Curso de Bacharelado em Biblioteconomia na Modalidade a Distância

Dilza Fonseca da Motta

Instrumentos de Representação Temática da Informação II

Semestre

4

Curso de Bacharelado em Biblioteconomia na Modalidade a Distância

Dilza Fonseca da Motta

Instrumentos de Representação Temática da Informação II

Semestre

4

Brasília, DF

CAPES

UAB UNIVERSIDADE ABERTA DO BRASIL

Rio de Janeiro

UFRJ faz 100 ANOS 1920 | 2020

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Faculdade de Administração e Ciências Contábeis

Departamento de Biblioteconomia

2018

Permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam o devido crédito ao autor e que licenciem as novas criações sob termos idênticos.

Presidência da República

Ministério da Educação

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Diretoria de Educação a Distância (DED)

Sistema Universidade Aberta do Brasil (UAB)

Universidade Federal do Rio de Janeiro (UFRJ)

Núcleo de Educação a Distância (NEAD)

Faculdade de Administração e Ciências Contábeis (FACC)

Departamento de Biblioteconomia

Leitor
Marcos Luiz Cavalcanti de Miranda

Comissão Técnica
Célia Regina Simonetti Barbalho
Helen Beatriz Frota Rozados
Henriette Ferreira Gomes
Marta Lígia Pomim Valentim

Comissão de Gerenciamento
Mariza Russo (*in memoriam*)
Ana Maria Ferreira de Carvalho
Maria José Veloso da Costa Santos
Nadir Ferreira Alves
Nysia Oliveira de Sá

Equipe de apoio
Eliana Taborda Garcia Santos
José Antonio Gameiro Salles
Maria Cristina Paiva
Miriam Ferreira Freire Dias
Rômulo Magnus de Melo
Solange de Souza Alves da Silva

Coordenação de Desenvolvimento Instrucional
Cristine Costa Barreto

Desenvolvimento instrucional
Kathleen da Silva Gonçalves

Diagramação
André Guimarães de Souza

Revisão de língua portuguesa
Licia Matos

Projeto gráfico e capa
André Guimarães de Souza
Patricia Seabra

Normalização
Dox Gestão da Informação

M921i Motta, Dilza Fonseca da.
Instrumentos de representação temática da informação II / Dilza Fonseca da Motta ; [leitor] Marcos Luiz Cavalcanti de Miranda. - Brasília, DF : CAPES : UAB ; Rio de Janeiro, RJ : Departamento de Biblioteconomia, FACC/UFRJ, 2018.
204 p. : il.

Inclui bibliografia.
ISBN 978-85-85229-24-5 (brochura)
ISBN 978-85-85229-25-2 (e-book)

1. Catalogação por assunto. 2. Ontologias. 3. Tesauros. I. Miranda, Marcos Luiz Cavalcanti de. II. Título.

CDD 025.4
CDU 025.4

Catalogação na publicação por: Solange Souza CRB-7 / 6646

Caro leitor,

A licença CC-BY-NC-AS, adotada pela UAB para os materiais didáticos do Projeto BibEaD, permite que outros remixem, adaptem e criem a partir desses materiais para fins não comerciais, desde que lhes atribuam o devido crédito e que licenciem as novas criações sob termos idênticos. No interesse da excelência dos materiais didáticos que compõem o Curso Nacional de Biblioteconomia na modalidade a distância, foram empreendidos esforços de dezenas de autores de todas as regiões do Brasil, além de outros profissionais especialistas, a fim de minimizar inconsistências e possíveis incorreções. Nesse sentido, asseguramos que serão bem recebidas sugestões de ajustes, de correções e de atualizações, caso seja identificada a necessidade destes pelos usuários do material ora apresentado.

LISTA DE FIGURAS

Figura 1 - Buscando informação 16

Figura 2 - Correspondência entre assuntos dos documentos e termos de indexação 18

Figura 3 - O processo de comunicação em ambientes informacionais 20

Figura 4 - Cena da entrevista 22

Figura 5 - Linguagens de indexação: LN e LD 25

Figura 6 - Diferenças básicas entre LN e LD 26

Figura 7 - Estudando as LD 27

Figura 8 - Cartão perfurado de *Herman Hollerith* 30

Figura 9 - Funções do vocabulário controlado 31

Figura 10 - O vocabulário controlado e suas regras 34

Figura 11 - Listas de cabeçalhos de assunto da *Biblioteca do Congresso dos Estados Unidos* 47

Figura 12 - *Real Gabinete Português de Leitura* (RJ) 48

Figura 13 - Parte do egípcio *Livro dos Mortos*, escrito em papiro 48

Figura 14 - Obra do artista flamengo *Jan van Eyck* (século XV) 49

Figura 15 - *Charles Cutter* 53

Figura 16 - Cena da reportagem sobre Sabino - SP 54

Figura 17 - Saber mais um pouco nunca é demais... 55

Figura 18 - Catálogo de fichas 56

Figura 19 - Línguas diferentes, sintaxes diferentes 59

Figura 20 - Tem tanta coisa que é raiz! Onde encaixar este livro?.... 62

Figura 21 - *Sherlock Holmes* 63

Figura 22 - Ordem alfabética: vantagem ou desvantagem na busca de informação? 65

Figura 23 - Página inicial do *site* da Biblioteca Digital do *Senado Federal* 71

Figura 24 - Manual da *LCSH* 72

Figura 25 - Fachada da *Biblioteca Nacional* do Rio de Janeiro 77

Figura 26 - Quem procura sempre encontra! 79

Figura 27 - Crítica pela crítica, NÃO! 81

Figura 28 - Pesquisando assuntos na *web* 85

Figura 29 - Cabeçalhos de assunto na internet 87
Figura 30 - Tesauro 95
Figura 31 - Tesouro ou tesauro? 96
Figura 32 - A palavra certa tem seu lugar! 97
Figura 33 - *Peter Mark Roget* criou o dicionário de ideias afins 98
Figura 34 - Invenções pós-Segunda Guerra Mundial 99
Figura 35 - Buscando e recuperando informações 100
Figura 36 - Navegando em rede 109
Figura 37 - Teoria e prática completam-se 111
Figura 38 - Qual o conceito de "manga"? 116
Figura 39 - Triângulo conceitual 117
Figura 40 - O conhecimento do objeto e suas propriedades permitiu a definição de baleia 118
Figura 41 - Características extrínsecas e intrínsecas dos conceitos .. 119
Figura 42 - Cena de vídeo sobre a produção do queijo minas 124
Figura 43 - Tesauro de Cultura Material dos Índios no Brasil 127
Figura 44 - Avaliar mais para errar menos 131
Figura 45 - Como escolher o software adequado ao meu tesauro? 133
Figura 46 - Não use gato por lebre! 134
Figura 47 - Tirando dúvidas 143
Figura 48 - Histórias são reveladoras... 144
Figura 49 - Cena do filme *A.I.: Inteligência Artificial* 145
Figura 50 - *Web* semântica: a resposta certa para cada pergunta .. 146
Figura 51 - Conceito e ontologias 147
Figura 52 - Vamos organizar o conhecimento? 149
Figura 53 - Problema? Olha a solução! 150
Figura 54 - *Aristóteles*: o primeiro autor sobre lógica 151
Figura 55 - Funcionalidade das ontologias 153
Figura 56 - Folksonomia: seus termos são chamados de *tags* 156
Figura 57 - Biblioteca 161
Figura 58 - Ontologia descrita com *OWL* 162
Figura 59 - Extra! Extra! *Software* para construir ontologias! 164
Figura 60 - O mundo conectado 173

