=1489.dwt&testeira=474&type=T&sub=2>

O critério pragmático diz respeito à relevância de determinados itens (traços semânticos, princípios de divisão) para realizar agrupamentos, considerando os valores das comunidades discursivas particulares. Ele tem como base a observação dos aspectos que melhor reúnem, para um determinado grupo social, um grupo de conceitos.

Se do ponto de vista da Terminologia clássica, de Wüster, a reunião dos conceitos se dá por características consideradas “essenciais”, na ótica da pragmática e da Terminologia Comunicativa as relações entre os conceitos são contextuais, ou seja, dependem do uso, da forma como se manifestam na linguagem, das “formas de vida” (para o II Wittgenstein), ou das referências das comunidades discursivas, segundo a Linguística Aplicada e a Terminologia contemporânea.

3.2.4. Critérios terminológicos

Os critérios terminológicos dizem respeito às segmentações ou recortes feitos no universo da língua pelas linguagens de especialidade, ou *Language for Special Purposes* (LSP), que dizem respeito a formas ou meios de expressão linguísticas (terminologia, fraseologia própria, traços estilísticos ou sintáticos) e também a um sistema conceitual próprio e um modo de organização do domínio visado.

Os conceitos não são necessariamente ligados a línguas particulares, mas são, entretanto, “submetidos à influência do contexto sociocultural que conduz frequentemente a categorizações diferentes”⁵ (ISO, 2000, p. 2). Assim, os critérios terminológicos se reportam ao modo como um determinado sistema de conceitos é recortado

⁵ “Concepts are not necessarily bound to particular languages. They are, however, influenced by the social or cultural background which often leads to different categorizations” (ISO, 2000, p. 2).

pela linguagem e reconhecido por uma comunidade discursiva. Os sistemas de conceitos representam um domínio organizado onde cada conceito se define relativamente a outros. Eles podem incluir variações de concepção e de designação oriundas das diferentes formas de organizar os objetos. De qualquer modo, é sempre possível harmonizar as diferenças entre descritores que os representam nas linguagens documentárias estabelecendo relações de equivalência entre eles, conforme sugere a norma de construção de tesouros e vocabulários, ISO 25964-1-2 (ISO, 2011, 2013), quer seja na mesma língua ou entre línguas diferentes.

Os critérios terminológicos fornecem referenciais interpretativos para os descritores, constituindo-se, assim, nos vínculos que permitem ligar a linguagem dos sistemas documentários e a de seus públicos.

Ex. definições de árvore em domínios diferentes:

Árvore (Linguística)

Árvore [Tree] – Um tipo particular de representação gráfica da estrutura de uma sentença. A maioria dos linguistas acredita que a estrutura de qualquer sentença é uma estrutura de constituintes, em que a sentença é constituída por algumas peças, e cada peça é constituída por partes menores, e assim sucessivamente até as partes menores de todas; além disso, cada peça pertence a alguma categoria sintética específica. Este tipo de estrutura pode ser representado de maneira eficaz pelo recurso gráfico das árvores⁶.

⁶ Fonte: Trask, R. L. (2004). *Dicionário de linguagem e linguística*. São Paulo: Contexto.

Árvore (Botânica)

Árvore [Arbor]. Planta lenhosa, geralmente superior a 5 metros de altura, com caule (tronco) indiviso até certa distância do solo, dividindo-se então em maior ou menor número de ramos⁷.

Árvore (Informática)

Uma árvore é uma coleção de nós, que são os elementos que contêm a informação que se pretende armazenar e de arestas que ligam os nós⁸.

As definições apresentadas constituem referências para os descritores.

4. Considerações finais

Neste capítulo apresentamos o conceito do termo “vocabulário controlado” que serve de “guarda-chuva” para abrigar diferentes tipos de instrumentos. Eles têm como princípio comum o controle do vocabulário utilizado em um dado domínio, com o intuito de permitir a organização do conhecimento e dos recursos de informação, tendo em vista facilitar seu gerenciamento e recuperação. Foram descritos os instrumentos: tesouros, sistemas de classificação, taxonomias, listas de cabeçalhos de assuntos e ontologias, que são desenvolvidos em distintos graus de representação e formalização de suas estruturas conceituais, para atender a diversos propósitos, funções e comunidades específicas.

⁷ Fonte: Fernandes, R. B. (2007). *Glossário de termos botânicos*. Coimbra: Universidade de Coimbra. Recuperado de https://www.uc.pt/herbario_digital/learn_botany/glossario

⁸ Fonte: Portal da Educação. (n.d.). *Árvores*. Recuperado de <https://www.portaleducacao.com.br/conteudo/artigos/informatica/arvores/67750>

No papel de sistemas de representação essas ferramentas são modelos que instrumentalizam a atividade de tornar visíveis os recursos informacionais e o conhecimento de um domínio. Tentamos evidenciar que os sistemas de classificação, as taxonomias e as ontologias são hospitaleiros em ambiente digital, uma vez que permitem agregação de dados em sua estrutura (metadados) de diferentes tipos, admitindo a marcação de documentos, o que garante que mecanismos de busca encontrem as informações desejadas, organizando-as, ordenando-as e disseminando-as. Por sua vez e pela sua natureza, os tesouros e as listas de cabeçalhos de assunto nem sempre estão atrelados a sistemas informatizados, sendo muitas vezes utilizados como documentos referenciais.

Também em comum podemos afirmar que todos os instrumentos de representação do conhecimento e da informação referidos possuem, como uma espinha dorsal, estruturas hierárquicas cuja origem lógica está em Aristóteles e Porfírio, que entenderam a representação da natureza como uma dicotomia, na qual há um objeto ou conjunto de coisas que pode ser subdividido em partes ou tipos. Nesse processo, há diversas regras de associação e distinção ordenadas e sistematizadas que constituem uma hierarquia.

Foram também expostas as contribuições da Semiótica, da Linguística, da Lógica e da Terminologia na construção de vocabulários controlados, sobretudo no que se refere à função sónica da linguagem documentária e aos elementos a serem considerados na organização que determinam os vínculos de significação entre eles e seus públicos.

Referências Bibliográficas

Agustín-Lacruz, M. C., Fujita, M. S. L., & Terra, A. L. S. (2014). Linguagens documentais para as bibliotecas escolares: O caso da Espanha, Portugal e Brasil. *Informação*

- & Sociedade: Estudos, 24(3), 83-97. Recuperado de <http://www.periodicos.ufpb.br/index.php/ies/article/view/20256>
- Almeida, M. B., & Bax, M. P. (2003). Uma visão geral sobre ontologias: Pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, 32(3), 7-20. Recuperado de <http://revista.ibict.br/ciinf/article/view/984/1023>
- Austin, D., & Dale, P. (1981). *Guidelines for the establishment and development of monolingual thesauri* (2nd ed., text rev.). Paris: Unesco.
- Austin, D., & Dale, P. (1993). *Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües* (B. A. Meto, trad.). Brasília: IBICT/SENAI.
- Baitello, N. (1994). Sociedade da informação. *São Paulo em Perspectiva*, 8(4), 19-21. Recuperado de http://produtos.seade.gov.br/produtos/spp/v08n04/v08n04_03.pdf
- Barbosa, M. A. (1995). Contribuição ao estudo de aspectos da tipologia de obras lexicográficas. *Ciência da Informação*, 24(3), 1-9. Recuperado de <http://revista.ibict.br/ciinf/article/view/572%20>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43. Recuperado de <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>
- Bezerra, S. M. F. (2013). *Variação linguística retratada nas canções de Luiz Gonzaga* (Trabalho de Conclusão de Curso não publicado). Universidade Federal da Paraíba, Campina Grande, Brasil.
- Bocato, V. R. C., & Fujita, M. S. L. (2006). Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. *Perspectivas em Ciência da Informação*, 11(2), 267-281. doi:10.1590/S1413-99362006000200010
- Bocato, V. R. C., Ramalho, R. A. S., & Fujita, M. S. L. (2008). A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. *Ibersid: Revista de Sistemas de Información y Documentación*, 2, 199-209. Recuperado de <https://ibersid.eu/ojs/index.php/ibersid/article/view/2235>
- British Standards Institution. (1987). *BS 5723: Guidelines for the establishment and development of monolingual thesauri*. London: Author.
- British Standards Institution. (2008). *BS 8723-5: Structured vocabularies for information retrieval: Guide interoperation between vocabularies and other components of information storage and retrieval systems*. London: Author.
- Burke, P. (2003). *Uma história social do conhecimento: De Gutemberg a Diderot*. Rio de Janeiro: Zahar.
- Campos, M. L. A. (2001). *Linguagem documentária: Teorias que fundamentam sua elaboração*. Rio de Janeiro: EdUFF.
- Campos, M. L. A., & Gomes, H. E. (2008). Taxonomia e classificação: O princípio de categorização. *DataGramaZero: Revista de Ciência da Informação*, 9(4). Recuperado de <https://brapci.inf.br/index.php/article/view/0000007750>
- Campos, M. L. A., Gomes, H. E., & Motta, D. F. (2004). *Manual de elaboração de tesouro*. Rio de Janeiro: BITI. Recuperado de <http://www.conexaorio.com/bitit/tesouro/index.htm>

- Cesarino, M. A. da N., & Pinto, M. C. M. F. (1978). Cabeçalho de assunto como linguagem de indexação. *Revista da Escola de Biblioteconomia da UFMG*, 7(2), 268-288. Recuperado de <http://hdl.handle.net/20.500.11959/brapci/71210>
- Chauí, M. (2003). A Filosofia como vocação para a liberdade. *Estudos Avançados*, 17(49), 7-15. doi:10.1590/S0103-40142003000300002
- Cintra, A. M. M., Tálamo, M. F. G. M., Lara, M. L. G. de, & Kobashi, N. Y. (2002). *Para entender as linguagens documentárias* (2ª ed.). São Paulo: Polis.
- Coyaud, M. (1966). *Introduction à l'étude des langages documentaires*. Paris: C. Klincksieck.
- Cros, R. C., Gardin, J.-C., & Lévy, F. (1968). *L'automatisation des recherches documentaires: Un modèle général, LE SYNTOL* (2nd ed., revue et augmentée). Paris: Gauthier-Villars.
- Cunha, M. B., & Cavalcanti, C. R. O. (2008). *Dicionário de Biblioteconomia e Arquivologia*. Brasília: Briquet de Lemos.
- Cutter, C. A. (1904). *Rules for a Dictionary Catalog* (4th ed.). Washington, DC: Government Printing Office.
- Dahlberg, I. (1978). A referent-oriented, analytical concept theory of interconcept. *International Classification*, 5(3), 142-151. doi:10.5771/0943-7444-1978-3-142
- Dahlberg, I. (1979). *Teoria da classificação, ontem e hoje* (H. B. Cox, trad.). Paper apresentado na Conferência Brasileira de Classificação Bibliográfica, Rio de Janeiro, Brasil. Recuperado de http://www.conexaorio.com/biti/dahlbergteoria/dahlberg_teoriam.htm
- Dodebei, V. (2002). *Tesouro: Linguagem de representação da memória documentária*. Niterói: Intertexto.
- Eco, U. (1984). *Conceito de texto*. São Paulo: EdUSP.
- Fujita, M. S. L., Santos, L. B. P., & Alves, R. V. (2018). Linguagens de indexação e linguagem documentária são sistemas de organização do conhecimento? Uma análise bardiana da variação terminológica. *Scire*, 24(2), 23-33. Recuperado de <https://www.ibersid.eu/ojs/index.php/scire/article/view/4577>
- Gardin, N. (1969). Le lexique intermédiaire: Un nouveau pas vers la cooperation internationale dans le domaine de l'information scientifique et technique. *Bulletin de l'UNESCO: à l'Intention des bibliothèques*, 23(2), 66-71.
- Gomes, H. E., & Marinho, M. T. (1984). Introdução ao estudo do cabeçalho de assunto. *BITI: Biblioteconomia, Informação & Tecnologia da Informação*. Recuperado de http://www.conexaorio.com/biticabecalho/cab_ass.htm
- Gómez-Pérez, A., Fernández, M., & Vicente, A. J. (1996). *Towards a method to conceptualize domain ontologies*. Paper presented at the Workshop on Ontological Engineering, Budapest, Hungary.
- Greimas, A. J. (1966). *Sémantique structurale: recherche de méthode*. Paris: Larousse.
- Grings, L. (2015). Controle de autoridades na Biblioteca Nacional do Brasil: breve histórico e práticas atuais. *Revista Brasileira de Biblioteconomia e Documentação*, 11(2), 139-154. Recuperado de <https://rbbd.febab.org.br/rbbd/article/view/433>
- Gruber, T. (1993). *What is an ontology?* Recuperado de <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

- Guarino, N. (Ed.). (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), *Formal Ontology in Information Systems: Proceedings of FOIS'98, Trento, Italy* (pp. 3-15). Amsterdam: IOS Press. Recuperado de <https://klevas.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKonceptinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>
- Guinchat, C., & Menou, M. (1994). *Introdução geral às ciências e técnicas da informação e documentação* (2ª ed.). Brasília: IBICT.
- Haav, H. M., & Lubi, T. L. (2001). A survey of concept-based information retrieval tools on the web. In A. Caplinskas & J. Eder (Eds.), *5th East-European Conference ADBIS' 2001: Professional Communications and Reports. Vol. 2. Advances in Databases and Information Systems* (pp. 1-11). Vilnius: Vilnius Gediminas Technical University. Recuperado de <https://www.ioc.ee/~helemai/HaavLubiADBIS2001.pdf>
- Hedden, H. (2016). *The accidental taxonomist* (2nd ed). Medford, NJ: Information Today, Inc.
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Washington, DC: The Digital Library Federation, Council on Library and Information Resources.
- Hutchins, W. J. (1975). *Languages of indexing and classification: A linguistic study of structures and functions*. Stevenage: Peter Peregrinus.
- Instituto Brasileiro de Informação em Ciência e Tecnologia. (1984). *Diretrizes para elaboração de tesouros monolíngues*. Brasília: Autor.
- International Federation of Library Associations and Institutions. (2005). *Guidelines for Multilingual Thesauri. Working Group on Guidelines for Multilingual Thesauri Classification and Indexing Section*. Den Haag, Netherlands: Author.
- International Organization for Standardization. (1986). *ISO 2788: British standard guide to establishment and development of monolingual thesauri*. Geneve: Author.
- International Organization for Standardization. (2000). *ISO 1087-1: Terminology work: Vocabulary. Part 1: Theory and application/Travaux terminologiques: Vocabulaire, Partie 1: Théorie and application*. Geneve: Author.
- International Organization for Standardization. (2011). *ISO 25964: Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneve: Author.
- International Organization for Standardization. (2013). *ISO 25964: Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneve: Author.
- Jasper, R., & Uschold, M. (1999). *A framework for understanding and classifying ontology applications*. Paper presented at IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends and Sixteenth International Joint Conference on Artificial Intelligence, Stockholm, Sweden.
- Lancaster, F. W. (1986). *Vocabulary control for information retrieval* (2nd ed.). Arlington, TX: Information Resources.
- Lancaster, F. W. (1993). *Indexação e resumos: Teoria e prática* (A. A. Briquet de Lemos, trans.). Brasília: Briquet de Lemos.

