

Do tratamento à

Organização da **Informação**

Reflexões sobre concepções,
perspectivas e tendências

Maria Graça Simões
Gercina Ângela de Lima
COORDENAÇÃO

IMPRESA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

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

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” (Bocato & 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: tesauros, sistemas de classificações, taxonomias, listas de cabeçalhos de assunto e ontologias.

2.1. Tesauros

O tesauro 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 tesauro 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 tesauros 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 tesau-ros 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 tesouro, 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 sígnica 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 morfologicamente, 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).

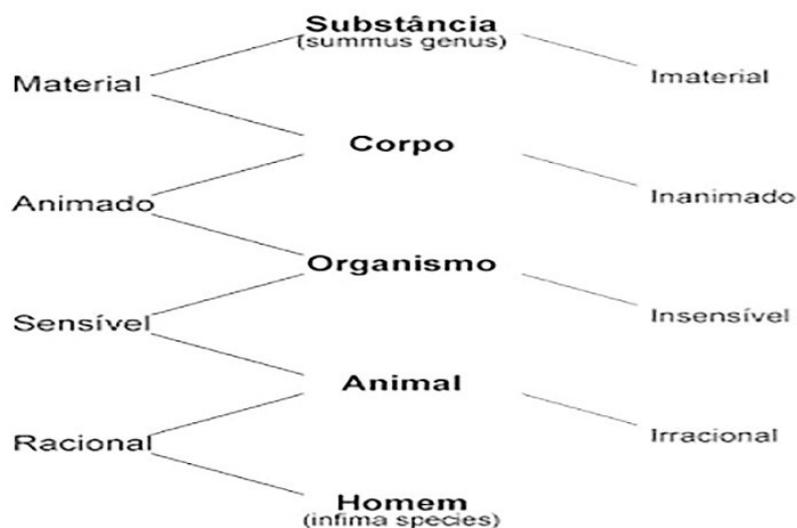


Figura 1: Árvore de Porfírio. Fonte: Burke, 2003, p. 82.

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 herdam 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 sógnica 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/bitl/dahlbergteoria/dahlberg_teorla.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.ibernid.eu/ojs/index.php/scire/article/view/4577>
- Gardin, N. (1969). Le lexique intermédiaire: Un nouveau pas vers la coopération 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/bitl/cabecalho/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 reengenhearia 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