Figura 61 - Reciclando o cabeçalho de assunto 174

Figura 62 - Repensando a recuperação da informação 176

Figura 63 - LD: inovar para melhorar..... 183

Figura 64 - *MeSH*: uma experiência americana na área médica 184

Figura 65 - O que é *MeSH*?..... 185

Figura 66 - Dados de um descritor do *MeSH*..... 186

Figura 67 - Instrumento de controle de assunto 187

Figura 68 - Trabalhar em rede é um grande negócio! 189

Figura 69 - Página de busca do *Nuovo Soggettario Thesaurus*..... 190

Figura 70 - Forma sistemática de apresentação dos termos do *Soggettario* 191

Figura 71 - Qual o conceito de "parada"? 192

Figura 72 - Modelo de etiquetagem (*tagging*)..... 193

Figura 73 - *Tags*..... 195

Figura 74 - *F. W. Lancaster*..... 197

Figura 75 - Não assuma que um tamanho serve para tudo! 199

LISTA DE QUADROS

Quadro 1 - Relação de assuntos para cabeçalhos de fichas 37

Quadro 2 - Lista alfabética com controle de sinônimos 38

Quadro 3 - Vocabulário de entrada e efeitos na recuperação da informação..... 39

Quadro 4 - As cinco leis da Biblioteconomia (*Ranganathan*)..... 51

Quadro 5 - Teorias para a construção de tesauro com base em conceito 103

Quadro 6 - Norma *ISO 25964* 111

Quadro 7 - Categorização do assunto "artesanato" 114

Quadro 8 - Categorias do *Classification Research Group*..... 115

Quadro 9 - Características de divisão..... 121

Quadro 10 - Etapas do planejamento de um tesauro..... 125

Quadro 11 - Categorias 130

Quadro 12 - Aplicação de ontologias em diversas áreas 154

Quadro 13 - Tipos de ontologias..... 156

Quadro 14 - Metodologias para a construção de ontologias 159

Quadro 15 - Metodologias para a construção de ontologias em grupo 160

Quadro 16 - Ferramentas para a construção de ontologias..... 164

Quadro 17 - Relação entre termos de áreas do conhecimento diversas 178

Quadro 18 - Como surgiu o *MeSH*?..... 184

Quadro 19 - Folksonomia: vantagens e desvantagens 194

Quadro 20 - Síntese comparativa de características do Vocabulário Sistematizado, do *Soggettario*, do *MeSH* e das folksonomias..... 198

SUMÁRIO

APRESENTAÇÃO 13

1 UNIDADE 1: REPRESENTAÇÃO TEMÁTICA E LINGUAGEM DOCUMENTÁRIA: QUAL A RELAÇÃO? 15

1.1 OBJETIVO GERAL 15

1.2 OBJETIVOS ESPECÍFICOS 15

1.3 PRÉ-REQUISITO 15

1.4 REPRESENTAR TEMAS... PARA QUÊ? 17

1.4.1 **Representação temática e indexação** 17

1.4.2 **Atividade** 19

1.5 PROCURANDO COMPREENDER UMA LINGUAGEM DOCUMENTÁRIA (LD) 20

1.5.1 **Para início de conversa: quem não se comunica...** 20

1.5.2 **Já teve oportunidade de conhecer uma LD?** 22

1.5.3 **Linguagem documentária: que linguagem é essa?** 24

1.5.4 **Linguagem documentária: ela é conhecida por tantos outros nomes!** 26

1.5.5 **Atividade** 28

1.6 VOCABULÁRIO DE UMA LD 29

1.6.1 **Funções do vocabulário controlado** 31

1.6.2 **Objetivos do vocabulário controlado** 33

1.6.3 **Como utilizar o vocabulário?** 34

1.6.4 **Atividade** 35

1.6.5 **Construindo um vocabulário controlado** 37

1.6.6 **Atividade** 40

RESUMO 41

SUGESTÃO DE LEITURA 42

REFERÊNCIAS 43

2 UNIDADE 2: REFLEXÕES SOBRE O CABEÇALHO DE ASSUNTO 45

2.1 OBJETIVO GERAL 45

2.2 OBJETIVOS ESPECÍFICOS 45

2.3 PRÉ-REQUISITO 45

2.4 LISTAS DE CABEÇALHOS DE ASSUNTO 47

2.5 TUDO TEM UMA HISTÓRIA 48

2.5.1 **Cabeçalho de assunto: o que é?** 52

2.5.2 **Como surgiu o cabeçalho de assunto?** 52

2.5.3 **A sistematização do cabeçalho de assunto** 53

2.5.4 **Princípios de *Cutter* para cabeçalhos de assunto** 57

2.5.5 **Atividade** 63

2.5.6 **Tipologia dos cabeçalhos de assunto** 65

2.5.7 **Características das listas de cabeçalhos de assunto** 65

2.5.8 **Funções do cabeçalho de assunto** 68
2.5.9 **Atividade** 68
2.6 CABEÇALHOS DE ASSUNTO NA PRÁTICA 69
2.6.1 **Elaboração de cabeçalhos de assunto** 70
2.6.2 **Panorama dos cabeçalhos na *LCSH*** 73
2.6.3 **Situação dos cabeçalhos de assunto no Brasil** 77
2.7 AVALIAÇÃO DOS CABEÇALHOS DE ASSUNTO DA *LCSH* 79
2.8 O QUE EXISTE ALÉM DA *LCSH*? 82
2.8.1 **Atividade** 83
2.9 OS CABEÇALHOS DE ASSUNTO EM TEMPOS DE *WEB*, COMO ESTÃO? 85
RESUMO 87
INFORMAÇÕES SOBRE A PRÓXIMA UNIDADE 88
PRÉ-REQUISITOS 88
SUGESTÃO DE LEITURA 89
REFERÊNCIAS 89
3 **UNIDADE 3: TESAURO** 93
3.1 OBJETIVO GERAL 93
3.2 OBJETIVOS ESPECÍFICOS 93
3.3 VOCÊ JÁ OUVIU FALAR EM TESAURO? 95
3.4 AFINAL, DE QUE SE TRATA? 95
3.4.1 **Origens do tesauro: um pouco de história** 96
3.4.2 **Dicionários ideológicos: a analogia entre as palavras** 98
3.4.3 **A trajetória do tesauro e as descobertas científicas nas guerras** 99
3.5 A BUSCA POR NOVAS FERRAMENTAS PARA A RECUPERAÇÃO DA INFORMAÇÃO 100
3.6 TESAURO DOCUMENTÁRIO 102
3.6.1 **O desenvolvimento dos tesauros nos continentes** 102
3.6.2 **Teorias que orientam a construção de tesauros** 103
3.6.3 **Componentes de um tesauro documentário** 104
3.6.4 **E as características de um tesauro, quais são?** 105
3.6.5 **Quais as funções de um tesauro?** 106
3.6.6 **Que tipos de tesauro existem?** 106
3.6.7 **Atividade** 106
3.7 ELABORAÇÃO DE UM TESAURO DOCUMENTÁRIO: BOTANDO A MÃO NA MASSA! 109
3.7.1 **Normas para a elaboração de tesauros** 110
3.7.2 **Teorias que fundamentam a elaboração de um tesauro documentário** 111
3.7.2.1 *Teoria da Classificação Facetada* 112
3.7.2.2 *Teoria Geral da Terminologia (TGT)* 115
3.7.2.3 *Teoria do Conceito* 116
3.7.3 **Características dos conceitos e sua classificação** 118
3.7.4 **Relações entre conceitos** 122

3.7.5 **Etapas da elaboração de um tesauro** 124
3.7.6 **Atividade** 128
3.8 AVALIANDO UM TESAURO 131
3.9 AVALIANDO *SOFTWARES* PARA ELABORAÇÃO DE TESAUROS 133
3.9.1 **Atividade** 136
3.10 CONCLUSÃO 137
RESUMO 138
SUGESTÃO DE LEITURA 138
REFERÊNCIAS 139
4 **UNIDADE 4: ONTOLOGIAS** 141
4.1 OBJETIVO GERAL 141
4.2 OBJETIVOS ESPECÍFICOS 141
4.3 MAIS UMA LINGUAGEM A SERVIÇO DA RECUPERAÇÃO DA INFORMAÇÃO 143
4.4 AFINAL, O QUE SÃO ONTOLOGIAS? DE ONDE VEM ESSA PALAVRA? 143
4.4.1 **Ontologias: um pouco de história** 144
4.4.2 **Quais as características das ontologias?** 148
4.4.3 **Componentes essenciais** 148
4.4.4 **Atividade** 152
4.5 É PARA COMER OU PASSAR NO CABELO? PARA QUE SERVEM AS ONTOLOGIAS? 153
4.6 ONTOLOGIAS: CONHECENDO SEUS TIPOS 155
4.6.1 **Atividade** 158
4.7 COMO ELABORAR ONTOLOGIAS? METODOLOGIAS, LINGUAGENS E FERRAMENTAS 159
4.7.1 **Metodologias para a construção de ontologias** 159
4.7.2 **OWL: uma linguagem para a representação de ontologias** 162
4.7.3 **Uma boa notícia: ferramentas para a construção de ontologias** 164
4.8 E AGORA, O QUE FAZER? DICAS PARA INICIAR A CONSTRUÇÃO DE UMA ONTOLOGIA 166
4.8.1 **Atividade** 167
4.9 CONCLUSÃO 168
RESUMO 169
SUGESTÃO DE LEITURA 169
REFERÊNCIAS 169
5 **UNIDADE 5: LINGUAGENS DE INDEXAÇÃO E RECUPERAÇÃO EM PRÁTICA** 171
5.1 OBJETIVO GERAL 171
5.2 OBJETIVOS ESPECÍFICOS 171
5.3 MIL CONEXÕES POR SEGUNDO 173
5.4 APRESENTAÇÃO DE OUTRAS LINGUAGENS DE INDEXAÇÃO/ RECUPERAÇÃO DA INFORMAÇÃO JÁ APLICADAS EM AMBIENTES INSTITUCIONAIS 173
5.5 CABEÇALHO DE ASSUNTO E NOVAS LINGUAGENS 174