- Lara, M. L. G. de (2000). *A terminologia (e as terminologias) e a documentação*. In M. Correia (Org.), *Terminologia e indústrias das línguas: actas do VII Simpósio Ibero-Americano de Terminologia* (pp. 593-602). Lisboa: Iltec/Riterm. Recuperado de <https://drive.google.com/file/d/110xeBbiGPKqa94nESCnKlFVVCMeXd5Ud/view>
- Library of Congress. (n.d.). *Introduction*. Recuperado de <https://www.loc.gov/aba/publications/FreeLCSH/LCSH40%20Main%20intro.pdf>
- Lorenzon, E. J. (2011). *Análise de domínio para avaliação de tesouros: Uma experiência com a cadeia produtiva do calçado no Brasil* (Tese de doutorado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, Brasil). Recuperado de <https://repositorio.unesp.br/handle/11449/103385>
- Maculan, B. C. M. S. (2015). *Estudo e aplicação de metodologia para reengenheria de tesouro: Remodelagem do Thesagro* (Tese de doutorado em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil). Recuperado de <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUBD-9ZKMUV>
- Marroni, G. N. B. (2006). *Identificação e delimitação de relações associativas em tesouros: Um estudo de caso na área do direito do trabalho* (Dissertação de mestrado em Ciência da Informação, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, DF, Brasil). Recuperado de <http://repositorio.unb.br/handle/10482/4919>
- Martins, M. C. B. (2014). *Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: O sistema no Instituto Politécnico de Portalegre*. Salamanca: Ediciones Universidad de Salamanca.
- Melton, J. (1964). *A use for the techniques of structural linguistics in documentation research*. Cleveland, OH: Western Reserve University.
- Mizoguchi, R., Vanwelkenhuysen, J., & Ikeda, M. (1995). Task ontology for reuse of problem solving knowledge. In N. J. I. Mars (Ed.), *Proceedings of ECAI'94: towards very large knowledge bases* (pp. 46-59). Amsterdam: IOS Press.
- Moreira, A. (2003). *Tesouros e ontologias: Estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético* (Dissertação de mestrado em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil). Recuperado de <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/LHLS-69UQKU>
- National Information Standards Organization. (n.d.). *ISO 25964: The international standard for thesauri and interoperability with other vocabularies*. Baltimore, MD: Author. Recuperado de <https://www.niso.org/schemas/iso25964>
- National Information Standards Organization. (R2010). *ANSI/NISO Z39.19-2005: Guidelines for the construction, format and management of monolingual controlled vocabularies*. Baltimore, MD: Author. Recuperado de <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>
- Noy, N. F., & McGuinness, D. L. (2005). *Desarrollo de ontologías – 101: Guía para crear tu primera ontología*. Stanford, CA: Author. Recuperado de <http://ocw.uc3m.es/ingenieria-informatica/sistemas-avanzados-de-recuperacion-de-informacion/ejercicios/ontology101-es.pdf/view>

- Oliveira, R. M. V. B. (2002). *Web Semântica: novo desafio para os profissionais da informação*. Paper apresentado no XII Seminário Nacional de Bibliotecas Universitárias, Recife, PE, Brasil. (Publicado em CD-ROM)
- Peirce, C. S. (1977). *Semiótica* (J. T. Coelho Netto, trad., J. Guinsburg, rev.). São Paulo: Perspectiva.
- Piedade, M. A. R. (1983). *Introdução à teoria da classificação* (2ª ed.). Rio de Janeiro: Interciência.
- Plosker, G. (2005). Taxonomies: Facts and opportunities for information professionals. *ABI/Inform Global*, 1(29), 58-60.
- Ranganathan, S. R. (1967). *Prolegomena to library classification*. Bombay, NY: Asia Publishing House.
- Rey-Debove, J. (1984). *Léxico e dicionário* (C. B. de Moraes, trad.). *Alfa*, 28, 45-69.
- Salgado, D. M. (2015). *O controle de autoridade sob a norma RDA: Análise da aplicação e implicações na construção de registros de autoridade* (Dissertação de mestrado em Ciência da Informação, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, Brasil). Recuperado de <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-19012016-105319/pt-br.php>
- Silva, A. S. (2006). Sociolinguística cognitiva e o estudo da convergência/divergência entre o português europeu e o português brasileiro. *Veredas – Revista de Estudos Linguísticos*, 10(1/2). Recuperado de <https://periodicos.ufjf.br/index.php/veredas/article/view/25230>
- Silva, D. L., Souza, R. R., & Almeida, M. B. (2008). Ontologias e vocabulários controlados: Comparação de metodologias para construção. *Ciência da Informação*, 37(3), 60-75. Recuperado de <http://revista.ibict.br/ciinf/article/view/1204>
- Smith, B. (2003). *Ontology and Informations Systems*. Buffalo, NY: State University of New York (SUNY), Dept. of Philosophy. Recuperado de <https://pdfs.semanticscholar.org/bf79/a0e32882ac9f8b49774f23396b248aee70fa.pdf>
- Soergel, D. (1974). *Indexing languages and thesauri: Construction and maintenance*. Los Angeles, CA: Melville.
- Soergel, D. (1997). *Functions of a thesaurus, classification and ontological knowledge bases*. [S.l.]: College of Library and Information Services, University of Maryland. Recuperado de <http://users.umiacs.umd.edu/~oard/teaching/878/spring99/readings/functclass.pdf>
- Svenonius, E. (2000). *The intellectual foundations of information organization*. Cambridge, MA: The MIT Press.
- Tálamo, M. de F. G. M. (2001). Terminologia e documentação. *Tradterm*, 7, 141-151. Recuperado de <http://www.revistas.usp.br/tradterm/issue/view/4127>
- Tálamo, M. de F. G. M., Lara, M. L. G. de, & Kobashi, N. Y. (1992). Contribuição da terminologia para a elaboração de tesouros. *Ciência da Informação*, 21(3), 197-200. Recuperado de <http://revista.ibict.br/ciinf/article/view/432/432>
- Tristão, A. M. D., Fachin, G. R. B., & Alarcon, O. E. (2004). Sistema de classificação facetada e tesouros: Instrumentos para organização do conhecimento. *Ciência da Informação*, 33(2), 161-171. Recuperado de <http://revista.ibict.br/ciinf/article/view/1058>

- Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-115. doi:10.1017/S0269888900007797
- Van Heijst, G., Schreiber, A. T., & Wielinga, B. J. (1997). Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46(2/3), 183-292. doi:10.1006/ijhc.1996.0090
- Vilela, M. (1997). O léxico do português: Perspectivação geral. *Filologia e Lingüística Portuguesa*, 1, 31-50.
- W3C. (2012). *Web Ontology Language (OWL)*. Recuperado de <https://www.w3.org/2001/sw/wiki/OWL>
- Yi, K., & Chan, L. M. (2010). Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. *Journal of the American Society for Information Science and Technology*, 61(4), 677-687. doi:10.1002/asi.21295

CAPÍTULO 7
APLICAÇÕES TECNOLÓGICAS EM REPRESENTAÇÃO
E ORGANIZAÇÃO DA INFORMAÇÃO
TECHNOLOGY APPLICATIONS IN INFORMATION
REPRESENTATION AND INFORMATION
ORGANIZATION

Ana Carolina Simionato Arakaki
Universidade Federal de São Carlos (UFSCar)
acsimionato@ufscar.br
ORCID: 0000-0002-0140-9110

José Eduardo Santarém Segundo
Universidade de São Paulo (USP)
santarem@usp.br
ORCID: 0000-0003-3360-7872

Silvana Aparecida Borsetti Gregorio Vidotti
Universidade Estadual Paulista (UNESP)
silvana.vidotti@unesp.br
ORCID: 0000-0002-4216-0374

Resumo: As aplicações tecnológicas no campo da Organização e Representação da Informação vêm possibilitando inovações às práticas e

teorias já consolidadas e ainda, colaboram para a formalização de estruturas de acesso, uso e reúso da informação. Dessa forma, este capítulo busca refletir sobre os fundamentos da representação e organização da informação em conjunto com as novas abordagens para os ambientes informacionais digitais. Foi adotada uma pesquisa de natureza teórica, de caráter qualitativa e exploratória. Como resultados, apresentam-se o impacto, as contribuições e as tendências das tecnologias na teoria e nos instrumentos do campo estudado. Além disso, é visto que a Ciência da Informação tem construído uma nova abordagem de pesquisa utilizando-se das tecnologias para se (re)programar, como exemplo dos temas abordados: Ontologias, SPARQL, FRBR, BIBFRAME, XML, RDF, OWL, *Linked Data*, *Dublin Core*. Portanto, considera-se que essa visão pode contribuir para uma convergência dos catálogos, da interoperabilidade de dados e dos modelos de recuperação baseados em *Linked Data* e fontes diversas de dados, construindo novos questionamentos para pesquisas e dinâmicas de se pensar em modelos de disseminação e acesso à informação.

Palavras-Chave: Organização da Informação. Representação da informação. Tecnologia e Informação.

Abstract: The technology applications in the field of organization and representation of information have enabled innovations to the practices and theories already consolidated and collaborate to the formalization of structures of access, use and reuse of information. In this way, this chapter aims to reflect on the fundamentals of the representation and organization of information in conjunction with the new approaches to digital informational environments. A theoretical research of qualitative and exploratory nature was adopted. The results present the impact, contributions and trends of technologies in the theory and tools of the studied field. Moreover, it can be seen that Information Science has built a new approach to research using the technologies to (re)program, as an example of the topics addressed: ontologies, SPARQL, FRBR, BIBFRAME, XML, RDF, OWL, *Linked Data*, *Dublin Core*. Therefore, it was thought that this view could contribute to a convergence of catalogs, data interoperability

and data-based recovery models and diverse data sources of *Linked Data*, giving rise to new questions for research and dynamics of thinking about models of dissemination and access to information.

Keywords: Information organization. Information representation. Technology and Information.

1. Introdução

O acesso aos recursos informacionais no ambiente digital é uma tarefa quase que instantânea em sistemas de informação, sendo que as condições necessárias para esse acesso decorrem das mais variáveis formas de tecnologia disponíveis e podem ser fundamentadas pelos conceitos da organização e representação da informação.

Nesse sentido, a organização está presente em diversas ações da vida em sociedade, como desde a organização sistêmica da atividade corpórea dos seres vivos, que é uma organização natural, até a organização funcional e administrativa de uma instituição. Conforme Rendon Rojas e Herrera Delgado (2010, p. 15, tradução nossa) afirmam, a organização da informação é “[...] um mundo como todo homem, cheio de sentido, mas sentidos de segunda ordem, ou seja, do mundo natural, mas sentidos de um mundo cultural que por sua vez é interpretado”.

Em relação aos recursos informacionais, além da recuperação da informação, a comunidade de usuários percebe outros interesses que condizem à representação da informação e que exigem o acesso a fragmentos de informações de certos documentos, a identificação do significado de certos dados, e o relacionamento entre objetos ou mesmo o uso de um vocabulário comum entre usuários e sistemas de recuperação de informação.

Na Ciência da Informação, a organização da informação é considerada como

[...] um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. Alguns tipos de representação da informação são construídos por meio de linguagens elaboradas especificamente para os objetivos da OI [Organização da Informação]. (Medeiros & Café, 2008, p. 5).

Logo, o escopo da organização da informação é disseminar os objetos de informação para os usuários, de forma que possibilite “[...] unir coisas semelhantes e diferenciar entre elas” (Svenonius, 2009, p. xi, tradução nossa). Essas ações envolvem as duas dimensões destacadas por Medeiros & Café (2008) que consistem nos processos descritivo e temático. A primeira dimensão, descritiva, trata dos elementos da forma dos documentos, possibilitando identificar os atributos dos objetos informacionais. A segunda dimensão, temática, corresponde ao conteúdo informacional, com foco na análise, identificação e representação do assunto sujeito a esses objetos.

Para que esse escopo seja contemplado, cada uma das dimensões, descritiva e temática, possui uma segmentação teórica com processos, métodos, padrões e instrumentos que permitem a identificação e a integração dos dados, tendo em vista o processo final, a recuperação da informação por meio dos atributos disponíveis, bem como os aspectos que envolvem a identificação do conteúdo de cada atributo.

A partir desse parâmetro, as Tecnologias de Informação e Comunicação (TIC) são utilizadas com as abordagens e as práticas já conhecidas no campo de Organização e Representação da Informação, possibilitando uma inovação às práticas já consolidadas e viabilizando as estruturas de acesso, uso e reúso da informação. Almeida e Crippa (2009, p. 110) afirmam que “organizar a informação

tornou-se o problema de nossos tempos”. Para os autores, a inserção dos novos recursos tecnológicos produziu alterações na forma de execução do trabalho, bem como nos processos relacionados, tornando-se urgente a necessidade de reestruturação e reconstrução da organização da informação, colocando-a em consonância com esse novo cenário.

Nessa perspectiva, explora-se nesse capítulo, discutir sobre os fundamentos da representação e organização da informação em conjunto com as novas abordagens para o ambiente informacional digital.

Em relação aos procedimentos metodológicos para o desenvolvimento da pesquisa, foi adotada a pesquisa de natureza teórica, de caráter qualitativa e exploratória, objetivando analisar as contribuições das tecnologias para a organização e representação da informação.

Dessa forma, na seção 2, são apresentadas algumas tecnologias, em especial da Web Semântica, como tendência na representação e organização da informação no ambiente Web. Na seção 3 é apresentado o impacto das tecnologias nos instrumentos de representação e organização da informação.

2. Tecnologias da Web Semântica

A estruturação semântica e sintática dos dados proposta em 2001 por Berners-Lee, Hendler, e Lassila, ficou conhecida como Web Semântica. A ideia era que os dados fossem utilizados para realizar descobertas, ampliar integração entre sistemas e reusados por meio de várias aplicações.

A Web Semântica é caracterizada por melhorias no tratamento informacional do conteúdo digital, bem como por proporcionar ao usuário a busca, localização e recuperação mais eficientes e sim-

plificadas do conteúdo que procuram. Foi definida por Berners-Lee et al. (2001) uma extensão da Web, de modo a tornar as máquinas capazes de realizar inferências e apresentar resultados mais significativos e semânticos aos usuários humanos e não humanos.

Como foi também denominada de Web 3.0, a Web Semântica é vista como um composto de tecnologias que se complementam e atribuem um valor a cada relacionamento. A proposta estrutural relacionada ao conteúdo significativo de páginas Web, cria um ambiente onde agentes possam realizar tarefas facilmente de uma página para outra e cada vez mais sofisticadas para os usuários (Berners-Lee et al., 2001).

Para isso, Berners-Lee et al. (2001) descrevem que os sistemas de conhecimento de representação dos dados normalmente são centralizados. Dessa forma, é exigido um compartilhamento com a mesma definição de conceitos comuns, podendo ser definidos e ligados de uma forma legível por máquinas.

A Web Semântica deve-se preocupar com os usuários, a descentralização e a abertura dos dados, para que as tecnologias atreladas sejam eficientes e a interação seja um meio universal para o intercâmbio de dados, de informações e de conhecimento.

Berners-Lee et al. (2001) projetaram a Web Semântica em uma arquitetura em camadas, conhecida como bolo de noiva, que possui os níveis de desenvolvimento estrutural, sintático, semântico, ontológico e lógico. Cada uma dessas camadas apresenta as tecnologias apropriadas para atender às necessidades de compartilhamento entre os agentes em diferentes aplicações de domínio, propiciando a interpretação dos conteúdos disponíveis na Web e com resultados mais eficientes (Alves, 2005; Berners-Lee et al., 2001; Catarino & Souza, 2012; Ferreira & Santos, 2013; Isotani & Bittencourt, 2015; Ramalho, 2006; Santos Neto et al., 2013; Silva & Santos, 2012; Souza & Alvarenga, 2004).