5.6 *VOCABULÁRIO SISTEMATIZADO* 176
5.6.1 **Características do *Vocabulário Sistematizado*** 178
5.6.2 **Objetivos do *Vocabulário Sistematizado*** 179
5.6.3 **Componentes do *Vocabulário Sistematizado*** 179
5.6.4 **Apresentação do *Vocabulário Sistematizado*** 180
5.6.5 **Atividade** 181
5.6.6 **Elaboração de *Vocabulário Sistematizado*** 182
5.7 PARA ONDE CAMINHAM AS LINGUAGENS DOCUMENTÁRIAS? 183
5.7.1 ***MeSH*** 184
5.7.1.1 *Objetivos* 185
5.7.1.2 *Caracterização* 185
5.7.1.3 *Componentes* 187
5.7.2 ***Nuovo Soggettario*: uma experiência na Itália** 187
5.7.2.1 *Objetivos* 188
5.7.2.2 *Caracterização* 188
5.7.2.3 *Componentes* 188
5.7.2.4 *Estágios de elaboração e aplicação do Soggettario* 189
5.7.2.5 *O Nuovo Soggettario em prática* 190
5.7.2.6 *Origem da sintaxe* 191
5.7.2.7 *Forma de apresentação do* Nuovo Soggettario 191
5.7.3 **Folksonomia: um instrumento de indexação colaborativo** 191
5.7.3.1 *Objetivos* 193
5.7.3.2 *Caracterização* 193
5.7.3.3 *Vantagens e desvantagens da folksonomia* 194
5.7.3.4 *Outras considerações* 196
5.7.4 **Atividade** 196
5.8 ESCOLHENDO UMA LINGUAGEM ADEQUADA PARA CADA CONTEXTO DE UM SRI 197
5.8.1 **Atividade** 200
5.9 CONSIDERAÇÕES FINAIS 201
5.10 CONCLUSÃO 202
RESUMO 202
SUGESTÃO DE LEITURA 203
REFERÊNCIAS 204

APRESENTAÇÃO

A recuperação da informação acontece após o tratamento de documentos em bibliotecas e outros serviços do gênero. Esse tratamento obedece a um ciclo e implica diversas etapas, como análise do documento, interpretação de seu conteúdo, condensação dos principais conceitos envolvidos e representação desses conceitos em uma linguagem para, então, a informação ser recuperada. Todas essas etapas têm um objetivo: organizar o conhecimento produzido e registrado em diversos tipos de documentos.

A disciplina "Instrumentos de Recuperação Temática da Informação II" diz respeito a ferramentas auxiliares para a indexação/recuperação de documentos, considerando seu conteúdo temático. Porém, no decorrer do curso todo, você estudará outras disciplinas também preocupadas com a organização do conhecimento, mas que abordarão aspectos diferentes da questão. Certamente você fará correlações com disciplinas como "Processos e Produtos de Representação Temática da Informação", "Análise da Informação", "Políticas de Organização e Representação da Informação", "Elementos Lógicos e Linguísticos na Organização e Representação da Informação" e, é claro, "Instrumentos de Representação Temática da Informação I", esta, mais próxima de nossa disciplina.

Um sistema de recuperação de informação é consultado sempre que se deseja satisfazer uma necessidade de informação. As formas de se fazer uma busca são várias, como também são variados os pontos de acesso para se chegar a um documento. E ambos dependem tanto dos objetivos de cada sistema como da necessidade de cada usuário. Mas é preciso que esses pontos de acesso sejam previstos desde o início do desenho dos sistemas. Assim, pode-se recuperar a informação pretendida.

Houve uma época em que o único ponto de acesso a um documento era o título, quando os registros ainda eram feitos em pedaços de pedras e colocados nas paredes dos mosteiros. É claro que a situação hoje é bastante diferente. Pode-se recuperar uma informação pelo título, pelo autor, pela data, pela cor, pela marca de um produto, enfim, por tantos aspectos quantos o sistema ache necessário para que o cliente fique satisfeito em sua busca. Tente o *Google* e confirme!

Em meio a tantas opções, existe uma maneira de representar e buscar informação que parece importantíssima para a maioria dos usuários dos sistemas, além de muito demandada atualmente: a busca por assunto. Quando se acessa um sistema, é comum querer recuperar documentos que tratem de determinados assuntos, que falem sobre eles, que digam respeito a um tema.

Nas bibliotecas e outros serviços de recuperação da informação, comumente encontram-se instrumentos desenvolvidos com o objetivo de representar e recuperar os assuntos contidos nos documentos. Esses instrumentos são chamados de linguagens documentárias. Atualmente, existem outras linguagens que extrapolam a indexação de documentos, servindo também como instrumento auxiliar no processamento de respostas mais precisas a perguntas pontuais dos usuários de um sistema de busca, como são as ontologias. Essas linguagens não se baseiam estritamente em palavras, mas se valem também de outros mecanismos, como regras de inferência e códigos computacionais. Elas também serão abordadas nesta disciplina.

A denominação "linguagem" é usada por se tratar de sistemas de símbolos usados na comunicação. Há linguagens documentárias que expressam seus assuntos por meio de notações formadas a partir de símbolos não linguísticos (números ou números e letras). São as chamadas linguagens notacionais, já abordadas na disciplina "Instrumentos de Representação Temática da Informação I". E há as que

representam seus assuntos por meio de palavras – as linguagens documentárias verbais, objeto de nossa disciplina.

As linguagens documentárias verbais são valorizadas numa busca por informação, na medida em que os assuntos nelas contidos são representados por palavras, símbolo usado pelo homem para se comunicar socialmente. Isso aproxima o sistema do usuário, tornando a comunicação mais fácil. Daí a importância da disciplina "Instrumentos de Representação Temática da Informação II", cujo objetivo é explorar o universo desses instrumentos.

Resumidamente, a proposta desta disciplina é contextualizar historicamente o aparecimento das linguagens de representação temática da informação e seu desenvolvimento, apresentar os fundamentos teóricos e práticos de cada uma delas, bem como os principais conceitos, princípios e normas que as envolvem.

Além disso, ao longo da apresentação da disciplina, você encontrará, intercaladas no texto, notas biográficas dos principais autores da área, referências a outros textos e mídias e explicações expandidas, caso queira aprofundar seu conhecimento. Não nos esquecemos de incluir também curiosidades, algumas delas bem interessantes e até inimagináveis!

Como o objetivo do curso é transmitir o conhecimento básico gerado em Biblioteconomia e Ciência da Informação, os exercícios de verificação e respectivas respostas não foram deixados de lado, para que você possa avaliar o que apreendeu.

Enfim, procuramos apresentar "Instrumentos de Representação Temática da Informação II" da forma mais atrativa possível, para que você aproveite ao máximo esta disciplina.

Pensamos em você o tempo todo. Mãos à obra! Sucesso!

UNIDADE 1

REPRESENTAÇÃO TEMÁTICA E LINGUAGEM DOCUMENTÁRIA: QUAL A RELAÇÃO?

1.1 OBJETIVO GERAL

Identificar a linguagem documentária (LD) como instrumento auxiliar do processo de representação temática em um sistema de recuperação de informação (SRI)[1], sua caracterização e influência na qualidade da recuperação da informação.

1.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

a) identificar o conceito de indexação e representação temática;

b) reconhecer o conceito de LD, sua trajetória histórica, função, componentes e tipos;

c) reconhecer o vocabulário como um dos componentes da LD e sua importância na entrada e saída dos dados de um SRI.

1.3 PRÉ-REQUISITO

Seria interessante você acessar a *web* e navegar em alguns *sites* de busca, como o *Google*, tentar pesquisar sobre assuntos de seu interesse e comparar os assuntos pesquisados com as respostas que obtiver nas buscas. Observe se as respostas foram satisfatórias, se recuperou muitas informações inúteis, e não deixe de atentar para o tempo gasto nas buscas.

[1] Neste documento, a expressão "linguagem documentária" será referida, também, como LD e "sistema de recuperação de informação", também como SRI.

Figura 1 - Buscando informação

Fonte: *Pixabay* (2017).[2]

[2] Disponível em: <https://pixabay.com/pt/biblioteca-livros-menina-2128813/>.

1.4 REPRESENTAR TEMAS... PARA QUÊ?

Como você viu na disciplina "Instrumentos de Representação Temática I", o conhecimento produzido pela sociedade precisa ser representado para ser transferido a outras pessoas. Essas representações são recortes da realidade e necessitam de padrões que possam comunicá-las.

Naquela disciplina, você aprendeu as linguagens documentárias notacionais, que são padrões para a representação de assuntos, por meio de notações formadas por diversos símbolos: números, números e letras, e outros sinais. Porém, na descrição de um documento para sua identificação e recuperação de seu conteúdo, a representação de outros aspectos é necessária. Um dos pontos de acesso ao documento mais procurados pelos usuários de um SRI é o assunto.

Em "Instrumentos de Representação Temática II", você aprenderá outra forma de representação dos documentos, por meio da linguagem verbal, que é mais um sistema de signos adequado para representar o conteúdo temático do documento.

1.4.1 Representação temática e indexação

Antes mesmo de falarmos das linguagens documentárias verbais, algumas explicações merecem ser revistas. Uma delas diz respeito ao que se entende por "representação temática".

Segundo o *Dicionário Eletrônico Houaiss da Língua Portuguesa* (Versão 3.0), a palavra "representação" tem várias acepções, mas quando aplicada à área de informação, que é nosso caso, "representar" pode ser conceituado da seguinte maneira: "aparecer numa outra forma, substituir, estar no lugar de; fazer as vezes de". Já a palavra "temática" tem a ver com tema, assunto (HOUAISS, 2009). Agora, podemos conceituar representação temática da seguinte forma:

> Representação temática é a apresentação dos temas tratados em um **documento** de outra forma – por meio de palavras ou expressões que substituam esses temas.

Atenção

Em nossas aulas, entenderemos "documento" como qualquer objeto de valor documental (livros, artigos de periódicos, patentes, fotografias, peças, papéis, filmes, construções etc.) contendo informações que elucidem, instruam, provem ou comprovem cientificamente algum fato, acontecimento, dito etc. Sempre que nos referirmos a documento, estaremos, então, tratando de itens de informação de uma coleção. Atualmente, a palavra usada em subs-

tituição a "documento" é "recurso informativo". Na internet, por exemplo, é comum encontrarem-se expressões como "descoberta de recursos" (*resources discovery*), referindo-se a documentos.

Ora, como vamos representar o tema, ou os temas, de um documento? Escolhendo palavras ou expressões que substituam, que fiquem no lugar dos assuntos tratados nesse documento. Esses substitutos nada mais são do que pontos de acesso que orientam o usuário na procura da informação desejada e que, quando consultados, devem ser o caminho para chegar ao próprio documento que contém tal informação. Por exemplo, se entro em um serviço de informação e quero um documento sobre "o ensino da música nas escolas brasileiras", é quase certo que, se a base de dados tiver um documento sobre esse assunto, ela me oferecerá os seguintes pontos de acesso: "música", "escola" e "Brasil." Nesse caso, esses assuntos representariam a informação desejada e seriam os pontos de acesso que me levariam a ela.

Para que um documento seja representado em uma base de dados, primeiro é preciso indexá-lo. E o que é **indexação**? É a atribuição de um ou mais pontos de acesso ao documento. No caso da indexação temática, é a atribuição de palavras ou expressões. Por isso, a função da indexação é possibilitar que um documento seja representado em uma base de dados.

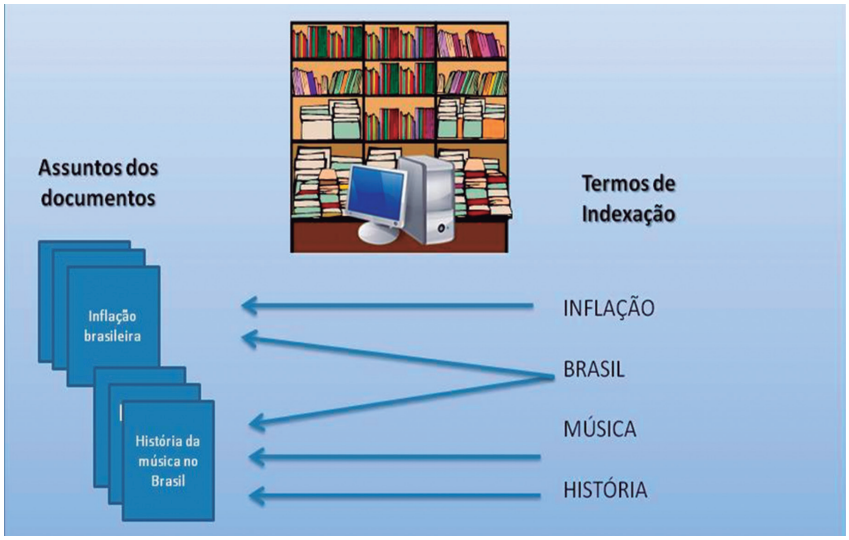
INDEXAÇÃO

De outra forma, *Lancaster* (2004, p. 1) define indexação assim: "é a construção de representações de documentos numa forma que se preste a sua inclusão numa base de dados."

Primeiro, cumprem-se todas as etapas que envolvem a indexação propriamente dita: análise do documento; seleção dos assuntos nele tratados; conversão desses assuntos em palavras ou expressões retiradas de algum vocabulário, seja da linguagem natural – a falada no dia a dia – ou de outras fontes de informação; atribuição ao documento das palavras/expressões correspondentes ao(s) assunto(s) nele tratados. Uma vez isso feito, o documento está pronto para ser representado em uma base de dados.

Na Figura 2, você poderá conferir os termos atribuídos aos assuntos tratados nos documentos. Você veria outras possibilidades? Que tal pensarmos, por exemplo, em "história da música"?

Figura 2 - Correspondência entre assuntos dos documentos e termos de indexação



Fonte: produção do próprio autor (2017).

Bem, à luz dessa explicação, estamos prontos para responder à pergunta: o que é **indexação temática**?

Indexação temática

É o processo de atribuir aos documentos palavras ou expressões que representam seus temas, para que estas sejam posteriormente incluídas em uma base de dados. Sua função é representar os assuntos de um documento e envolve as seguintes etapas: análise e interpretação do documento, seleção dos assuntos nele tratados, conversão desses assuntos em palavras ou expressões (retiradas do próprio documento ou de um vocabulário construído para tal fim) e indexação propriamente dita (correlação entre os assuntos e os documentos).

Você deve estar se perguntando: para que tudo isso? Obviamente, isso é feito com um único objetivo: atender o usuário de um SRI. Claro, todo sistema é desenvolvido tendo uma clientela em mente. Não se esqueça de que o documento é para ser usado! Uma vez atribuídos, os termos serão incluídos numa base de dados de representações dos documentos e serão recuperados posteriormente, em buscas feitas no sistema.

1.4.2 Atividade

A partir do que foi estudado até agora, enumere o que você entendeu ser a indexação e sua função em um SRI.

No quadro a seguir, comece enunciando uma dessas ideias. A partir das informações dadas no início da seção 1.4.1 ("Representação temática e indexação"), inclua, pelo menos, outras duas.

INDEXAÇÃO E SUA FUNÇÃO EM UM SRI
1. A indexação é um processo que torna possível recuperar a informação desejada.
2.
3.

Resposta comentada

Você pode ter pensado em outras ideias. Veja se elas estariam incluídas nas seguintes. Se não, são pontos sugeridos para você refletir.

a) A indexação temática faz parte de uma série de processos que envolvem o tratamento de um documento em um SRI.

b) Representar a informação contida em documentos é mais econômico do que disponibilizar a informação em texto integral.

c) Mesmo as bases de dados de textos integrais necessitam ter seus textos indexados de alguma forma, para que o usuário possa chegar até eles.

1.5 PROCURANDO COMPREENDER UMA LINGUAGEM DOCUMENTÁRIA (LD)

Todo serviço de informação se presta a dar informações a seus usuários. Então, de um lado, temos bases de dados, com informações tratadas e prontas para ser disponibilizadas. Do outro, o usuário, com suas necessidades de informação. A chave para a maior satisfação do usuário é a interatividade que deve haver entre ele e o sistema, ou seja, a capacidade que o sistema tem de se comunicar com o usuário. Quanto maior a interatividade, melhor. De nada adianta uma imensa base de dados, se o processo de comunicação é ineficaz.

1.5.1 Para início de conversa: quem não se comunica...

Figura 3 - O processo de comunicação em ambientes informacionais

Fonte: produção do próprio autor a partir de *Campos* (2001) e apostilas de aulas.