Como exemplo de aplicação no contexto da Web Semântica, um motor de busca pode identificar se a expressão de busca ‘rosa’ se refere à coloração ‘rosa’, ou à botânica ‘*Rosoideae*’, ou ainda, ao nome próprio de uma pessoa. Neste sentido, motores de busca como o *Wolfram Alpha* (<https://www.wolframalpha.com/>) revitalizam competências entre máquinas e usuários (Santarém Segundo & Simionato, 2016).

Santarém Segundo (2014, p. 3864) afirma que as tecnologias da Web Semântica “[...] estão diretamente relacionadas ao processo de construção da informação e ao armazenamento, constituindo assim ambientes que possam ter conjunto de dados ligados semanticamente.”

Berners-Lee (2006) definiu quatro princípios que servem como sugestões de estruturação de dados ligados. Esses princípios são:

- I. Use URIs como nomes para as coisas;
- II. Use HTTP URIs para que as pessoas possam procurar esses nomes;
- III. Quando alguém procurar um URI, forneça informações úteis, usando os padrões (RDF, SPARQL);
- IV. Inclua links para outros URIs, para que eles possam descobrir mais coisas.

Considerando a disponibilização dos dados sob licença aberta, Berners-Lee (2006) utilizou um sistema que classifica em até cinco estrelas a disponibilização de *Linked Open Data*:

1 estrela: dados disponíveis na Web em qualquer formato;

2 estrelas: dados estruturados legíveis por máquina, ainda que em formato proprietário (por exemplo, em *excel* em vez de digitalização de uma imagem de uma tabela);

3 estrelas: dados estruturados legíveis por máquina em formato não proprietário (por exemplo, *csv*);

4 estrelas: além dos requisitos das estrelas anteriores, utiliza padrões abertos da W3C como RDF e SPARQL para identificar as coisas, de modo que as pessoas possam relacionar seu material;

5 estrelas: todas as demais estrelas e link de seus dados para dados de outras pessoas para fornecer contexto.

De acordo com Nowack (2009), a base do *Linked Data* está apoiada pelas tecnologias das camadas da arquitetura da Web Semântica. Entre as camadas discutidas por Santos e Alves (2009) estão as características internacionais, camada sintática, camada de dados e camada ontológica.

A camada de características internacionais é a base de toda estrutura da Web e conseqüentemente da Web Semântica. Entre as tecnologias que fundamentam essa camada estão: *Uniform Resource Identifier* (URI) e *Unicode*.

O *Unicode* é um padrão que permite aos computadores representar e manipular, de forma consistente, texto de qualquer sistema de escrita existente. Já o URI é uma sequência de cadeia de caracteres de forma única de nomear e identificar itens. Um URI pode ser um *Uniform Resource Locators* (URL) ou um URN. O URL apresenta informações do local em que o recurso pode ser obtido.

O URN fornece a identificação por meio do nome de um recurso. Como alternativa do URL ou URN, foi criado o *Persistent Uniform Resource Locators* (PURL). O PURL é uma iniciativa da *Online Computer Library Center* (OCLC) que gera um identificador único e persistente que garante a identificação do recurso, mesmo que o endereço da Web mude ao longo do tempo (Online Computer Library Center [OCLC], 2016). Entre os principais URI para identificação de recursos, pode-se citar: DOI utilizado para identificar documentos digitais; *Handle*, utilizado em repositórios digitais para identificação da publicação; *International Standard Book Number* (ISBN), um número internacional normalizado para livros;

International Standard Serial Number (ISSN), utilizado para identificar publicações periódicas; *Archival Resource Key* (ARK), criado no contexto da curadoria digital que abarca arquivos, bibliotecas e museus; *Open Researcher and Contributor ID* (ORCID), para identificação de pessoas; *International Standard Name Identifier* (ISNI), padronizado pela ISO 27729 para identificar nomes de pessoas, produtores, editores, agregadores (Arakaki, 2016).

A camada sintática está atrelada às linguagens de marcação. Usualmente é utilizada a *eXtensible Markup Language* (XML), que é “[...] conjunto de regras sintáticas para a criação, desenvolvimento e utilização de linguagens de marcação.” (Ferreira & Santos, 2013, p. 42). Entretanto, outras linguagens mais simplificadas já existem e facilitam o processo de interoperabilidade entre as informações compartilhadas.

A camada de dados apresenta como tecnologia o *Resource Description Framework* (RDF). Ele possibilita relacionamentos por meio da tríade: recurso, propriedade e valor. De acordo com Santarém Segundo (2014), um recurso é uma informação que pode ser identificada por um URI ou conjunto de caracteres, as propriedades são elementos ou atributos que representam as características do recurso e os valores são os dados que representam a informação a ser descrita. A partir desse grafo (recurso, propriedade e valor), o RDF estrutura dados na Web e as descrições de recursos em *Linked Data*, por meio de uma arquitetura que define ligações explícitas entre dados dos recursos e a fonte dos dados.

Já a camada ontológica abarca tecnologias como RDF *Schema*, OWL e SKOS. O RDF *Schema* possibilita a criação de hierarquias, classes, propriedades, subclasses e subpropriedades. O *Simple Knowledge Organization System* (SKOS) é um modelo de dados utilizado para expressar a estrutura básica de um sistema de organização do conhecimento de vocabulários no ambiente da Web.

A *Web Ontology Language* (OWL) consiste em uma linguagem para definir e instanciar Ontologias na Web.

Atrelado à camada de dados e à camada ontológica está o SPARQL. O SPARQL é uma linguagem que proporciona consultas semânticas em bancos de dados capaz de consultar e manipular dados publicados em RDF (Santarém Segundo & Coneglian, 2016).

3. Impactos das tecnologias na organização e representação da informação

Os ambientes digitais proporcionaram impactos significativos na definição e evolução dos conceitos e tecnologias. Nesse sentido, os instrumentos de representação e de organização da informação estão se adequando às novas estruturas e ferramentas tecnológicas e aos novos ambientes digitais.

Fomentando essa discussão, o grupo de trabalho da W3C, o *W3C Library Linked Data Incubator Group*, desenvolveu um relatório discutindo os impactos tecnológicos para bibliotecas (Baker et al., 2011; Arakaki, 2016) e levantaram cinco pontos fundamentais:

1. Apesar das bibliotecas gerarem dados, não há uma relação com outras fontes de dados na Web;
2. Os padrões de bibliotecas são criados apenas para o universo bibliográfico, por exemplo, o formato MARC;
3. Não há prática de uso de identificadores para representar as informações para os dados produzidos em bibliotecas;
4. A Web Semântica e a Biblioteconomia e Ciência da Informação possuem conceitos semelhantes para terminologias diferentes. Assim, a compreensão mútua deveria ser incentivada;
5. Entretanto, as grandes mudanças tecnológicas dependem ainda dos fornecedores dos sistemas.

Diante dessas observações, pode-se destacar a necessidade de repensar e reestruturar as práticas e instrumentos de representação frente as ferramentas tecnológicas, principalmente quando são abordadas, por exemplo: novas tipologias de documentos, novas formas de acesso, novos suportes para armazenamento, além da abertura dos dados das bibliotecas, arquivos e museus.

A preocupação dos eixos da Arquivologia, Biblioteconomia e Museologia em oferecer o acesso, uso e reúso aos recursos informacionais faz com que os profissionais da área se apropriem dos princípios de organização e representação da informação para desenvolverem instrumentos para os mais diversos tipos de recursos informacionais. Logo, esses princípios são fundamentais para que a informação seja ordenada, estruturada ou inserida de alguma forma, caso contrário, ela permanecerá amorfa e inutilizável (McGarry, 1999).

Nesse sentido, a representação se refere ao conjunto de convenções sintáticas e semânticas que tornam possível descrever as coisas, de uma forma intrínseca e extrínseca, contudo, necessita de padrões para formalizar e adaptar a apresentação prevista e proposta ao usuário, no sentido de aperfeiçoar suas buscas, acesso e localização, isto é, de forma mais formal, sistemática e institucional (Glushko, 2014; Simionato, 2017).

A representação da informação é uma ação complexa e Marcondes (2001, p. 63) elucida que:

Representar objetos e conceitos na mente, manipular estas representações antes de agir na realidade de acordo com os resultados da manipulação mental prévia e socializar estas experiências, sob a forma de representações orais ou escritas, para outros membros da espécie, é uma das mais fundamentais características do gênero humano.

Por esse motivo, os instrumentos de representação da informação são auxílios funcionais na tradução da complexidade do recurso informacional para as estruturas documentais e que irão caracterizar a principal forma de comunicação entre o profissional e o usuário. Santos e Pereira (2014, p. 8) detalham que essas estruturas documentais denotam de uma

[...] estrutura de complexidade que a conduz para um tratamento específico, a partir de formas de representação e do uso das tecnologias disponíveis para a persistência de dados e o instanciamento de registros descritivos que garantirão acesso às informações e aos recursos disponíveis.

Nessa estrutura, que oferece a persistência de dados e o instanciamento, está a desconstrução de um registro bibliográfico pela representação abstrata e não mais de um item, como isso vem sendo realizado há anos na Biblioteconomia e Ciência da Informação. Assim, os modelos conceituais propiciam a reestruturação do registro monolítico para uma estrutura de relações, com relacionamentos entre entidades e atributos.

Os modelos conceituais tiveram impactos significativos, principalmente nas estruturas dos instrumentos de descrição. As ISBDs (*International Standard Bibliographic Description*), que tiveram suas publicações iniciais individualizadas para cada tipo de material na década de 1960, foram reunidas em um único instrumento de descrição em 2011, denominada de ISBD consolidada.

Baseada principalmente na proposta do *Functional Requirements for Bibliographic Records* (FRBR), a ISBD focou na descrição da entidade de manifestação de uma obra, ainda baseada nas áreas descritivas. Por conta da unificação de todas as ISBDs, foi necessário ainda a criação de uma nova área descritiva (área 0), para descrição da forma do conteúdo e do tipo de mídia.

Paralelo a essas mudanças, a AACR2r (*Anglo-American Cataloguing Rules*) também sofreu alterações em sua estrutura. Quando o grupo gestor da AACR2 começou os estudos de atualização da AACR3 (3ª edição), observou-se que, a partir da proposta do FRBR, não seria possível manter a mesma estrutura que, assim como a ISBD, é estruturada por capítulos que estão vinculados à tipologia dos materiais descritos. Devido às alterações da estrutura, optou-se por alterar o nome para *Resource Description and Access* (RDA).

A RDA foi projetada para o ambiente digital, com escopo que extrapola as bibliotecas, e diferente da AACR2r, a estrutura da RDA foi criada em Partes (I e II), que contém capítulos (cada tipo de material tem um capítulo), áreas, elementos e regras. Já na RDA, não há mais a divisão por tipo de material: ela está estruturada em Partes, que estão relacionadas a atributos e relacionamentos, seções, capítulos, elementos e regras.

Com as novas propostas (modelos conceituais e RDA), começa-se a questionar a estrutura em que estava sendo utilizadas, descritas e compartilhadas as informações pelo formato *Machine Readable Cataloging* (MARC21).

Embora seja amplamente utilizado no ambiente de bibliotecas, o padrão MARC 21 apresenta dificuldades quanto ao seu uso no ambiente *Web*, principalmente no que se refere à sintaxe necessária para compartilhar seus registros e promover a interoperabilidade, devido a estrutura de seu esquema de metadados. (Arakaki et al., 2017, p. 3).

A partir dessa evolução dos registros, baseada em padrões internacionalmente reconhecidos, ressalta-se a principal preocupação em multidimensionar as possibilidades de acesso, uso e reúso dos dados bibliográficos. Sendo essa relação entre usuário e o catálogo, é denominada como '*user tasks*', descritas por Bertolini et al. (2016,

p. 8, grifo nosso) na Declaração dos Princípios Internacionais de Catalogação:

[...] **Encontrar** recursos bibliográficos numa coleção como resultado de uma pesquisa, utilizando atributos e relações entre recursos [...] Identificar um recurso bibliográfico ou agente [...] **Selecionar** um recurso bibliográfico que seja apropriado às necessidades do utilizador (usuário) [...] **Adquirir** ou **obter** acesso a um item descrito [...] ou adquirir um item por meio de compra, empréstimo, etc. [...] **Navegar** num catálogo ou para além dele (quer dizer, através da organização lógica dos dados bibliográficos e de autoridade [...]).

Essas tarefas que o usuário deve usufruir do catálogo são o início para o desenvolvimento do *Functional Requirements for Bibliographic Records* (FRBR), publicado em 1998. Após o desenvolvimento do FRBR, outras abstrações foram conceituadas, isto é, para dados de autoridade foi publicado, em 2009, o *Functional Requirements for Authority Data* (FRAD) e, em 2010, para dados de assunto, o *Functional Requirements for Subject Authority Data* (FRSAD).

As funções dos modelos FRBR, FRAD e FRSAD ressaltam a necessidade da evolução dos registros nos catálogos, refletindo as entidades e atributos em uma nova configuração da representação, como conceituado por Svenonius (2009, p. 71): “O princípio da representação requer que as descrições bibliográficas sejam construídas para refletir as entidades do meio bibliográfico e representar elas mesmas.” Para que isso aconteça, a abstração para um modelo deve ser definida pelos seus dados como ‘coisas’ (ou entidades) com as quais outras ‘coisas’ funcionam e como eles se relacionam um com o outro de forma superior (Coyle, 2016).

O *Functional Requirements of Authority Data* (FRAD) publicado em 2009, define os requisitos funcionais para os registros de auto-

ridade. Os dados de autoridade representam os pontos de acesso controlados e outra informação de que as instituições utilizam para coletar obras de uma determinada pessoa, família, organização, edições corporativas ou múltiplas de um título.

O *Functional Requirements for Subject Authority Data* (FRSAD) propõe os relacionamentos entre dados para autoridade de assunto ligadas aos recursos bibliográficos e são disponibilizados para os usuários. O núcleo do FRSAD é definido por uma obra que tem como assunto ou é assunto de um thema, ou vice-versa, e um thema que tem como definição ou é definido por um nomen, ou vice-versa.

O FRSAD apresenta características da modelagem entidade-relacionamento e orientado a objetos, apresentando como entidades: thema, uma superclasse de todas as entidades FRBR, ou seja, uma superentidade que permite a modelagem de relacionamentos e atributos em um nível mais geral e abstrato; e nomen, definido como qualquer sinal ou sequência de sinais (caracteres alfanuméricos, símbolos, som, etc.).

Como já descrito por Riva, Le Bœuf, & Žumer (2017), os três modelos são necessários para completude do sistema bibliográfico; no entanto, todos foram criados em uma estrutura de modelagem de entidade-relacionamento e adotaram diferentes pontos de vista e soluções diferentes para problemas comuns. Dessa forma, os três modelos apresentados foram conjugados no modelo de referência conceitual denominado de FRBR *Library Reference Model* (FRBR-LRM) que visa a abrangência para os dados bibliográficos, como entendido em um sentido amplo e geral.