Um problema na recuperação da informação é que, comumente, o usuário chega a um serviço para buscar a informação de que necessita e faz sua pergunta ao sistema usando a linguagem natural (LN) – a linguagem corrente, falada no dia a dia, que é cheia de **polissemias**. Soma-se a isso o fato de que, muitas vezes, ele nem sabe o que quer realmente e suas buscas tornam-se insatisfatórias!

Polissemia

Multiplicidade de sentidos de uma palavra ou locução (por exemplo: "prato" pode significar: vasilha, comida, iguaria, receptáculo de balança ou instrumento musical. "Pé-de-moleque" pode ser: doce, tipo de calçamento) (HOUAISS, 2009).

De fato, muitos serviços utilizam a linguagem natural para indexar suas informações, simplesmente selecionando palavras/expressões dos textos para **atribuí-las aos documentos com o objetivo de representá-los em bases de dados**. Muito comumente, indexadores fazem essa seleção de sua própria cabeça, a partir do conhecimento que têm do assunto.

Esses processos são feitos sem que nenhum controle, seja de sinônimos, de homógrafos ou qualquer outro, seja exercido sobre a linguagem natural; é o uso livre do signo linguístico, com significados que podem variar de contexto para contexto. Talvez o maior e melhor exemplo de indexação desse tipo seja o da *web*.

É indiscutível o apoio dos recursos disponíveis na rede para a resolução de nossos problemas atuais, o que não invalida críticas quanto à qualidade da informação nela recuperada. Sem perder isso de vista, investimentos maciços vêm sendo feitos na *web* semântica, na tentativa de refazer a maneira de organizar conteúdos. Você verá mais detalhes sobre *web* semântica na parte referente às ontologias (Unidade 4).

Curiosidade

A *web* semântica é uma nova forma de tecnologia para fazer a *web* entender significados, ficando, assim, mais inteligente. A ideia é fazer uma *web* em que as informações não sejam somente estocadas, mas compreendidas pelos computadores, para trazer ao usuário aquilo que ele busca.

A *web* semântica permitirá, dessa forma, tornar o conteúdo semântico da *rede* interpretável pelo homem e pela máquina. Seus primeiros usos já estão distinguidos. Essa iniciativa é particularmente promissora nos domínios verticais (comércio, viagens, habitação, emprego, entre outros).

Por exemplo, no ramo de viagens, idealmente, o sistema da *web* semântica deve ser capaz de dar uma resposta completa para uma pergunta do tipo: "eu quero férias na Toscana neste verão. Tenho um orçamento de 4 mil euros. E nós temos um filho de 8 anos. Qual o hotel adequado?". Atualmente, responder a tais perguntas vai exigir a triagem em listas distintas de hotéis e de locação de carros. Com a *web* semântica, a solicitação chamará uma resposta coerente, meticulosamente reunida. O sistema trabalhará por você: ele classificará todos os comentários e encontrará, por dedução, o bom hotel.

Multimídia

Quer saber um pouco mais sobre *web* semântica? Acesse o *link* a seguir e assista à entrevista com a especialista *Martha Gabriel*: <https://www.youtube.com/watch?v=i4GG4etWjR8>.

Figura 4- Cena da entrevista

Fonte: *Youtube* (2009).

O computador trouxe com ele a esperança de resolução de todos os problemas relacionados à necessidade de informação, afetando rotinas de ambientes informacionais. Nos anos 1990, com a intensificação do uso de novas tecnologias, sabe-se que várias bibliotecas brasileiras abandonaram seus sistemas de recuperação de informação tradicionalmente usados e passaram a indexar os documentos de suas coleções utilizando somente a linguagem natural. Mais tarde, descobriram que, talvez, essa não tenha sido a melhor solução, já que houve aumento significativo de discrepâncias entre perguntas dos usuários e respostas dos sistemas.

Apesar do uso intensivo de novas tecnologias, o controle sobre o vocabulário empregado para indexar e descrever os assuntos de documentos pertencentes a coleções de bibliotecas e outros serviços de informação, como o vocabulário de empresas, ou mesmo o de documentos de coleções privadas, continua a ser necessário e usado por meio das linguagens documentárias. Como fazer esse controle?

Agora que já estudamos o que é representação temática, podemos falar dos instrumentos usados neste processo, objetivo principal de nossa disciplina. Vamos começar identificando a natureza desse instrumento.

1.5.2 Já teve oportunidade de conhecer uma LD?

Com tantas disciplinas estudadas até agora, é bem provável que você já tenha conhecido uma LD sem saber do que se tratava. Se já utilizou uma *Classificação Decimal de Dewey (CDD)*, por exemplo, então a resposta é sim!

Classificar e dar nome às coisas não é tarefa fácil, embora seja uma das mais antigas. O homem sempre teve necessidade de classificar, numa tentativa de ordenar o mundo que o cerca.

Uma das noções que aprendemos primeiro, entre outras, é a de direita e esquerda. Nas escolas, sempre aprendemos a levantar a mão direita ou a esquerda, a entrar na fila da direita ou da esquerda. Quando você identifica a natureza de determinado objeto, você está fazendo nada mais do que uma classificação, ou seja, você cria uma classe à qual o objeto deve pertencer, sempre de acordo com suas características. Da mesma forma, quando você indexa um livro sob determinado termo, você o está classificando. Por exemplo, um livro trata de laranjas, tangerinas e abacaxis. Qual o termo mais adequado para representar esse livro numa base de dados? Se seu serviço de informação possui um vocabulário para padronizar a linguagem usada, ou seja, se ele possui uma LD, é quase certo que seu vocabulário contenha o termo "frutas cítricas" para nomear uma grande classe.

Porém, nem sempre há consenso entre os profissionais que classificam os objetos: uns classificam em uma classe, outros, em outra. Por que isso acontece? As razões podem ser várias, desde a falta de definições adequadas dos termos, até o próprio entendimento que o profissional tem do objeto a ser classificado.

Esse fato acontece em relação às listas de cabeçalhos de assunto, "que, de um modo geral, os autores não consideram como linguagem documentária" (CAMPOS, 2001, p.12), por julgar os esquemas de classificação e os tesauros como LD mais relevantes (CAMPOS, 2001).

Muito menos por sua natureza e função e mais por sua apresentação formal, as LD podem ser:

a) notacionais: possuem notações para expressar os assuntos, como os códigos numéricos (ex.: *CDD*); alfanuméricos (ex.: *Classificação Decimal Universal* (*CDU*), *Classificação dos Dois Pontos – Colon Classification, Classificação da Biblioteca do Congresso dos Estados Unidos – Library of Congress*);

b) verbais: baseiam-se, exclusivamente, em palavras/termos para expressar os assuntos.

Em nossa disciplina, serão estudadas apenas as linguagens verbais, já que as demais são objetos de outra disciplina ("Instrumentos de Representação Temática da Informação I"). Aqui, pontualmente, serão consideradas as seguintes LD:

a) lista de cabeçalhos de assunto;

b) tesauro;

c) outras iniciativas desenvolvidas como instrumentos para representação temática dos documentos em geral (vocabulário sistematizado, folksonomia, *Soggetario*).

Modernamente, surgiram as ontologias – instrumentos igualmente auxiliares no processamento de respostas precisas a perguntas pontuais dos usuários de um sistema de busca. Por suas características peculiares, elas serão abordadas na Unidade 4.

1.5.3 Linguagem documentária: que linguagem é essa?

Você está estranhando essa expressão "linguagem documentária"? Tem razão. Quando falamos de linguagem, lembramos imediatamente da língua que usamos para nos comunicar. Nossa comunicação utiliza sistemas simbólicos e um dos símbolos usados é a palavra, expressa por meio da linguagem. Mas quando a linguagem comum – a natural – é convertida para outra, de forma organizada, para representar assuntos de documentos, aí já estamos diante de outra linguagem – artificial –, chamada de linguagem documentária. Linguagem documentária é, então, uma expressão usada em sentido figurado. Ela é chamada de linguagem por ser constituída por um sistema de signos para comunicação.

O conceito de linguagem documentária surge no século XX, mais precisamente nos anos 1940, com a teorização da área de indexação/recuperação da informação. E qual é sua principal função? É intermediar o assunto dos documentos e a necessidade de informação do usuário. Nesse sentido, tal atividade se insere em um quadro maior, que denominamos "organização do conhecimento".

O avanço do conhecimento e, consequentemente, o aumento de publicações especializadas exigiram a criação de serviços especializados de informação técnico-científica, pois seus usuários passaram a requerer maior rapidez e exatidão na recuperação das informações ali armazenadas. Nesses serviços, o acesso ao documento por meio dos assuntos é uma forma de busca bastante procurada pelos pesquisadores. Aqui é que se insere a importância da indexação por assuntos para a representação dos documentos.

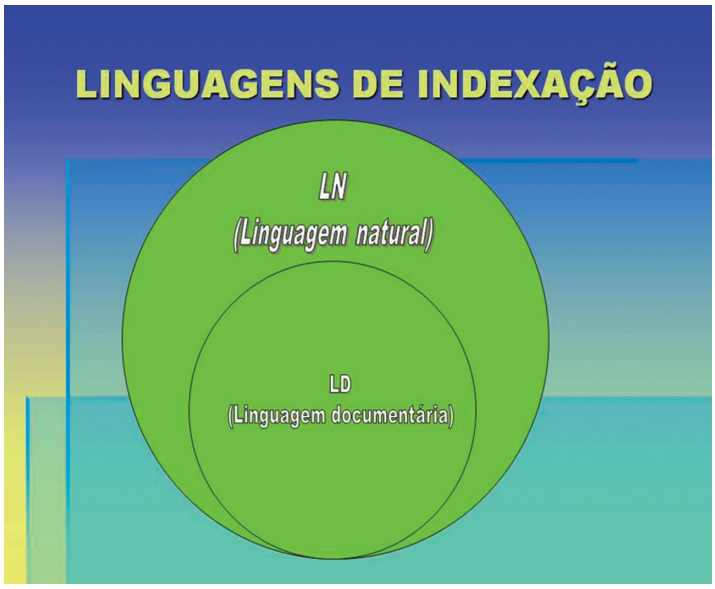
Já dissemos, quando falamos de indexação, que esse processo tem por finalidade verificar os assuntos contidos nos documentos e rotulá-los com palavras ou expressões (pontos de acesso) que representam os assuntos, substituindo-os. Pois bem, ao conjunto dessas palavras ou expressões chamamos de linguagem documentária.

Em contraposição à linguagem natural, pode-se dizer que linguagens documentárias são linguagens artificiais, pois são construídas para um fim específico, baseadas em regras formuladas antes mesmo de seu uso. Por vezes, essa linguagem é também conhecida por linguagem de indexação, embora consideremos tal designação inadequada. Por que inadequada? Porque a linguagem natural também é uma linguagem de indexação, com a diferença de que nenhum controle sobre os termos é feito; ainda hoje, ela é usada por vários serviços de informação na indexação de seus documentos. A LD é, então, apenas um dos tipos de linguagem de indexação.

Vamos ver se você compreendeu. Se eu fizer a seguinte pergunta: "podemos dizer que a linguagem documentária é um tipo de linguagem de indexação com termos extraídos da linguagem natural, porém submetidos a um controle, no mínimo, de sinônimos e homógrafos?", que resposta você dará? Se pensou em "sim", parabéns, acertou na mosca! Em muitos casos, as LD apresentam outras peculiaridades, como veremos adiante.

A Figura 5, a seguir, ajuda a mostrar o universo da linguagem documentária no âmbito da linguagem natural:

Figura 5 - Linguagens de indexação: LN e LD



Fonte: produção do próprio autor (2017).

Nada melhor para uma boa definição do que resumir as principais características do objeto a ser definido. Podemos caracterizar uma LD como:

a) linguagem artificial: é uma linguagem construída artificialmente, a partir das palavras/expressões retiradas da linguagem natural – a falada no dia a dia; é baseada em regras definidas antes mesmo de seu uso. Por exemplo, em um vocabulário sobre economia, o termo "comércio exterior" poderia vir acompanhado das seguintes orientações em uma "nota de aplicação":

> **COMÉRCIO EXTERIOR**
>
> Indexar sob este termo documentos que tratem exclusivamente do comércio de bens e serviços entre determinados países. Para o comércio entre todos os países do mundo, indexar sob o termo "comércio mundial."
>
> Diante dessa orientação, um documento sobre exportação de minério de ferro brasileiro para o Japão seria indexado com os termos: "minério de ferro", "Brasil", "comércio exterior" e "Japão". E um documento sobre o comércio, no mundo, de produtos manufaturados em 2014 deveria ser indexado com os termos: "comércio mundial", "produtos manufaturados" e "2014";

b) linguagem que tem como componentes: vocabulário predeterminado, regras para emprego dos termos e, às vezes, sintaxe – para determinar a ordem dos elementos do termo de indexação;

c) instrumento que exerce controle de aspectos verbais e conceituais da língua, como quando tenta reduzir a polissemia da linguagem natural, controlando sinônimos (ex.: economia informal = economia oculta = economia subterrânea = criptoeconomia), quase sinônimos (ex.: champanhe e espumante) e homógrafos (ex.: tênis, o esporte, e tênis, a peça de vestuário);

d) ferramenta de auxílio na indexação e recuperação de documentos;

e) instrumento que tem impacto direto na qualidade da informação recuperada.

Agora, podemos finalmente dizer o que entendemos por **linguagem documentária**:

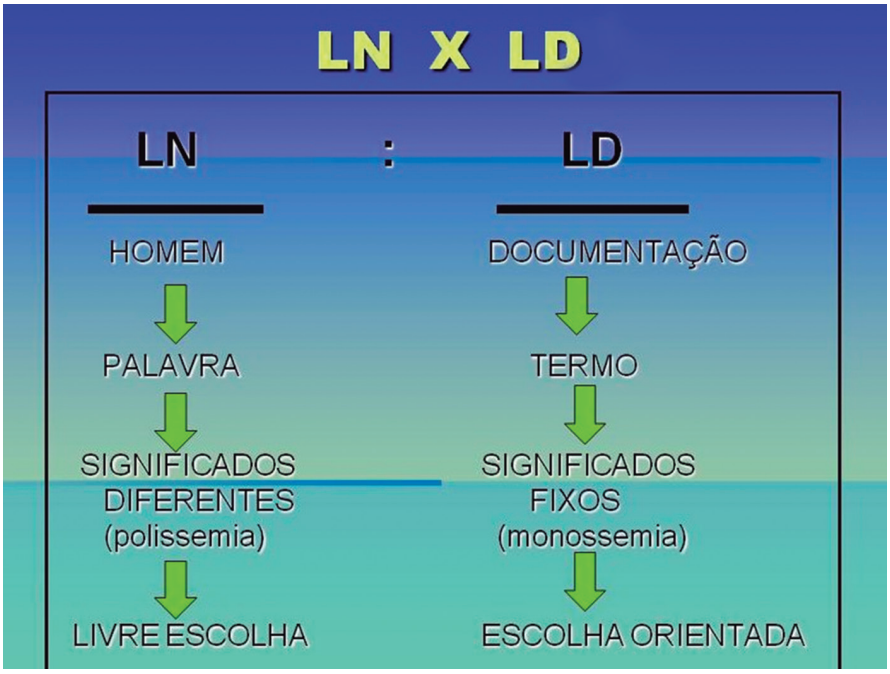
> **Linguagem documentária**
>
> É uma linguagem de indexação desenvolvida artificialmente, composta de um vocabulário que contém regras para controle verbal e conceitual dos termos – e, por vezes, sintaxe –, empregada nos SRI. Tem como funções orientar a indexação dos assuntos dos documentos tratados e auxiliar nas buscas feitas aos sistemas, objetivando alcançar maior coincidência entre perguntas e respostas aos usuários.

1.5.4 Linguagem documentária: ela é conhecida por tantos outros nomes!

Não se esqueça de que você poderá encontrar na literatura da área outros nomes dados às linguagens documentárias, como: linguagem de indexação (embora tenhamos esclarecido que a LD é apenas uma das linguagens de indexação, que também se indexa a partir da LN); vocabulário controlado (mas o vocabulário é apenas um de seus componentes); linguagem descritora (porque os termos em determinadas linguagens são também chamados de "descritores"); linguagens de recuperação da informação (embora as LD sejam usadas também na indexação, na entrada de dados, e não só na hora da recuperação). A literatura ainda se refere às LD como codificações documentárias, linguagens de informação e listas de assuntos autorizados.

A Figura 6 ilustra as diferenças básicas entre "linguagem natural" e "linguagem documentária":

Figura 6 - Diferenças básicas entre LN e LD



Fonte: produção do próprio autor (2017).

Com a quantidade de informação à disposição do público atualmente, principalmente depois do advento da internet, os provedores de informação não poupam esforços para melhorar a qualidade da informação recuperada. Como já dito anteriormente, a *web* semântica é o maior exemplo disso. O grande desafio é lidar com o significado das palavras. Então, qualquer iniciativa que tente reduzir a ambiguidade própria da linguagem usada pelos falantes é louvável. Resumindo, a linguagem documentária tem algumas vantagens também nesse sentido, cumprindo a função de reduzir ambiguidades semânticas. Como já dito antes, o controle de sinônimos e homógrafos é o mínimo que pode ser feito para esse fim.