O modelo FRBR-LRM define uma única entidade de nível superior, nomeada como a RES – ‘coisa’, uma generalização da entidade do FRSAD, além das necessidades da relação de assunto. Entre as entidades que constituem de obra, expressão, manifestação, item, agente, nome, lugar, evento, pessoa, entidade coletiva, família e

organização, esta última dividida em tipos de grupos e identificador do tipo de nome.

A experiência anterior na criação de vocabulários da IFLA para os modelos conceituais da família FR indicou a estrutura do LRM como um documento altamente estruturado, por exemplo, em a tarefa de especificar *namespaces* para uso com aplicativos *Linked Data*, para que seja mais fácil e reduzir as possíveis ambiguidades.

O contexto mudou desde que o modelo FRBR foi originalmente desenvolvido e novas necessidades surgiram, particularmente em termos de reutilização de dados em aplicações, fazendo desta consideração uma parte integrante do planejamento inicial da apresentação de a definição do modelo.

Dessa forma, buscando representar desde a abstração ou ideia para construção de um livro até a materialização, foi proposto o *Bibliographic Framework Initiative* (BIBFRAME) baseado nas tecnologias da Web Semântica e dos modelos conceituais. O BIBFRAME apresenta um modelo de dados baseado na estrutura da família FR, que contempla entidades como Obra e Expressão do FRBR, para entidade Obra no BIBFRAME. A Obra no BIBFRAME corresponde à abstração do recurso descrito. A entidade Manifestação do FRBR, corresponde à entidade Instância do BIBFRAME e abarca a materialização da Obra. Por fim, Item do FRBR, corresponde à entidade Item no BIBFRAME, que abrange informações individuais do recurso informacional (Arakaki et al., 2017).

A partir dessa estrutura do BIBFRAME (Obra, Instância e Item), foi desenvolvido um vocabulário estruturado em classes, subclasses, propriedades e subpropriedades para descrever os recursos informacionais.

Paralelo ao BIBFRAME, outros padrões também aderiram ao movimento da Web Semântica como foi o caso do *Dublin Core* e do *Schema.org*. O *Dublin Core* surgiu na década de 1990, com o propósito de localizar recursos na Web e, no contexto da Ciência da

Informação, foi amplamente adotado principalmente em repositórios digitais. Em quase 25 anos de criação, com uma estrutura simples e flexível, o *Dublin Core* teve grandes influências das tecnologias da Web Semântica, o que proporcionou a manutenção e relevância do padrão para o desenvolvimento de diversos outros padrões (Arakaki, Alves, & Santos, 2018).

Já o *Schema.org* surge da iniciativa de buscadores como *Google* e *Yahoo*, para criar um padrão que facilite a identificação de recursos no ambiente da Web. Tem sido utilizado em diversos contextos, em especial, promove a ligação de dados no *WorldCat*, que é uma base de dados gerenciada pela OCLC.

Há diversos outros padrões que auxiliam na descrição de recursos no ambiente digital e que adotam práticas e ferramentas das tecnologias da Web Semântica, além do BIBFRAME, o *Dublin Core* e o *Schema.org* que cada vez mais estão mais próximos dos conceitos de Ontologias, no que diz respeito a proporcionar as inferências das informações prestadas em consultas mais avançadas com SPARQL.

Diante desse cenário, os vocabulários controlados já adotaram o SKOS para representar seus vocabulários no contexto da Web Semântica. O SKOS tem propriedades que estruturam o vocabulário controlado como: termo preferido; termo alternativo; termo oculto; termo genérico; termo específico; termo relacionado; nota de escopo; definição; exemplo; e nota histórica. Alguns exemplos que utilizam o SKOS são: *The Food and Agriculture Organization Thesaurus* (AGROVOC), um vocabulário controlado nas áreas de Agricultura, Pesca, Alimentação e áreas correlatas e o *UK Archival Thesaurus*, um tesouro utilizado na área de arquivos do Reino Unido (Catarino, 2014).

Além das concepções da descrição temática, é importante ressaltar a relação da organização e representação da informação como elementos fundamentais da Arquitetura da Informação. Coneglian, Roa-Martínez, Ferreira, Vidotti, & Santarém Segundo (2019) e Vidotti,

Coneglian, Roa-Martínez, Vecchiato, & Santarém Segundo (2019) realizam uma abordagem das possibilidades de uso das tecnologias no contexto da Arquitetura da Informação. Assim, no Sistema de Organização, um dos elementos da Arquitetura da Informação, por exemplo, para a estrutura dos conteúdos informacionais, pode ser utilizadas as estruturas da XML, o RDF, o RDF *Schema*, e as ontologias construídas em OWL, SPARQL e SWRL.

Dessa forma, Coneglian et al. (2019) exemplificam o Sistema de Organização como forma de aplicação de tecnologias da Web Semântica no *Knowledge Graph*: “O *Knowledge Graph* é uma aplicação de busca do Google que possibilita a pesquisa por um determinado conjunto de caracteres, ou seja, a busca por entidades.” (Coneglian et al., 2019, p. 28). Nessa perspectiva, os referidos autores afirmam que a

[...] incorporação das tecnologias RDF, OWL e SPARQL no Sistema de Organização permitirá a identificação do significado dos conteúdos informacionais e do perfil de cada usuário, e propiciará a personalização do ambiente para cada tipo de indivíduo, para cada tipo de informação, e/ou para cada informação especificamente (p. 28).

Já o Sistema de Rotulagem da Arquitetura da Informação, pode utilizar tecnologias como XML, RDF, OWL, Microdados e RDFa, que podem auxiliar na representação dos conteúdos informacionais por meio de rótulos. Os autores exemplificam o Sistema de Rotulagem com aplicações da Web Semântica com o *Rich Snippets*. O *Rich Snippets* é uma descrição enriquecida que contém informação adicional para os usuários (Coneglian et al., 2019).

No Sistema de Navegação, o uso do *Linked Data* é fundamental, pois na apresentação dos resultados, pode-se apresentar as ligações e relacionamentos de um recurso com outros recursos, com o

apoio de tecnologias como RDF, OWL e SPARQL. Como exemplo, foi abordado o projeto *LodLive*, que explora o uso da aplicação do *Linked Data* para procurar recursos de RDF (Coneglian et al., 2019).

No Sistema Busca, são utilizadas tecnologias como OWL, RDF, RDF *Schema*, SPARQL e SWRL. Para explicação da busca em uma aplicação, Coneglian, Roa-Martínez, Ferreira, Vidotti, & Santarém Segundo (2019) utilizam o mecanismo de busca *Explorer*.

Por fim, no Sistema de Representação, estão relacionados vocabulários controlados, tesouros, taxonomias, *folksonomias*, ontologias e metadados, que podem utilizar as tecnologias como o RDF, a OWL, a XML e o SKOS. Para discutir a aplicação das tecnologias no Sistema de Representação, Coneglian et al. (2019) exemplificam com a *Europeana*, que é um ambiente digital que reúne objetos culturais de bibliotecas, arquivos, museus e galerias espalhadas pela Europa.

4. Reflexões finais

Há um conjunto de recursos, ferramentas e objetos citados neste texto que não existiam antes da Internet, antes dos modelos de conexão em rede, antes do novo contexto tecnológico em que o mundo se encontra atualmente. Todas essas novidades têm impactado diretamente os modos de construir ciência.

A Ciência da Informação tem construído uma nova abordagem de pesquisa utilizando-se das tecnologias para (re)programar o que vem sendo e o que será desenvolvido na área nos próximos anos. Neste texto é possível notar que temas como: Ontologias, SPARQL, FRBR, BIBFRAME, XML, RDF, OWL, *Linked Data*, *Dublin Core* e tantos outros recursos tecnológicos, já são parte da rotina de quem estuda e trabalha na área.

Essas aplicações tecnológicas estão convergindo para o campo da organização e representação da informação, principalmente em

tendências para os catálogos, para a interoperabilidade de dados, para os novos modelos de recuperação baseados em *Linked Data* e fontes diversas de dados, construindo novas perguntas de pesquisa, novas dinâmicas de se pensar modelos de acesso ao conhecimento.

É certo que os processos e as ferramentas tecnológicas têm possibilitado novas perspectivas e oportunidades no âmbito da Ciência da Informação, alterando paradigmas tanto do ponto de vista de seus processos como de seus objetos.

Ao observar o que temos neste texto é possível afirmar que há um reposicionamento da área de Ciência da Informação, baseado principalmente nas Tecnologias de Informação e Comunicação que a amparam, e pensando mais especificamente na representação da informação, vemos uma substancial mudança de paradigma – a abordagem inicial feita a partir do item sobrepõe-se para uma nova estruturação realizada pela origem intelectual ou artística de um recurso informacional.

Nesse momento, é importante que os esforços e iniciativas sejam comuns, para que o modelo FRBR-LRM tenha a mesma aplicação e comunicação entre os grupos responsáveis pelo BIBFRAME e RDA. Assim, temas comuns sobre interoperabilidade e qualidade dos dados poderão ser discutidos com maior eloquência e profundidade. Da mesma forma, é importante que esses instrumentos de representação possam dialogar com os princípios da Web Semântica, no caso o *Linked Data* e *Linked Open Data*, para que corporifique realmente a identificação e formalização entre os recursos disponíveis em catálogos e suas possibilidades de recursos na Web.

Considera-se que a integração de dados bibliográficos, e também a integração de dados patrimônios culturais na iniciativa *Linking Open Data*, bem como os esforços voltados especificamente para a publicação de dados, justificam-se pela criação de novos serviços, pela construção de parcerias entre instituições (que não precisam

necessariamente de pertencer ao mesmo país), mas principalmente pela sustentabilidade que pode ser alcançada com dados abertos e disponíveis para acesso, uso e reúso.

Pensando neste novo modelo de dados, talvez este texto possa trazer uma nova pergunta a área e também ao pensar da representação da informação: “Sendo as fontes de dados ligados os recursos emergentes para acesso ao conhecimento e as bibliotecas, com seus catálogos, a estrutura de confiabilidade e qualidade para fornecer informação, não seria a união destes contextos, em forma de *Linked Data*, um novo paradigma para constituir silos importantes de conhecimento?”.

Referências Bibliográficas

- Almeida, M. A. de, & Crippa, G. (2009). De Bacon à Internet: Considerações sobre a organização do conhecimento e a constituição da Ciência da Informação. *PontodeAcesso*, 3(2), 109-131. doi:10.9771/1981-6766rpa.v3i2.3284
- Alves, R. C. V. (2005). *Web semântica: Uma análise focada no uso de metadados* (Dissertação de mestrado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo, Brasil). Recuperado de <http://hdl.handle.net/11449/93690>
- Arakaki, F. A. (2016). *Linked Data: Ligação de dados bibliográficos* (Dissertação de mestrado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo, Brasil). Recuperado de <https://repositorio.unesp.br/handle/11449/147979>
- Arakaki, F. A., Alves, R. C. V., & Santos, P. L. V. A. da C. (2018). Dublin Core: State of art (1995 to 2015). *Informação & Sociedade: Estudos*, 28(2), 7-20. doi:10.22478/ufpb.1809-4783.2018v28n2.38012
- Arakaki, F. A., Galeffi, L. F., Alves, R. C. V., & Santos, P. L. V. A. da C. (2017). BIBFRAME: Tendência para a representação bibliográfica na web. *RBBB. Revista Brasileira de Biblioteconomia e Documentação*, 13, 2231-2249. Recuperado de <https://rbbd.febab.org.br/rbbd/article/view/995>
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., ... Zeng, M. (2011). Library Linked Data Incubator Group Final Report. Cambridge, MA: W3C. Recuperado 4 de dezembro de 2016, de <https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>
- Berners-Lee, T. (2006). *Linked Data: Design Issues*. Recuperado 4 de dezembro de 2016, de <https://www.w3.org/DesignIssues/LinkedData.html>

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43.
- Bertolini, M. V., Bothmann, R. L., Rodríguez, E. E., Galeffi, A., & McGarry, D. (Eds.) (2016). Declaração dos Princípios Internacionais de Catalogação (PIC). *Pela Seção de Catalogação da IFLA e as Reuniões de Especialistas sobre um Código de Catalogação Internacional.*, 21. Recuperado de https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-pt.pdf
- Catarino, M. E. (2014). Simple Knowledge Organization System: Construindo sistemas de organização do conhecimento no contexto da Web Semântica. *Informação & Tecnologia (ITEC)*, 1(1), 17-28. Recuperado de <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/19307/10980>
- Catarino, M. E., & Souza, T. B. de. (2012). A representação descritiva no contexto da web semântica. *Transinformação*, 24(2), 77-90. doi:10.1590/S0103-37862012000200001
- Coneglian, C. S., Roa-Martínez, S. M., Ferreira, A. M. J. F. da C., Vidotti, S. A. B. G., & Santarém Segundo, J. E. (2019). Tecnologías de la Web Semántica en arquitectura de la información. *Revista Interamericana de Bibliotecología*, 42(1), 23-35. doi:10.17533/udea.rib.v42n1a03
- Coyle, K. (2016). *FRBR, before and after: A look at our bibliographic models*. Chicago, IL: ALA Editions.
- Ferreira, J. A., & Santos, P. L. V. A. da C. (2013). O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. *Informação & Sociedade: Estudos*, 23(2), 13-23. Recuperado de <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/15436>
- Glushko, R. J. (2014). *The discipline of organizing: Core concepts edition*. Sebastopol, CA: O'Reilly Media. Recuperado de <https://www.overdrive.com/search?q=70B2FB07-B1BE-458C-8BC5-5E7FBBEF946E>
- Isotani, S., & Bittencourt, I. I. (2015). *Dados Abertos Conectados: Em busca da Web do Conhecimento*. São Paulo: Novatec Editora.
- Marcondes, C. H. (2001). Representação e economia da informação. *Ciência da Informação*, 30(1), 61-70.
- McGarry, K. (1999). *O contexto dinâmico da informação: Uma análise introdutória*. Brasília: Briquet de Lemos.
- Medeiros, M. B. B., & Café, L. M. A. (2008). Organização da Informação ou Organização do Conhecimento? In *Anais do IX Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB)*, São Paulo, Brasil. Recuperado de enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142
- Nowack, B. (2009). *The Semantic Web Technology Stack (not a piece of cake...)* [Photo]. Recuperado de <https://www.flickr.com/photos/jalbertbowdenii/15136351742/>
- Online Computer Library Center. (2016). *PURL*. Recuperado 19 de setembro de 2019, de <https://www.oclc.org/research/themes/data-science/purl.html>
- Ramalho, R. A. S. (2006). *Web Semântica: Aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação* (Dissertação de mestrado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, Brasil). Recuperado de <http://hdl.handle.net/11449/93709>

- Rendon Rojas, M. A., & Herrera Delgado, L. B. (2010). Bases filosóficas de la organización de la información. *Perspectivas em Ciência da Informação*, 15(1), 3-17. Recuperado de <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/957>
- Riva, P., Le Bœuf, P., & Žumer, M. (Eds.) (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. Den Haag, Netherlands: IFLA. Recuperado de https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf
- Santarém Segundo, J. E. (2014). Web Semântica: Introdução A Recuperação De Dados Usando Sparql. In *Anais do XV Encontro Nacional De Pesquisa Em Ciência Da Informação (ENANCIB)*, Belo Horizonte, MG, Brasil (pp. 3863-3882). Recuperado de <https://www.brapci.inf.br/index.php/article/view/0000015784/5ec5e71f1f6137b7f1a5f19b89496cb5/>
- Santarém Segundo, J. E., & Coneglian, C. S. (2016). Web Semântica e Ontologias: Um estudo sobre construção de axiomas e uso de inferências. *Informação & Informação*, 21(2), 217-244. doi:10.5433/1981-8920.2016v21n2p217
- Santarém Segundo, J. E., & Simionato, A. C. (2016). Uma abordagem sobre a estrutura do Geonames e suas contribuições para o Linking Open Data. In *Anais do XVII Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB)*, Salvador, BA, Brasil (pp. 117-137). Recuperado de <https://www.brapci.inf.br/index.php/res/download/53740>
- Santos Neto, A. L. dos, Marcondes, C. H., Pereira, D. V., Fonseca, E. R. da, Souza, I. V. P. de, Barbosa, N., ... Castro Martins, S. de. (2013). Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: Um caso machadiano. *Transinformação*, 25(1), 81-87. doi:10.1590/S0103-37862013000100008
- Santos, P. L. V. A. da C., & Alves, R. C. V. (2009). Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. *DataGramaZero: Revista de Ciência da Informação*, 10(6). Recuperado de <https://www.brapci.inf.br/index.php/article/view/0000008341/cc97dab7d8e1528baa04134fcd6c21b9>
- Santos, P. L. V. A. da C., & Pereira, A. M. (2014). *Catálogo: Breve história e contemporaneidade*. Niterói: Intertexto.
- Silva, R. E. da, & Santos, P. L. V. A. da C. (2012). Requisitos Funcionais para Registros Bibliográficos (FRBR): Considerações sobre o modelo e sua implementabilidade. *RBBB. Revista Brasileira de Biblioteconomia e Documentação*, 8(2), 116-129. Recuperado de <https://rbbd.febab.org.br/rbbd/article/view/214>
- Simionato, A. C. (2017). Família FR, BIBFRAME e Linked Data: Integração de dados bibliográficos. In Z. R. Zafalon & P. R. Dal'Evedove (Orgs.), *Perspectivas da representação documental: Discussão e experiências* (pp. 211-226). São Carlos: CPOI. Recuperado de https://www.researchgate.net/profile/Zaira_Zafalon/publication/320619115_Perspectivas_da_representacao_documental/links/59f13763aca272cdc7ce0a1f/Perspectivas-da-representacao-documental.pdf
- Souza, R. R., & Alvarenga, L. (2004). A Web Semântica e suas contribuições para a ciência da informação. *Ciência Da Informação*, 33(1), 132-141. doi:10.1590/S0100-19652004000100016
- Svenonius, E. (2009). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.

Vidotti, S. A. B. G., Coneglian, C. S., Roa-Martínez, S. M., Vecchiato, F. L., & Santarém Segundo, J. E. (2019). Web, Web Semântica e Web Pragmática: Um posicionamento da Arquitetura da Informação. *Informação & Sociedade: Estudos*, 29(1), 195-214. doi:10.22478/ufpb.1809-4783.2019v29n1.44358

CAPÍTULO 8
DA ORGANIZAÇÃO À DESCOBERTA
DO CONHECIMENTO: NOVOS CAMINHOS
PARA A RECUPERAÇÃO DA INFORMAÇÃO
FROM KNOWLEDGE ORGANIZATION
TO KNOWLEDGE DISCOVERY: NEW WAYS
FOR INFORMATION RETRIEVAL

Maria Cristina Vieira de Freitas
Universidade de Coimbra (UC)
Faculdade de Letras/CEIS20
cristina.freitas@fl.uc.pt
ORCID: 0000-0002-8849-8792

Olívia Pestana
Universidade do Porto (UP)
Faculdade de Letras/CITCEM
opestana@letras.up.pt
ORCID: 0000-0002-5485-3143

Maria Aparecida Moura
Universidade Federal de Minas Gerais (UFMG)
Escola de Ciência da Informação
mamoura@ufmg.br
ORCID: 0000-0003-2670-923X

Resumo: Partindo de uma abordagem introdutória da Organização do Conhecimento, identificam-se alguns dos aspectos essenciais relacionados com as origens e a evolução do campo científico, a partir de finais do século XIX até início do século XXI, salientando-se algumas das principais abordagens, dos estudos prévios e das tendências das pesquisas nessas matérias. Tendo por base esta abordagem, segue-se a apresentação de recentes tendências de sistemas de descoberta do conhecimento em ambientes digitais, os quais desafiam os profissionais na antecipação da oferta de informações qualitativas e conexas às necessidades contextuais dos usuários da informação. Neste âmbito, serão apresentados os conceitos de convergência e distância semântica e sistemas de descoberta do conhecimento e suas implicações nos processos de organização da informação em contexto.

Palavras-Chave: Organização do Conhecimento. Similaridade semântica. Sistemas de descoberta do conhecimento.

Abstract: Starting from an introductory approach to knowledge organization, some of the main aspects related to its origins and the evolution of the scientific field, from the end of the 19th century until the beginning of the 21st century, as well as some of the main approaches, studies and research trends in this area, are identified. Based on this approach, we present the recent trends of knowledge discovery systems in digital environments, which challenge professionals in anticipating the supply of qualitative information and related to the contextual needs of information users. In this context we will present the concepts of convergence and semantic distance and knowledge discovery systems as well as their implications in the processes of information organization in context.

Keywords: Knowledge organization. Semantic similarity. Knowledge discovery systems.

1. Introdução

A história da Organização do Conhecimento (OC), interpretada como disciplina, é de traçado difícil. Trata-se de uma área de estudos com uma natureza complexa, porque é tecida juntamente com várias outras áreas do conhecimento. Diversos estudos situam a sua origem e evolução em balizas que se encontram entre a segunda metade do século XIX, quando é sistematizada e associada aos estudos de Biblioteconomia e de Documentação, até o final do século XX, momento em que sofre uma expressiva influência, ocasionada pelos desenvolvimentos notados nas tecnologias de informação, nos ambientes e conteúdos digitais, mas também nas mais diversas áreas científicas.

A intensificação da presença de distintos atores sociais nos mais diversos ambientes informacionais, associada à intensa digitalização e circulação da informação em fluxos não homogêneos, exige a adoção de novos e dinâmicos procedimentos de organização da informação e do conhecimento em contexto. De igual modo, surgem novos caminhos na recuperação da informação com a implementação de sistemas de descoberta do conhecimento em ambientes digitais, desafiando os profissionais, e os sistemas, na antecipação da oferta de informações qualitativas e conexas às necessidades contextuais dos usuários da informação.

Com base nesse enquadramento, neste capítulo são realizados intentos de sistematização de alguns aspectos relacionados com as origens, a evolução, os estudos e algumas das tendências mais visíveis de investigação em OC, finalizando com uma análise da sua aplicação em uma das áreas com as quais tem uma proximidade reconhecida: a Recuperação da Informação (RI) (Smiraglia, 2012; Hjørland, 2008). Esta área, cada vez mais explorada pelas ciências da computação, é simultaneamente causa e consequência da evolução dos estudos em OC.

Para que se tenha alguma percepção do percurso exploratório realizado para a produção deste capítulo, descrevem-se, sumariamente, os conteúdos presentes nas suas subdivisões e os recursos metodológicos que suportam a sua construção intelectual.

Assim, em uma primeira parte intenta-se uma análise da literatura que subjaz aos atuais estudos sobre OC, sob diferentes ângulos, sem a pretensão de atingir a exaustividade, o que extrapolaria largamente os limites impostos à sua produção. Nesse intuito, abordam-se as origens, a evolução e as inter-relações da OC, de modo a proporcionar uma visão panorâmica do seu campo de estudo, finalizando com a exposição de alguns dos estudos prévios e das tendências de investigação mais significativas e visíveis no seu entorno.

Em uma segunda parte, apresentam-se questões que atualmente integram os ambientes de exploração de conteúdos digitais e que ilustram o impacto da OC no terreno da Recuperação da Informação (RI), com vistas à sua descoberta. Esta análise tem como foco a questão da rastreabilidade da informação, da similaridade semântica e da sua relação com a OC e com os sistemas tecnológicos de descoberta do conhecimento. Serão apresentados os conceitos de convergência e distância semântica e sistemas de descoberta do conhecimento, bem como suas implicações nos processos de organização da informação em contexto.

Com vistas à elaboração da primeira parte deste capítulo foi constituído um corpus de textos identificados através da revisão da literatura recuperada pela pesquisa nas bases de dados *Library, Information Science & Technology Abstracts* (LISTA) e *Library & Information Science Source*, no decorrer do mês de janeiro de 2019. As expressões de busca usadas nessa plataforma restringiram a pesquisa a textos nos idiomas português, inglês e espanhol e foram limitadas aos anos de 2000 a 2018. Foram, ainda, consultadas as atas dos eventos promovidos pela ISKO (*International Society for Knowledge Organization*), considerando os eventos internacionais

e os dos capítulos Brasil e Espanha/Portugal. Para além desta pesquisa, ainda foram recuperados e analisados alguns trabalhos citados em artigos do corpus, para consolidação do exposto na análise. Os autores considerados mais expressivos, quer pelo número de textos recuperados, quer pelo volume de citações realizadas por outros autores (ainda que estas não tenham sido contabilizadas para o efeito) e também no âmbito da elaboração desta parte do capítulo, foram Birger Hjørland e Ingetraut Dahlberg, o que condiz com o fato de serem estes, efetivamente, alguns dos autores mais recorrentes nos estudos dedicados a estas matérias e, logicamente, com uma maior visibilidade e recuperação em bases de dados especializadas. Finalmente, os textos selecionados foram submetidos a um processo de análise de conteúdo, apoiada em software de análise qualitativa de dados Atlas.ti¹, o qual oferece uma concretização relativamente simples dos procedimentos de análise e é ajustado aos objetivos que determinaram as regras de produção deste capítulo.

A segunda parte deste capítulo, pela menor dimensão da literatura existente sobre a matéria e dada a especificidade e a atualidade da abordagem, alicerçou-se em uma busca exploratória mais ampla de literatura, convergindo na definição dos principais conceitos subjacentes aos mecanismos dos sistemas de descoberta, os quais se enquadram nos propósitos da RI.

2. Organização do Conhecimento: das origens às tendências de investigação

Origens, evolução e inter-relações do campo de estudo

A abordagem das origens da OC exige a observância da falta de consenso ao longo das últimas décadas quanto à sua assunção

¹ Disponível em: <https://atlasti.com/>.

ou como uma área do domínio da Ciência da Informação (CI) (ver, por ex. Smiraglia, 2011), ou como uma ciência autônoma com fortes laços interdisciplinares com a Ciência da Informação (ver, por ex. Dahlberg, 2006). Pando e Almeida (2015), em uma abordagem terminológica e epistemológica, vão mais além discutindo os conceitos de organização da informação e organização do conhecimento, revelando que ainda não há um uso concreto destes termos. Este fato vem, também, contribuir para a inexistência de uma direção consensual deste campo temático. Não obstante, há linhas cronológicas, a partir das quais podemos traçar uma evolução.

O **estatuto de campo de estudo** atribuído à Biblioteconomia, uma das áreas fortemente relacionadas com a OC na ótica, então, de diversos autores, dever-se-á, sobretudo, à publicação da primeira edição da classificação de Melvin Dewey, datada do final da década de 1870, sendo este o período no qual também emergem e proliferam na Europa e na América as designadas Escolas de Biblioteconomia (Hjørland, 2003). Uma outra via de relacionamento, que não exclui a anterior, vincula a OC aos clássicos estudos de Paul Otlet e de Henri La Fontaine, iniciados no dealbar do século XX e, portanto, no contexto de formação e no decurso das então designadas Ciências Documentais (Dahlberg, 2006). A par com estas relações, é sobejamente conhecido que o termo que designa a área – organização do conhecimento – aparece e se consolida em obras consideradas fundacionais, publicadas na primeira trintena do século XX, por nomes de grande envergadura associados à classificação bibliográfica: Charles A. Cutter, Ernest Cushing Richardson, W. C. Berwick Sayers, S. R. Ranganathan e Henry Bliss². Por volta desta época e amparada nesses igualmente clássicos estudos, a OC ter-se-á

² O termo terá sido utilizado por Henry Bliss em duas obras fundamentais, publicadas, respectivamente, em 1929 e em 1933: *The Organization of Knowledge and the System of the Sciences* e *The Organization of Knowledge in Libraries* (Guimarães, Santos, Cândido, & Pinho, 2014). Também a obra *Prolegomena to Library Classification*,

consolidado enquanto campo teórico e, portanto, com pretensões científicas e acadêmicas, desvinculando-se de uma etiqueta que a posicionava até então no rol de um saber eminentemente prático e empírico (Almeida, Grácio, & Castanha, 2017; Hjørland, 2016; Guimarães, Santos, Cândido, & Pinho, 2014; Ferreira & Guimarães, 2011; Gunjal, 2012; Smiraglia, 2012; Fujita, 2008; Dahlberg, 2006).

A partir do início da década de 1970, após uma consolidação que decorre das excepcionais demarcações prévias, da rapidez que caracterizou as transformações tecnológicas e dos esforços teóricos que se prolongam até os anos 1990, a área ganha um novo impulso, dado pelas iniciativas extraordinárias de Dagobert Soergel e de Ingetraut Dahlberg. Nesse período, serão cravados dois outros marcos fundamentais e decisivos para a sua especialização científica. O primeiro refere-se à criação da revista *International Classification* (1973-1974) que, no início dos anos 1990, passa a designar-se *Knowledge Organization Journal – KO* (1993)³, tendo este vindo a ser o veículo internacional com o maior prestígio nessa matéria até os dias atuais. O segundo associa-se ao estabelecimento, no final dos anos 1980, da *International Society for Knowledge Organization – ISKO* (1989)⁴, sociedade científica que, desde então, assume o papel preponderante e a responsabilidade pela concertação de ações de caráter associativo, ao nível nacional, regional e internacional, e pela realização periódica de eventos científicos nacionais e internacionais, nos quais são discutidos temas e problemas importantes para a área (Hjørland, 2016; Guimarães, Santos, Cândido, & Pinho, 2014; Ferreira & Guimarães, 2011; Smiraglia, 2012; Fujita, 2008; Dahlberg, 2006; Hjørland, 2003).

publicada em 1931 por Ranganathan, representa um enorme contributo para este mesmo fim (Gunjal, 2012).

³ Disponível em: <http://www.isko.org/ko.html>.

⁴ Disponível em: <http://www.isko.org/about.html>.

Deste modo, do início ao fim do século XX, a OC conhece alguns dos momentos que têm sido apontados como os mais significativos para a sua evolução e consolidação. Primeiramente, alcança um desenvolvimento e um reconhecimento acadêmico, institucionalizando-se e generalizando-se o seu ensino em programas espalhados por escolas e por universidades, em diversos países da Europa e da América. Posteriormente, atinge uma relativa maturidade como campo e como tema de pesquisa, e torna-se uma área de toque da Ciência da Informação, contribuindo, inclusivamente, para o seu desenvolvimento científico (Hjørland, 2016; Guimarães, Santos, Cândido, & Pinho, 2014; Gunjal, 2012; Hjørland, 2008; Broughton, 2008).