Multimídia

Mergulhe e vá fundo!

Figura 7 - Estudando as LD

Fonte: produção do próprio autor (2017).

Se você se interessou por esse assunto, o livro *Linguagem documentária: teorias que fundamentam sua elaboração*, de *Maria Luiza de Almeida Campos*, editado pela *EdUFF* em 2001, dispõe de uma quantidade enorme de informações que podem, no mínimo, matar sua curiosidade. Disponível no *link* da referência a seguir:

- CAMPOS, M. L. de A. **Linguagem documentária***:* teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001. Disponível em: <http://www.uff.br/ppgci/editais/linguagem.pdf>. Acesso em: 13 nov. 2014.

1.5.5 **Atividade**

Leia o texto a seguir e identifique dois termos que possibilitem conexões com outros conteúdos. Em seguida, faça, nos espaços disponíveis 1 e 2, dois pequenos textos sobre os termos que você identificou.

Todo movimento existente nos Sistemas de Recuperação de Informação tem por princípio geral possibilitar a seu usuário o acesso à informação/documentos. Nestes Sistemas, vários são, atualmente, os instrumentos utilizados para representar o conhecimento de uma dada área do saber. Estes instrumentos são denominados, de uma forma geral, linguagens documentárias, como o Tesauro e os Esquemas de classificação, para citar apenas os mais relevantes.

Fonte: CAMPOS, M. L. de A. **Linguagem documentária:** teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001. p. 17.

1

2

Resposta comentada

Eu escolhi os seguintes termos: tesauro e esquema de classificação. Será que você escolheu um desses dois?

a) Sobre tesauro, o que podemos dizer sucintamente é que ele, tal como o conhecemos hoje, é uma linguagem documentária verbal que, como as demais linguagens documentárias, é usada tanto para a indexação como para a recuperação da informação. Uma característica que torna o tesauro diferente das demais LD é que ele, tradicionalmente, não se destina a cobrir todas as áreas do conhecimento. De acordo com seu conteúdo, ele pode ser multidisciplinar (voltado para um assunto e para áreas do conhecimento relativas a esse assunto) ou especializado (uma disciplina específica do conhecimento humano);

b) Quanto a esquemas de classificação, o que podemos dizer é que são a representação gráfica do conjunto de agrupamentos de assuntos coordenados e subordinados por determinadas características chamadas de classes, com suas divisões, seções etc.

1.6 VOCABULÁRIO DE UMA LD

Qualquer tipo de linguagem tem em comum dois componentes: o vocabulário e o conjunto de prescrições e regras que determinam o uso considerado correto de uma língua.

O mesmo ocorre com uma linguagem documentária: ela tem vocabulário, regras para sua aplicação e determinadas LD também têm sintaxe. Por que devemos nos concentrar no vocabulário controlado? Porque já vimos que, na indexação a partir da linguagem natural, o vocabulário é o usado em nosso dia a dia; sobre este, não há controle algum.

Quanto às regras para sua utilização, elas são peculiares a cada termo e poderão ser observadas nos exemplos desta disciplina.

Um vocabulário qualquer é um conjunto dos vocábulos de uma língua; é seu léxico. Porém, o vocabulário de uma linguagem documentária tem certas peculiaridades. Ele deve ser construído, estruturado, padronizado com o propósito imediato de controlar a língua que se usa correntemente, fazendo a ponte entre o sistema e o usuário. Há dois planos de controle: o plano da língua e o plano conceitual.

No plano da língua (verbal), esse controle se dá para regular o uso de:

a) equivalências (sinônimo e homógrafos. Ex.: Carretel USE **Bobina**; Aipim USE **Mandioca**; Cabo USE **Fio** (nesse caso, para diferenciar de "cabo" com o significado de posto militar);

b) forma (Século XX, **e não** séc. 20);

c) concordância de número (Cabeçalhos de assunto, **e não** cabeçalhos de assuntos).

O vocabulário das linguagens documentárias pode ser classificado de diversas formas, de acordo com determinados pontos de vista, como veremos a seguir:

a) vocabulários pré-coordenados e pós-coordenados:

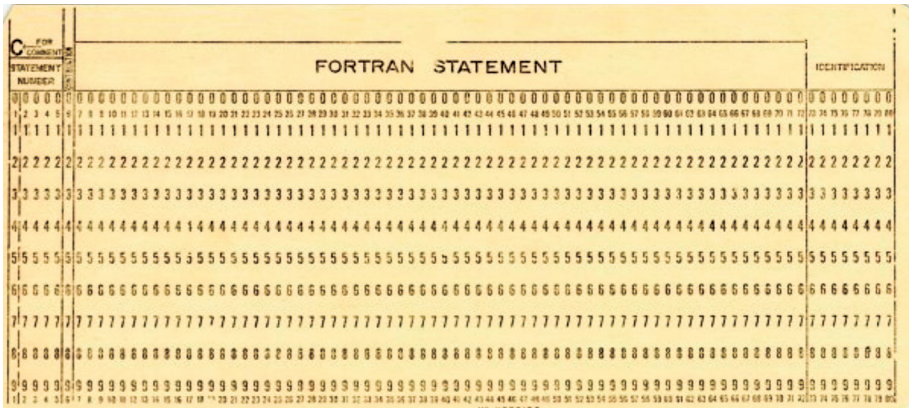
- pré-coordenado: quando a combinação entre os assuntos abordados nos documentos é feita a priori, na hora de entrada dos dados no sistema; assim, os termos de indexação já vêm prescritos, indicados para uso. Ex.: beneficiamento de café;
- pós-coordenado: quando a combinação dos termos é feita na hora da busca. Ex.: na busca de um documento sobre "produção da soja brasileira", os seguintes termos, constantes do vocabulário, deverão ser combinados: soja + produção + Brasil.

Na atualidade, muitos sistemas de recuperação são parcialmente pré e pós-coordenados. Conceitos que aparecem frequentemente juntos podem ser combinados em um termo pré-coordenado e esse termo pode, ainda, ser coordenado com outros, fora da pesquisa. Se uma busca é feita sobre "radioterapia de neoplasmas do fígado", por exemplo, os seguintes termos constantes de um vocabulário médico podem ser combinados: neoplasmas do fígado + radioterapia.

Explicativo

Para saber mais...

Figura 8 - Cartão perfurado de *Herman Hollerith*



Fonte: *Wikimedia Commons* (2005).[3]

Os sistemas pré-coordenados já existiam antes de 1940 e eram destinados a catálogos de fichas. Como exemplos, podem ser citados os sistemas desenvolvidos por *Battem* (cartões *peek-a-boo*: uso da coincidência ótica); *Mooers* (cartões perfurados nas margens),

[3] Disponível em: <https://commons.wikimedia.org/w/index.php?curid=59230>.

Taube (fichas Unitermo: uso de um assunto sendo representado por uma única palavra).

Atualmente, acredita-se que os sistemas pré-coordenados já sejam todos gerenciados automaticamente.

b) vocabulários enumerativos e sintéticos:

- enumerativos: só permitem empregar termos independentes e não autorizam a combinação de termos para expressar algo mais complexo. Relacionam ou enumeram os termos que o indexador pode utilizar na indexação dos assuntos dos documentos, sem possibilitar flexibilidade para a criação de novos termos por meio da combinação dos enumerados;
- sintéticos: além de relacionar os termos que o indexador pode empregar na indexação, oferecem regras pelas quais esses termos podem ser combinados de modos variados para formar novos termos, mais específicos (LANCASTER, 1972).

1.6.1 Funções do vocabulário controlado

Figura 9 - Funções do vocabulário controlado

Fonte: produção do próprio autor (2017).

A Figura 9 evidencia as duas funções principais de um vocabulário controlado:

a) na entrada de dados: auxilia o indexador a converter os assuntos tratados em um documento para o vocabulário do sistema;

b) na saída de dados: auxilia o usuário a converter sua necessidade de informação, expressa com suas palavras, nos termos usados no vocabulário.

A essa duas funções, somam-se outras especificidades igualmente importantes, a saber:

a) reduz ambiguidades semânticas: sinônimos e homógrafos podem ser controlados.

   Ex.:

   - INDEXAÇÃO (ECONOMIA)
   - INDEXAÇÃO (BIBLIOTECONOMIA)
   - PG **USE** PROGRESSÃO GEOMÉTRICA

b) possibilita a representação de um documento em uma base de dados;

c) permite a coincidência entre a linguagem do indexador e a do pesquisador;

d) propicia buscas genéricas. No exemplo a seguir, o usuário é lembrado de que, sob a expressão "operação matemática", ele pode encontrar todas elas, sendo que uma delas pode ser a mais adequada a sua pesquisa.