Dado o seu aspecto híbrido, a OC abarca diversas disciplinas, sendo marcada pela coexistência de diferentes correntes de pensamento, o que no presente leva à constante necessidade de efetuar uma vigilância e uma prospecção aprofundada sobre quais são as suas próprias bases metodológicas (e quais não são), uma vez que o mapeamento dos seus contributos endógenos e exógenos será condição *sine qua non* para a sua autocompreensão enquanto área de estudo, reconhecendo-se, no entanto, que a dificuldade em realizar um tal exercício residirá em uma estreita dependência da perspectiva pela qual ela venha a ser realizada (Hjørland, 2016; Smiraglia, 2012).

Nesse sentido, algumas críticas denunciam o fato de os seus desenvolvimentos, na prática, estarem mais orientados pelos progressos observados em outras áreas, ou seja, no meio exógeno (*e.g.*, tecnologias de informação, de onde virá o seu estreito relacionamento com o armazenamento e, essencialmente, a recuperação da informação⁵), do que pelos seus próprios progressos enquanto área

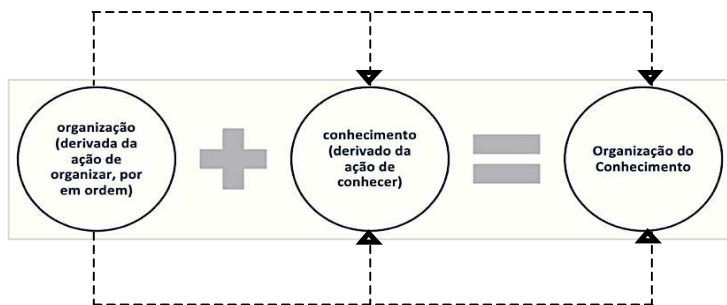
⁵ Conforme podemos ver na segunda parte deste trabalho, a respeito dos sistemas de descoberta do conhecimento.

científica *per se*, sendo esse, naturalmente, o seu meio endógeno (Hjørland, 2003, 2016). No entanto, essa aparente falta de uma precisão científica, que à partida pode ser interpretada como uma fragilidade, é também compensada pela diversidade conceitual e pelos inegáveis resultados obtidos ao nível aplicacional, sendo essa também considerada a sua força (Fujita, 2008).

Todo esse hibridismo (académico, técnico e científico) seria, então, nesta perspectiva, um reflexo do seu relacionamento próximo, o que tem, por muitos autores, sido entendido como áreas científicas afins (Biblioteconomia, Ciências Documentais, Ciência da Informação e Tecnologias de Informação, para apenas citarmos algumas). Assim, nascida no meio das teorias destinadas à organização e à estruturação sistemática dos conceitos, das quais é francamente devedora, com vistas à sua tradução em modelos representacionais, ela se vem consolidar como um “domínio” que em um certo sentido busca o autoconhecimento (Almeida, Grácio, & Castanha, 2017; Hjørland, 2016; Guimarães, Santos, Cândido, & Pinho, 2014; Gomes, 2009; Dahlberg, 2006).

O **objeto de estudo** da OC constitui-se da junção de dois conceitos (Fujita, 2008; Dahlberg, 2006) vinculados em uma aparentemente simples equação (Figura 1):

Figura 1: Elementos definidores da organização do conhecimento.



Fonte: Elaboração própria, com base em Dahlberg (2006) e em Fujita (2008).

Em uma simples lógica, dispor em uma certa ordem o que se conhece pode querer dizer estruturar os “registros”, ou seja, as inscrições do conhecimento (Chernyi, 1997 citado por Gomes, 2009). No entanto, obedecendo a uma racionalidade mais complexa, a OC não se limita ao terreno dos objetos (concretos), evidenciando-se justamente pelo fato de ir além e de lidar com os conceitos (abstratos), isto é, com ideias ou noções anteriores aos objetos que procura organizar e representar (Brascher & Café, 2008; Fujita, 2008; Dahlberg, 2006).

Esta viragem teórico-metodológica é o que possivelmente a torna tão particularmente complexa e seguramente será o que a distingue de outras áreas muito próximas como, por exemplo, a Organização da Informação (OI) (Brascher & Café, 2008). Invocamos aqui, novamente, o trabalho de Pando e Almeida (2015) que nos evidencia a complexidade terminológica entre OC e OI sem aparente consolidação. Na ótica da Organização do Conhecimento, diversos autores sublinham que este – o conhecimento – reside na mente das pessoas e que como tal *a priori* nem pode ser gerido nem pode ser organizado, tratando-se nesse caso de matéria de estudo das Ciências Cognitivas (Hjørland, 2008). Assim, a Ciência da Informação, ao lidar com a organização dos recursos informacionais, que são os artefatos através dos quais o conhecimento é exteriorizado e compartilhado com, e por, pessoas, estaria, em um sentido estrito, a organizar recursos de informação e, nessa acepção, a organizar informação. No entanto, ao considerar, não a manifestação concreta ou material do conhecimento, mas o que a antecede e que nesse caso é a ideia ou a noção (abstrata), prevaleceria uma outra lógica: a de organizar o conhecimento. Portanto, a nosso parecer, o que está em causa no ato de delimitar as fronteiras entre as áreas é perspectivar o conceito (OC) ou perspectivar o objeto (OI). De uma forma clarividente, Brascher e Café (2008, p. 6) resumem ambos os processos do seguinte modo:

Em nossa visão, temos dois tipos distintos de processos de organização, um que se aplica às ocorrências individuais de objetos informacionais – o processo de organização da informação – e outro que se aplica a unidades do pensamento (conceitos) – o processo de organização do conhecimento.

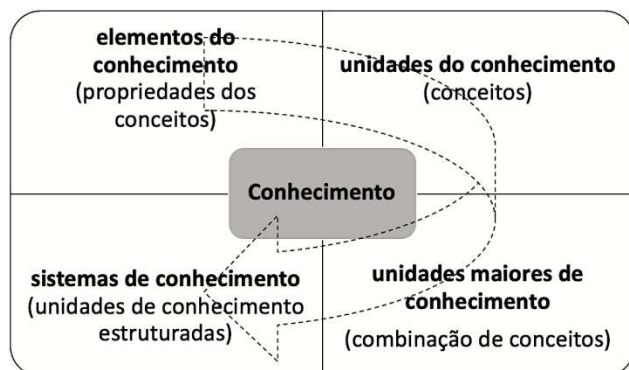
Deste modo, a OI compreende a organização de objetos para o seu arranjo sistemático em coleções (físicas, nesse caso). A OC compreende a construção de “modelos”, que se constituem em abstrações da realidade (ou seja, representações do conhecimento) (Brascher & Café, 2008, p. 6).

Ainda, em uma perspectiva que nos parece complementar à anterior, por oposição ao conhecimento individual que pelo fato de se apresentar para uma mente ou um cognóscio não seria à partida organizável, simplesmente por não ser possível exercer um controlo técnico e científico sobre o que concretamente uma pessoa conhece (ou seja, o estoque individual de conhecimento), compreende-se que o objeto de estudo da OC venha a ser o conhecimento exteriorizado, i.é., socializado, compartilhado, documentado (Guimarães, Santos, Cândido, & Pinho, 2014; Barité, 2001 citado por Fujita, 2008). Nessa acepção de “corpo coletivo de conhecimento”, ou seja, de uma construção social que é fruto de um processo e que deste modo se torna abarcável e registrável em documentos, ou seja, em objetos portadores de informação, será o conhecimento uma entidade passível de submissão a um processo de controlo e de organização e, evidentemente, de gestão⁶.

⁶ Neste ponto, a OC entra em uma zona de confluência teórica com outras áreas ou domínios do conhecimento, entre as quais, a Gestão do Conhecimento. Trata-se esta de uma visão lata da OC, conforme veremos mais adiante. Para uma leitura de autores que apresentam discussões já consideradas clássicas sobre essa matéria, sugere-se consultar: Choo (1996), Allen (1996) e Buckland (1995).

Na sua abordagem do conhecimento como sendo o objeto de estudo da OC, Dahlberg (2006, p. 12) o define sob uma perspectiva que, apesar de individual e, portanto, cognitivista, a nosso parecer é multidimensional e inter-relacionada. Este contributo (Dahlberg, 2006, p. 12) é citado em alguns dos textos consultados (cf. Gomes, 2009; Brascher & Café, 2008; Fujita, 2008) (Figura 2).

Figura 2: O objeto de estudo da organização do conhecimento e as suas relações.



Fonte: Elaboração própria, com base em Dahlberg, 2006, p. 12.

Esse modelo matricial (Figura 2) permite compreender que, do ponto de vista teórico e metodológico, a área de organização do conhecimento tem como objeto de estudo, principalmente, e acima de tudo, o que Dahlberg (2006, p. 12) designa “unidades de conhecimento”, ou seja, os conceitos, no seu entrecruzamento e na articulação com outras três dimensões que a autora também considera importantes para a sua compreensão e a sua diferenciação: os elementos do conhecimento, as unidades maiores de conhecimento e os sistemas de conhecimento. Assim, por esta lógica, toda a forma de organizar conhecimento – e que resulta na construção de sistemas de conhecimento – fundamenta-se nessas unidades de conhecimento (ou conceitos), bem como nas suas diferentes combinações (unidades maiores de conhecimento).

Com efeito, de acordo com Smiraglia (2012), entre a teoria que suporta a produção desses sistemas de conhecimento e os seus produtos (i.é., os Sistemas de Organização do Conhecimento – SOC, entre os quais e, principalmente, estão: classificações, taxonomias, ontologias e tesouros) existe atualmente um “discurso bastante sólido”, ou seja, um corpo de conhecimentos acumulados e validados. Esses produtos serão frutos da criação de regras ou teorias usadas para descobrir uma “ordem natural” para o conhecimento, ou seja, uma sua forma de organização, ou mesmo para impor uma sequência útil para a sua descoberta.

Essas questões levam-nos ao tema das **delimitações do campo**, bem como à questão dos **modelos de investigação** comumente adotados nas pesquisas realizadas na área da OC e que em muito ficam a dever aos esforços teóricos e epistemológicos de Birger Hjørland⁷, alguns dos quais tentaremos sumariar na sequência.

Quanto ao primeiro aspecto, Smiraglia (2012) pondera que a classificação de áreas epistemológicas que se encontram na confluência de, e com, outras áreas, como é o caso da OC, é um exercício exigente, que implica a seleção e a clarificação do método que preside à respectiva classificação, assumindo-se os riscos inerentes. Não obstante os constrangimentos e face à necessidade de fortalecer o campo e a respectiva comunidade epistêmica, proliferam-se os intentos. No entender de Hjørland (2016, 2008, 2003), por exemplo, em um **sentido estrito**, a OC demarca-se como um campo de estudo vinculado à *Library and Information Science*. Em um **sentido lato**, extrapola estes mesmos limites (Quadro 1).

⁷ No estudo desenvolvido para uma dissertação de mestrado, de um total de 35 autores mais citados na área da Organização do Conhecimento, Birger Hjørland aparece em primeiro lugar, com 29 citações, no período de 2004 a 2013 (Lima, 2015, p. 49).

Quadro 1: Comparativo das demarcações da OC, com base no modelo proposto por Hjørland (2016, 2008, 2003)

OC em sentido estrito	OC em sentido lato
Objetivos: devotada aos processos e aos sistemas usados para organizar documentos, bem como as suas representações e conceitos (conhecimento).	Objetivos: definida como atividade intelectual e social de organização do conhecimento, nas mais diversas disciplinas, instituições e sistemas simbólicos, teóricos e conceituais.
Áreas nucleares: situa-se predominantemente no campo da <i>Library and Information Science</i> (admitindo-se também grandes influências das áreas dedicadas às TIC).	Áreas nucleares: esta dimensão cobre campos tais como a sociologia do conhecimento, a história social do conhecimento, entre outros.
Tarefas: atividades tais como descrição bibliográfica, indexação e classificação, levadas a efeito por instituições de informação e também por produtores de conteúdos bibliográficos (<i>e.g.</i> , bases de dados bibliográficas).	Tarefas: as preocupações da OC, nesse caso, giram em torno de questões tais como: a) a forma pela qual o conhecimento é organizado na sociedade; e b) como o conhecimento é organizado em disciplinas académicas (por exemplo, as taxonomias usadas na Biologia).
Visão: no âmbito desta comunidade epistêmica de pequeno espectro predomina a visão de organizar para recuperar (RI) conhecimento (em registos bibliográficos, índices de citação, textos completos e na Internet).	Visão: no âmbito desta comunidade epistêmica de largo espectro predomina uma visão de atividade intelectual da OC.

Na interpretação de Fujita (2008, p. 8), “Hjørland (2008) distingue a Organização do Conhecimento em sentido específico, como a organização cognitiva ou intelectual do conhecimento e, em sentido amplo, como a organização social do conhecimento”.

Ambas as demarcações levam a constatar a existência de várias teorias sobre a organização do conhecimento, dentro e fora do próprio domínio da OC (Hjørland, 2016, 2008, 2003). E o reconhecimento dessas diferentes abordagens e teorias, tanto dentro como fora da OC, torna ainda mais estratégicas as escolhas relacionadas com a base de suporte a adotar quer no seu ensino, quer na sua pesquisa. Estas perspectivas, contudo, não se autoexcluem, porque as teorias que se conectam com a OC em um sentido mais abrangente são as mesmas que servem de base para organizar o conhecimento no seu sentido mais estrito, ou seja, no âmbito da própria *Library and*

Information Science. Sob essa lógica interpretativa que, conforme Gunjal (2012, p. 81), é comum a vários autores, as teorias sobre a OC que são desenvolvidas no interior da CI tornam-se de algum modo dependentes do conhecimento que é produzido no exterior da própria CI.

Ainda, Hjørland (2016, p. 476 e ss.) enumera alguns dos contributos teóricos que se desdobram na perspectiva de cada uma dessas abordagens, algumas das quais associam-se à evolução da própria área da OC e a alguns dos seus mais renomados teóricos (Quadro 2).

Quadro 2: Abordagens desenvolvidas no interior e no exterior da OC

A. Abordagens desenvolvidas no interior da OC	B. Abordagens desenvolvidas no exterior da OC
A.1 Abordagens práticas e intuitivas (ex. Dewey Decimal Classification)	B.1 Abordagens bibliométricas
A.2 Abordagens baseadas em consensos (ex. consensos científicos e educacionais propostos por Henry Bliss)	B.2 Abordagens em RI
A.3 Abordagens de análises de facetas (S. R. Ranganathan)	B.3 Outras abordagens (ontológica, de gênero, etc.)
A.4 Abordagens cognitivas e abordagens baseadas nos usuários	
A.5 Análise de domínio e abordagens epistemológicas	

Fonte: Elaboração própria, com base em Hjørland, 2016, p. 476 e ss.

Hjørland (2003, p. 107) reconhece que há toda uma teia de relações entre as diferentes abordagens e que isso tem implicações nos métodos utilizados. Em uma visão geral, começa por dizer que, quanto à natureza, as pesquisas não devem ser classificadas de uma forma reducionista ou dicotômica. Estudos há, no âmbito da OC, em que ambos os enfoques – qualitativo e quantitativo – podem estar envolvidos concomitantemente. Também refere que as questões de ordem epistemológica e, mais recentemente, as ontológicas,

perpassam os estudos em OC e conseqüentemente interferem com a sua classificação⁸.

O mesmo Hjørland (2008) advoga que a OC não pertence ao domínio exclusivo da *Library and Information Science* – isto porque o conhecimento é classificado e organizado em várias áreas (por exemplo, a Biologia, a Sociologia, a Filosofia, etc.), obedecendo a princípios teóricos consensuais ou não – e cita diversos exemplos de domínios correlatos⁹, reforçando que estes campos possuem, ao mesmo tempo, perspectivas disciplinares diferentes, mas compartilham bases teóricas e epistemológicas comuns, pelo que as diversas áreas científicas podem beneficiar-se mutuamente desta fertilização cruzada¹⁰.

Toda essa base de argumentação usada, ao mesmo tempo, no desenvolvimento e na classificação das pesquisas em OC é, portanto, inclusiva e não exclusiva. Não entra, pois, em desacordo com propostas ditas tradicionais (cf. modelo de Dahlberg, 2006, fig. 2), que elegem a Teoria do Conceito e a Análise Conceitual como bases teórico-metodológicas gerais da OC e que têm como objeto de estudo o conceito, as suas propriedades e relações, visando a sua inclusão em um domínio, junto de outros conceitos, e a sua representação em uma estrutura sistematicamente organizada (Van den Heuvel & Smiraglia, 2010; Vickery, 2008 citado por Brascher & Café, 2008; Dahlberg, 2006). Contrariamente, a integra e a aperfeiçoa.

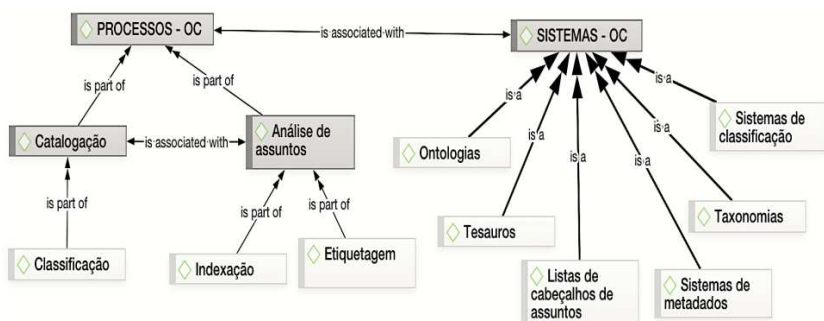
⁸ Para uma comparação entre estas duas vertentes da OC e as tendências atuais, consulte-se: Chaudhry (2016).

⁹ Nomeadamente, Informática, Linguística, Teoria do Conhecimento, Teoria das Organizações Sociais, entre outras (Hjørland, 2008).

¹⁰ A título ilustrativo, Hjørland (2008) menciona duas abordagens que se encaixam nessa descrição, dentro e fora da OC: a) a abordagem ontológica de Heinrich Herre, desenvolvida para representar o conhecimento em domínios específicos e baseada em métodos “onto-axiomáticos” e b) a abordagem de gênero de Jack Andersen, aplicada à OC.

Os conteúdos de alguns dos textos consultados também revelam alguns exemplos práticos do que são reconhecidamente os principais **resultados dos estudos** desenvolvidos em OC, em mais de um século de investigação e de intensa produção científica (Smiraglia, 2012; Hjørland, 2003). Tanto os processos de organização do conhecimento como os sistemas de organização do conhecimento resultam de ações que envolvem a organização intelectual do conhecimento que, no parecer de Hjørland (2008), também pode ser denominada organização cognitiva do conhecimento ou organização do conhecimento em um sentido restrito (Fujita, 2008, p. 8) (Figura 3).

Figura 3: Esquema das relações entre processos e sistemas de organização do conhecimento.



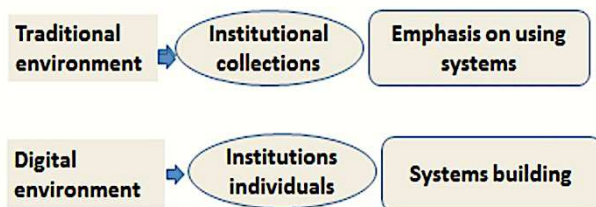
Fonte: Elaboração própria, com base em Hjørland (2003) e em Smiraglia (2012).

Como se infere, trata-se de uma miríade de resultados teóricos e empíricos, entre os quais, normas, orientações, regras, teorias e aplicações, que vieram a associar-se aos desenvolvimentos tecnológicos que se fizeram sentir fortemente, e sobretudo, a partir da década de 1990 e que proporcionam, na atualidade, um “retrato muito claro” do que a OC significou no decorrer do século XX.

A modo de resumo, Chaudhry (2016, p. 102) refere que no atual século a OC se reparte entre a ênfase dada à utilização dos sistemas nos ambientes tradicionais, que se instituíram no decorrer do século

passado, e o desvio no foco para os ambientes digitais, nos quais a tônica se desloca para a construção de sistemas. E complementa que essas mudanças contextuais levam ao uso de abordagens holísticas quer no ensino, quer na pesquisa em OC (Figura 4).

Figura 4: Mudanças contextuais em OC.



Fonte: Chaudhry, 2016, p. 102.

3. Tendências na aplicação da OC em RI: colaboração, similaridade semântica e sistemas de descoberta do conhecimento

Monitoramento científico e vestígios digitais

A cultura epistêmica tem passado por inúmeras transformações que impactam os modos de produzir e acessar o conhecimento na contemporaneidade. Isso se deve, em grande parte, pela ampliação da cooperação internacional, a intensificação dos processos de digitalização da informação e pelo uso de algoritmos na identificação de latências e percursos que tornaram possível a sugestão de parcerias mediadas por filtros de colaboração.

A colaboração científica relaciona-se à interação e ao desenvolvimento de atividades intelectuais e científicas compartilhadas entre dois ou mais pesquisadores e envolve o compartilhamento de ideias, recursos econômicos ou estruturas laboratoriais.

Atualmente, as possibilidades de monitoramento das práticas culturais e científicas em redes passaram a exigir um olhar da orga-

nização da informação e do conhecimento com ganhos qualitativos para os processos criativos em rede.

Um dos aspectos em que essa articulação é perceptível são as redes latentes. As redes latentes são o resultado de uma análise sobre a probabilidade de conexão entre pesquisadores. A colaboração latente pode ser percebida na associação entre os tipos de publicações que são adotados na pesquisa, no padrão temático desenvolvido, nas práticas científicas e no compartilhamento de descritores palavras-chave e referências acadêmicas.

De acordo com Boutain, Gallezot e Duvernay (2008), redes latentes são baseadas em infometria induzida e ajudam a revelar novas associações entre dois ou mais autores através da análise relacional.

As possibilidades de agregação a partir da latência presumida entre pesquisadores tornaram oportunos os esforços no sentido da articulação e da pesquisa entre a infometria e a organização do conhecimento.

Nesse âmbito, encontram-se também os estudos de Descoberta do Conhecimento (DC). A descoberta de conhecimento é um processo orientado pela combinação de diferentes metodologias que visam identificar em um volume de dados, padrões regulares e inéditos que possam representar novidade e utilidade em um domínio específico.

Béjar (2019) destaca que a descoberta do conhecimento possui como características o padrão, a validade, a novidade, a utilidade e a compreensibilidade.

O padrão responde por qualquer formalismo de representação capaz de descrever as características comuns dos dados; a validade refere-se à capacidade de se prever o comportamento de novas informações com um grau de certeza; a novidade é a possibilidade de o conhecimento identificado referir-se a algo que não se saiba a respeito no domínio ou a qualquer conhecimento prévio descoberto; a utilidade revela se conhecimento descoberto é útil ou permite realizar ações que geram algum benefício a uma dada comunidade,

de acordo com critérios estabelecidos; e a compreensibilidade vincula-se ao reconhecimento e interpretação da descoberta por especialistas do domínio.

O percurso para a descoberta de conhecimento em bases de dados envolve: o estudo do domínio, a criação do conjunto de dados, o pré-processamento dos dados, a redução da dimensão, a seleção do objetivo da descoberta, a seleção das metodologias adequadas, a mineração de dados, a avaliação dos resultados e a interpretação e, finalmente, o uso do conhecimento.

Os espaços heterotópicos articulados em torno de diferentes práticas discursivas em rede são os ambientes propícios para a aplicação de metodologias destinadas à descoberta do conhecimento em virtude dos inúmeros vestígios digitais que agregam e propagam.

Nos últimos anos, os vestígios digitais obtidos de modo voluntário ou aleatório tornaram-se estratégicos identificadores de conteúdo de valor científico ou cultural, além de funcionarem como *commodities* no mercado econômico.

Rapidamente, observa-se a consolidação de uma nova modalidade de economia pautada no rastreamento dos vestígios deixados pelos usuários em suas incursões na Web, uma espécie de economia do vestígio realizada em articulação com a economia do compartilhamento.

Conforme Serres (2012), a temática do vestígio, anterior ao contexto digital podia ser estudada de diferentes perspectivas: como marca psíquica associada à memória e à imaginação, como indício ou como marca do passado. Como índice, refere-se a uma marca sensível de um fenômeno que dá origem a processos de significação. Nesse sentido, o vestígio é sempre dependente de um balizador perceptivo. A análise do vestígio, conforme Ricoeur citado por Serres (2012), envolve: a individualidade do objeto de análise, o caráter indireto do deciframento e a conjectura.

A individualidade do objeto de análise envolve a observação dos fenômenos ou de documentos singulares. O caráter indireto do deciframento vincula-se à ideia de que o conhecimento opera por vestígios e documentos e a conjectura inclui os comportamentos cognitivos e as operações intelectuais de análise, comparação e classificação.

Conforme ressalta Serres (2002), o estudo dos vestígios deixados pelos usuários em situações de busca da informação coloca em evidência a serendipidade deste ator social e nos permite compreender o seu processo abduativo¹¹.

Nesse contexto, os vestígios digitais são analisados como interação homem-máquina e como evento. Da perspectiva das interações homem-máquina apresenta-se com a marca das ações humanas no ambiente digital e do ponto de vista dos eventos é considerado “vestígio de uma sequência de eventos ordenados temporalmente ou não”.

Os vestígios são considerados, conforme Merzeau citado por Serres (2012, p. 89), como uma categoria de presença digital que se organiza em: vestígios declarativos, comportamentais, documentais e presença dos outros. Os vestígios declarativos são frequentemente encontrados em blogs e redes sociais; os comportamentais encontram-se disponíveis no registro das ações realizadas na Web, como os cookies, por exemplo. Os vestígios documentais revelam uma intencionalidade e se organizam em torno da nossa produção documentária em rede. Finalmente os vestígios da presença dos outros associam-se à construção da reputação pessoal explicitada nos ambientes digitais.

¹¹ Para um aprofundamento acerca das particularidades que se acercam ao fenômeno da serendipidade na Internet, sugere-se a consulta do estudo de Eva Sandri (2013).

Para Serres (2012, p. 89), os vestígios digitais possuem pelo menos sete características que podem ser repertoriadas. São elas: materialidade, intencionalidade, temporalidade, calculabilidade, interpretabilidade, reflexividade e memorização (Figura 5).

Figura 5: Características dos vestígios digitais.



Fonte: Elaboração própria, com base em Serres, 2012, p. 89.

Do ponto de vista da **materialidade**, constata-se que os ambientes digitais se organizam em infraestrutura tecnológica densa e complexa marcada por lastros materiais. No que se refere à **intencionalidade**, nota-se que os vestígios digitais não são sempre espontâneos pois se realizam em uma infraestrutura concebida com o propósito de recolher os vestígios das interações e dos usos específicos. A **temporalidade** vincula-se à decupagem de sequências discretas de eventos e se revela como o produto de um processo ou ação já realizada. A **calculabilidade** refere-se à possibilidade de uso dos vestígios digitais para calcular e prever comportamentos dos usuários a partir de dispositivos, cada vez mais sofisticados e integrados. A **interpretabilidade** é a leitura das possibilidades de significação apresentadas nos vestígios encontrados em *logs* de conexão, capturas de telas e em outras conexões tornadas visíveis pelos atores

e sua rede de interação. A **reflexibilidade** permite a análise e a reposição de ações futuras baseadas na reflexão das ações passadas e registradas em históricos de navegação e em linhas de tempo de redes sociais, por exemplo. Finalmente, a **memorização** potencializa o registro do curso das ações e dos eventos. Nesse sentido, ela tende a acentuar as características da reflexibilidade, pois registra de forma discreta as ações e os eventos, desejados ou não pelos atores sociais (Serres, 2012, p. 89 e ss.).

Assim, com os vestígios digitais,

nós passamos da ordem do dado para a construção, da raridade para a abundância, do índice ao símbolo, da singularidade à calculabilidade, da localização à deslocalização, da fixidez à volatilidade, da marca do passado ao registro do presente, da conservação ao dinamismo dos vestígios. (Serres, 2012, p. 93)

O compartilhamento de trilhas e vestígios nos ambientes digitais abriram também a possibilidade de que algoritmos, adotados em diferentes serviços de informação dispostos na Web, possam presumir, a partir de tais movimentações, a identidade dos sujeitos em interação.

Conforme analisa Cheney-Lippold (2011, p. 165),

[a] infraestrutura de rede da Internet, com sua capacidade tecnológica de rastrear os movimentos dos usuários em diferentes sites e servidores, deu origem a uma indústria de empresas de análise da Web que acumulam ativamente informações sobre indivíduos e aperfeiçoam algoritmos de computador para entender esses dados. O produto de muitas dessas empresas é uma “nova identidade algorítmica”, uma formação de identidade que funciona através de algoritmos matemáticos para inferir categorias de identidade em seres anônimos. Ele usa modelos de estatísticas

comuns para determinar o sexo, a classe ou a raça de uma maneira automática, ao mesmo tempo em que define o próprio significado de sexo, classe ou raça.

A articulação entre os conceitos de código e de categorização é fundamental para a compreensão dos desdobramentos da ideia de identidade algorítmica. O Código, que anteriormente estava associado à noção de regulação do uso do ciberespaço, hoje vincula-se à arquitetura das fronteiras e à regulação dos fluxos informacionais que trafegam na Internet. Nesse aspecto, o código pode orientar as formas de utilização da informação em tais arquiteturas.

Conforme destaca Cheney-Lippold (2011, p. 167),

[o]s códigos são objetos culturais incorporados e integrados dentro de um sistema social cuja lógica, regras e funcionamento explícito funcionam para determinar as novas condições de possibilidades de vida dos usuários. Como uma variável como X vem a ser definida, então, não é o resultado de um fato objetivo, mas é antes uma consequência tecnologicamente mediada e culturalmente situada.

A categorização orienta a consolidação de bases de dados de intenções que se nutrem das regularidades das expressões e comportamentos adotados pelos usuários na Internet.

A identidade algorítmica decorre, então, do monitoramento sistemático e pervasivo das atividades mundanas realizadas pelos usuários através de serviços de informação articulados em rede. Dessa perspectiva, a identidade algorítmica produz uma presença duplicada do sujeito na sociedade regulando em primeira instância o consumo de produtos e serviços, mas também a conformação, por esse *bias*, de determinadas visões de mundo por ele adotadas ou manifestas. Nesse contexto, certas categorias outrora fixas como

gênero, por exemplo, passam a se orientar pelo fluxo das novas identificações associadas aos termos mirando, sobretudo, a economia de escala proporcionada.

Em face da colaboração, da economia dos vestígios e da identidade algorítmica, surgem, cada vez mais, oportunidades inauditas de atuação no campo da organização do conhecimento. Percebe-se que, nesses contextos, continua sendo um desafio o volume de dados para alinhar, a heterogeneidade da representação dos conceitos e a qualidade para a obtenção de alinhamento semântico entre os diferentes SOC disponíveis ou em construção.

Similaridade semântica e organização do conhecimento

Em face desses desafios, os estudos associados à identificação da similaridade semântica em textos integrais surgem como possibilidade de atualização e alinhamento semântico-semiótico em SOC.

De acordo com Slimani (2013), a similaridade semântica é um método computacional que busca simular o julgamento humano utilizado para estimar a semelhança ou a distância semântica entre termos tendo como o propósito identificar conceitos que possuem características comuns. Com tal propósito, a similaridade semântica adota com frequência hierarquias de conhecimento dispostas em taxonomias, ontologias ou sistemas de classificação.

O autor destaca que a similaridade semântica é amplamente utilizada em aplicações de sistemas de recuperação de informação semântica baseados em conhecimento. O método apoia a compreensão da correspondência ideal entre termos, a desambiguação e a identificação do tipo de relação existente entre conceitos.

Dentre os métodos de determinação de medidas semânticas, destacam-se os baseados na estrutura ou na contagem de arestas, no conteúdo da informação, em características e o híbrido.

As medidas baseadas na estrutura computam a similaridade com base em uma dada ontologia. Nesse caso, a similaridade entre conceitos é medida em função do percurso que liga os termos na hierarquia da estrutura semântica adotada, a densidade e a profundidade (ontologias, taxonomias, mapas conceituais).

As medidas baseadas em conteúdo da informação usam o conteúdo informativo para aferir a semelhança semântica entre conceitos. Conforme destaca Slimani (2013), o valor do conteúdo informativo de um conceito é calculado com base na frequência do termo em uma determinada coleção de documentos.

As medidas baseadas em características tomam como referência as relações entre os termos, suas propriedades e características. O índice de similaridade é obtido pela descrição das propriedades do conceito ou pelas relações estabelecidas com outros termos presentes na hierarquia.

As medidas híbridas combinam as características estruturais, o conteúdo informacional, as características e as propriedades dos conceitos. A medida híbrida requer a adoção de ontologias para o estabelecimento de parâmetros que orientem a precisão.

Conforme destacado, as medidas semânticas tomam como parâmetro uma estrutura taxonômica ou ontológica; entretanto, as possibilidades atuais de monitoramento online das práticas linguageiras e discursivas em tempo real possibilitam a recuperação da informação ao mesmo tempo em que atualizam a estrutura semântica orientadora. Nessa perspectiva, uma das possibilidades mais utilizadas pela indústria do infoentretenimento tem sido as etiquetas, ou *hashtags*, adotadas pelos usuários de seus produtos e serviços.

Além disso, é possível, por exemplo, combinar estratégias de descoberta de conhecimento e de similaridade semântica no monitoramento da presença digital em práticas discursivas e linguageiras. Tal medida pode apoiar a produção de SOC mais dinâmicos,

expressivos e articulados às necessidades informacionais de uma dada comunidade cultural ou epistêmica, conforme assinalado na Figura 6.

Figura 6: Monitoramento da presença digital em práticas discursivas e languageiras.



Fonte: Elaboração própria.

Nesse sentido, pode-se reduzir o *gap* ocasionado pelas garantias adotadas em modelos clássicos de SOC, visto que as métricas propostas no contexto da similaridade semântica e de mineração de dados, visando a descoberta do conhecimento, podem apoiar simultaneamente a recuperação da informação e a atualização da estrutura semântica adotada.

Finalmente, cabe ressaltar que as alterações e as novas dinâmicas exigidas para a atualização dos SOC em tempos de identidade algorítmica não se vinculam apenas às *affordances* tecnológicas, mas exigem um olhar menos ortodoxo na identificação dos contextos de estudo e observação e na adoção de fundamentos teóricos e métodos de abordagem no campo da Organização do Conhecimento.

4. Conclusão

As origens da OC, pese embora o traçado difícil e complexo, situam-se, essencialmente, entre a segunda metade do século XIX, nos estudos de Biblioteconomia e de Documentação, até o final do século XX, fortemente influenciada por outras áreas como a Linguística, a RI e a Informática, impactadas pelos avanços irreversíveis provocados pela tecnologias de informação e pela Internet, gerando novos ambientes de interação e conteúdos digitais em áreas científicas com as quais a OC partilha uma significativa afinidade.

O objeto de estudo da OC é bipartido: organização e conhecimento. O conhecimento é visto em duas acepções: a) individual e intangível e, sob essa perspectiva, não organizável; b) social e coletivo e, sob essa perspectiva, tangível e passível de ser objetivado pela OC.

As abordagens em OC podem ser resumidas a duas perspectivas principais de orientação das pesquisas no campo: a) sentido estrito e b) sentido lato. No sentido estrito, limitam-se ao uso de teorias e de posturas epistemológicas viradas à produção de SOC, usados, principalmente, no âmbito da *Library and Information Science*. No sentido lato, extrapolam-se esses limites, avançando-se para outras áreas, campos e sistemas conceituais existentes em variadas esferas da sociedade. No interior da CI, a finalidade principal ao estudar a OC seria desenvolver melhores serviços de informação, mas podemos, naturalmente, concluir que a OC pode oferecer um contributo efetivo às outras áreas do conhecimento com as quais mantém ligações. O ensino e a pesquisa em OC tem de considerar também as potencialidades desta vertente.

Os métodos usados em OC são diversos e dependem, sensivelmente, da natureza, das posturas epistemológicas e dos relacionamentos observados dentro e fora do seu campo de estudo.

Os resultados em OC são muito variados, porque se confundem com os produtos obtidos nas suas etapas históricas, sobretudo, ao longo do século XX. Entretanto, destacam-se os chamados SOC, que serão os frutos mais visíveis dos processos de organização do conhecimento.

Evidenciou-se, neste trabalho, que dentro das tendências da aplicação da OC, mais especificamente no âmbito da Recuperação da Informação (RI), encontramos trabalhos que surgem no contexto dos estudos da similaridade semântica e no desenvolvimento dos sistemas de descoberta do conhecimento.

Futuros estudos poderiam ser desenvolvidos com o intuito de sistematizar e aprofundar as análises sobre quais seriam essas relações da OC com outras áreas (OC no sentido amplo), especialmente no que diz respeito aos seus desenvolvimentos atuais no que tange às aplicações e aos produtos produzidos a partir dessa fertilização cruzada.

Referências Bibliográficas

- Allen, B. (1996). *Information tasks: toward a user-centered approach to information systems*. San Diego, CA: Academic Press.
- Almeida, C. C., Grácio, M. C. C., & Castanha, R. G. (2017). Colaboração científica na temática Ética em Organização e Representação do Conhecimento: um estudo no periódico Knowledge Organization. In F. A. Pinho & J. A. C. Guimarães (Org.), *Memória, tecnologia e cultura na organização do conhecimento* (pp. 334-340). Recife: UFPE. Recuperado de <http://isko.marilia.unesp.br/index.php/isko/isko/schedConf/presentations?searchInitial=G&track=>
- Béjar, J. (2019). *Knowledge discovery* [PowerPoint Presentation]. Recuperado de <http://www.cs.upc.edu/~bejar/URL/material/01-DataMiningIntro.pdf> .
- Boutain, E., Gallezot, G., & Duvernay, D. (2008). Co-publications scientifiques, analyse des réseaux latents. In *XVIème Congrès de la SFSIC, Les sciences de l'information et de la communication: affirmation et pluralité, Compiègne, France*. Recuperado de http://archivesic.ccsd.cnrs.fr/docs/00/34/17/46/PDF/SFSIC_2008-Boutin_Gallezot_Duvernay.pdf.
- Brascher, M., & Café, L. (2008). Organização da informação ou organização do conhecimento? In *Anais do IX Encontro Nacional de Pesquisa em Ciência da*

- Informação (ENANCIB)*, São Paulo, Brasil. Recuperado de <http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142>.
- Broughton, V. (2008). Henry Evelyn Bliss: the other immortal, or a prophet without honour? *Journal of Librarianship and Information Science*, 40(1), 45-58. doi:10.1177/0961000607086620
- Buckland, M. (1995). *Information and information systems*. New York, NY: Praeger.
- Chaudhry, A. S. (2016). Re-conceptualization of knowledge organization: imperatives of networked resources and digitization. *International Journal of Knowledge Content Development & Technology*, 6(2), 93-108. doi:10.5865/IJKCT.2016.6.2.093
- Cheney-Lippold, J. (2011). A new algorithmic identity soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), 164-181. doi:10.1177/0263276411424420
- Choo, C. W. (1996). The knowing organization: how organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, 16(5), 329-340. doi: 10.1016/0268-4012(96)00020-5
- Dahlberg, I. (2006). Knowledge organization: a new science? *Knowledge Organization*, 33(1), 11-19.
- Ferreira, G. M., & Guimarães, J. A. C. (2011). Análise da presença e articulação da concepção teórica da catalogação de assunto na literatura científica da International Society Knowledge Organization - ISKO: uma análise de domínio. *Revista EDICIC*, 1(4), 188-207. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=3867021>
- Fujita, M. L. S. (2008). Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 1(1), 1-32. Recuperado de <https://www.brapci.inf.br/index.php/article/view/000007781/4b1334ae427706488f783f9454eab8e5>
- Gomes, H. E. (2009). Tendências da pesquisa em organização do conhecimento. *Pesquisa Brasileira em Ciência da Informação*, 2(1), 60-88. Recuperado de <https://brapci.inf.br/index.php/article/view/0000007771/a9da848c/>
- Guimarães, J. A. C., Santos, A. Y., Cândido, G. G., & Pinho, F. A. (2014). A dimensão temática da pesquisa em organização do conhecimento. *Scire*, 20(2), 19-25. Recuperado de <https://www.ibersid.eu/ojs/index.php/scire/article/download/4177/3797/0>
- Gunjal, B. (2012). *Knowledge organisation systems in digital libraries a case of ETDS* (Doctoral thesis in Library and Information Science, Department of Studies in Library and Information Science, University of Mysore, Karnataka, India). Recuperado de <http://hdl.handle.net/10603/76438>
- Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2), 87-111.
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2-3), 86-101. doi:10.5771/0943-7444-2008-2-3-86
- Hjørland, B. (2013). Theories of knowledge organization: theories of knowledge. *Knowledge Organization*, 40(3), 169-181. doi:10.5771/0943-7444-2013-3-169

- Hjørland, B. (2016). Knowledge organization (KO). *Knowledge Organization*, 43(6), 475-484. doi:10.5771/0943-7444-2016-6-475
- Lima, L. S. (2015). *Produção científica em organização do conhecimento: uma análise de domínio via citações de autores* (Dissertação de mestrado em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, Brasil). Recuperado de <http://hdl.handle.net/11449/128079>
- Pando, D. A., & Almeida, C. C. de (2015). Organização da informação e do conhecimento no contexto da Ciência da informação: da análise terminológica à reflexão epistemológica. In *XII Congreso ISKO España y II Congreso ISKO España-Portugal: Organización del conocimiento para sistemas de información abiertos* (pp. 556-569). Murcia: Universidad de Murcia. Recuperado de http://www.iskoiberico.org/wp-content/uploads/2015/11/ISKOEP2015_Actas.pdf
- Sandri, E. (2013). La sérendipité sur Internet: égarement documentaire ou recherche créatrice? *Cygne Noir*, 1, 1-17. Recuperado de <http://www.revuecygnoir.org/numero/article/la-serendipite-sur-internet>
- Serres, A. (2002). *Quelle(s) problématique(s) de la trace?* Texte d'une communication prononcée lors d'un séminaire du CERCOR (actuellement CERSIC), le 13 décembre 2002, sur la question des traces et des corpus dans les recherches en Sciences de l'Information et de la Communication. Recuperado de https://archivistic.ccsd.cnrs.fr/sic_00001397/document
- Serres, A. (2012). Problématiques de la trace à l'heure du numérique. *Sens-Dessous*, 10(1), 84-94. Recuperado de <https://www.cairn.info/revue-sens-dessous-2012-1-page-84.htm>.
- Slimani, T. (2013). Description and evaluation of semantic similarity measures approaches. *International Journal of Computer Applications*, 80(10), 25-33. Recuperado de <https://arxiv.org/abs/1310.8059>
- Smiraglia, R. P. (2011). Domain coherence within knowledge organization: people, interacting theoretically, across geopolitical and cultural boundaries. In P. McKenzie, C. Johnson, & S. Stevenson (Eds.), *Exploring Interactions of People, Places and Information: Proceedings of the 39th Annual CAIS/ACSI Conference, University of New Brunswick, Fredericton, N. B., Canada*. doi:10.29173/cais601
- Smiraglia, R. P. (2012). Knowledge organization: some trends in an emergent domain. *El Profesional de la Información*, 21(3), 225-227. doi:10.3145/epi.2012.may.01
- Van den Heuvel, C., & Smiraglia, R. P. (2010). Concepts as particles: metaphors for the universe of knowledge. In C. Gnoli & F. Mazzocchi (Eds.), *Paradigms and conceptual systems in Knowledge Organization: Proceedings of the 11th International ISKO Conference, Rome, Italy* (pp. 50-56). Würzburg: Ergon-Verlag.

MARIA DA GRAÇA SIMÕES foi Professora Auxiliar da Faculdade de Letras da Universidade de Coimbra, no Curso de Ciência da Informação Arquivística e Biblioteconómica, onde assegura as cadeiras relacionadas com a Organização do conhecimento (desde 1998). Doutorou-se em Ciência da Informação pela Universidade de Salamanca (2010), Mestrado (2007) em Ciência da Informação pela Faculdade de Letras da Universidade de Coimbra, Pós Graduação em Ciências Documentais pela Faculdade de Letras da Universidade de Coimbra (1991), Pós Graduação em Estudos Europeus pela Faculdade de Direito da Universidade de Coimbra (1993) e Licenciatura em História pela mesma Universidade (1989). Foi autora de diversas monografias e artigos publicados em revistas nacionais e internacionais, essencialmente, sobre Representação da informação e organização do conhecimento.

GERCINA ÂNGELA DE LIMA é Pós-doutora em Estudos sobre metodologias para o processo de indexação de documentos textuais, pelo Departamento de Biblioteconomia y Documentación, Facultad de Humanidades, Comunicación y Documentación, Universidad Carlos III, Madrid, Espanha, e pela Escola de Artes e Comunicação da USP. Doutora em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação PPGCI-ECI/UFMG. Mestre em Science in Library Service - Clark Atlanta University, USA. Bibliotecária pela Escola de Biblioteconomia da UFMG. Professora Titular no Departamento de Organização e Tratamento da Informação, na ECI/UFMG. Bolsista de Produtividade (PQ) do CNPq, Nível 2. Bolsista do Programa Pesquisador Mineiro- PPM/ FAPEMIG-MG. Coordenadora do Grupo de Pesquisa MHTX - Modelagem Conceitual para Organização e Representação da Informação Hipertextual, registrado no CNPq desde 2004. Pesquisadora no Programa de Pós-Graduação em Gestão & Organização do Conhecimento-PPGGOC/ECI/ UFMG. Atualmente é Membro do Scientific Advisory Board da International Society for Knowledge Organization (ISKO), e do Centro de Estudos Interdisciplinares do Século XX da Universidade de Coimbra, Portugal

Série Investigação

•

Imprensa da Universidade de Coimbra

Coimbra University Press

2020

1 2



9 0