   Ex.:

OPERAÇÃO MATEMÁTICA
Termo específico
ADIÇÃO
DIVISÃO
MULTIPLICAÇÃO
SUBTRAÇÃO

e) evita a dispersão de assuntos com significados relacionados. Repare, por exemplo, nos seguintes assuntos constantes de um serviço de informação de uma loja de roupas masculinas e femininas: blusa, cardigan, jardineira, lingerie, minissaia, pantalona, saia-calça. Numa lista arranjada apenas em ordem alfabética, esses assuntos ficariam distantes uns dos outros, dispersos semanticamente, já que "blusa" ficaria na letra "b", "cardigan" na letra "c", "jardineira" na letra "j" e assim por diante. De fato, todos eles são tipos de artigos de vestuário e, se estivessem agrupados de acordo com seu significado, a busca de um usuário poderia ficar mais completa. Nesse caso, o usuário seria lembrado dos assuntos relacionados ao de sua busca inicial existentes no sistema. As listagens meramente alfabéticas não dispõem de artifícios para possibilitar a reunião de assuntos correlatos. Observe como o exemplo dado estaria disposto numa linguagem documentária:

PEÇA DE VESTUÁRIO
Termo específico
BLUSA
CARDIGAN
JARDINEIRA
LINGERIE
MINISSAIA
PANTALONA
SAIA-CALÇA

Um pesquisador que esteja interessado em buscar informações sobre "peça de vestuário" saberá, de antemão, que tipos de peças de vestuário existem no sistema de informação. Vamos imaginar que a ele interesse, sobretudo, informação sobre "lingerie" e "pantalona." O agrupamento dos termos pelo significado indicará que existe informação sobre esses assuntos, mais específicos do que "peça de vestuário", e que a busca pode ser feita diretamente por eles;

f) restringe o número de termos de entrada do vocabulário para uso. O controle de sinônimos cumpre essa função por meio das referências do tipo USE. Elas orientam o usuário a fazer a busca sob um único termo, o escolhido para representar o assunto no sistema. Imagine um vocabulário com vários sinônimos de "aipim" como termos de entrada: macaxeira, macaxera, mandioca, mandioca-doce, mandioca-mansa! Por onde começar a busca?;

g) expressa a linguagem em nível conceitual (o do conceito), e não verbal (o da palavra), fixando um só significado para cada termo;

h) minimiza o esforço intelectual na atribuição de termos de indexação; racionaliza o trabalho: o indexador não precisa perder tempo pensando em que termo usar, porque o termo a ser usado já foi decidido a priori.

O vocabulário controlado, por ser um instrumento-padrão, revela-se de importância fundamental em qualquer SRI que pretenda melhorar a qualidade do resultado das buscas nele feitas.

1.6.2 Objetivos do vocabulário controlado

Falamos das funções do vocabulário. Mas quais são seus objetivos em um SRI? Sintetizando, podemos citar três:

- promover a consistência da representação dos assuntos de um documento por indexadores (o mesmo assunto deve ser representado sempre pelo mesmo termo) e da busca feita pelo usuário. Isso é conseguido por meio do controle de termos/expressões sinônimas ou homógrafas. Tomando-se os exemplos dados anteriormente, a partir do momento em que o vocabulário orienta para usar BOBINA e não Carretel, o indexador não pode deixar de incluir todos os documentos a respeito desse assunto sob um único termo, qual seja: BOBINA. Por sua vez, o usuário que quiser buscar informação sobre Carretel será orientado a fazer a busca no termo BOBINA. Dá para imaginar indexadores de uma rede nacional de bibliotecas, um em cada estado do Brasil, indexando um documento sobre "pau-brasil", sem um vocabulário controlado? Olhe só quantos sinônimos o *Dicionário Houaiss da Língua Portuguesa* menciona para essa árvore: arabutã, arubatã, árvore-do-brasil, brasilaçu, brasilete, brasileto, brasil-rosado, ibirapiranga, ibirapitá, ibirapitanga, ibirapuitá, imbirapatanga, muirapiranga, orabutã, pau-de-pernambuco, pau-de-tinta, pau-pernambuco, pau-rosado, pau-vermelho, sapão. Neste caso, seria provável que várias entradas diferentes fossem dadas para o mesmo documento, se o sistema de informação fizesse a indexação dos documentos a partir da linguagem natural. Você concorda? E se todos os indexadores estivessem usando um vocabulário para controlar a entrada dos termos? O que

aconteceria? Todos esses sinônimos seriam remetidos para um só e a indexação seria consistente. O usuário seria o maior beneficiado, porque teria acesso à literatura toda sobre "pau-brasil" existente em todas as bibliotecas daquela rede;

- agregar termos com significados relacionados, propiciando buscas abrangentes. Ex.: "Seda", ver também: fibra natural, bicho-da-seda, tecido de seda, crepe da China, tecido de seda, carda, fiação. Repare que se esses termos fossem colocados numa simples ordem alfabética, embora com significados relacionados, ficariam distantes uns dos outros. Ora, a explicação é fácil, na verdade, óbvia: a ordenação alfabética não se preocupa com juntar significados. Sua proposta é listar assuntos seguindo a ordenação do alfabeto da língua pertinente a eles. E cada assunto começa com uma letra do alfabeto diferente. Voltando ao exemplo, de que forma o usuário se lembraria de todos os assuntos relacionados a "seda"? Percorrendo a lista alfabética toda? Ainda assim, seria necessário que ele tivesse relativo conhecimento do assunto pesquisado para identificá-los como sendo de interesse;
- proporcionar a comunicação entre indexador e usuário: o vocabulário controlado possibilita a convergência entre a linguagem usada pelo usuário e a usada pelo indexador para um único termo. Ele é o ponto comum que faz com que indexador e usuário "falem" a mesma linguagem.

1.6.3 Como utilizar o vocabulário?

Figura 10 - O vocabulário controlado e suas regras

Fonte: *Pixaba*y (2014).[4]

Às vezes, o vocabulário controlado também possui regras para uso dos termos na indexação. Ex.: no termo "saia de franjas", o vocabulário dá a seguinte orientação para indexação:

[4] Disponível em: <https://pixabay.com/pt/controle-trabalho-oficial-forma-427510>.

"Quando o artefato for caracterizado por matéria-prima, indexar pelo artefato **E** pela matéria-prima.

Ex.: SAIA DE FRANJAS DE PALHA; SAIA DE FRANJAS DE MIÇANGAS.

No plano conceitual, dependendo do tipo de linguagem documentária, além dos controles citados que se dão em nível verbal, o vocabulário pode incluir, ainda, uma estrutura de relações conceituais (termos genéricos, específicos, partitivos e associativos).

Ex: CAVALO.

Termo genérico (TG)

MAMÍFEROS

Essas linguagens e suas respectivas estruturas serão vistas detalhadamente mais adiante, em outra Unidade deste Curso.

1.6.4 **Atividade**

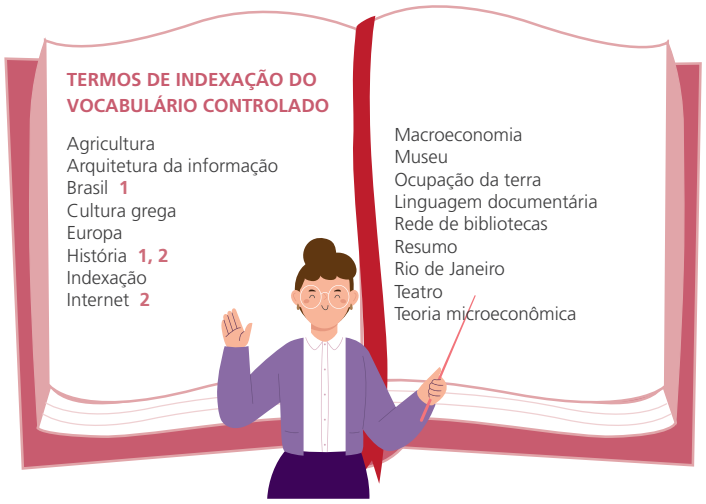
Praticando a indexação com LD

Imagine um serviço de indexação de um SRI, com um grupo de indexadores. O sistema dispõe de uma LD para indexar os assuntos dos documentos.

Um indexador, depois de analisar os documentos, cujos títulos estão listados na figura A, atribuiu a eles os temas/assuntos relacionados na figura B. Faça a correspondência entre documentos e termos de indexação. Os dois primeiros já estão feitos. Agora, continue!

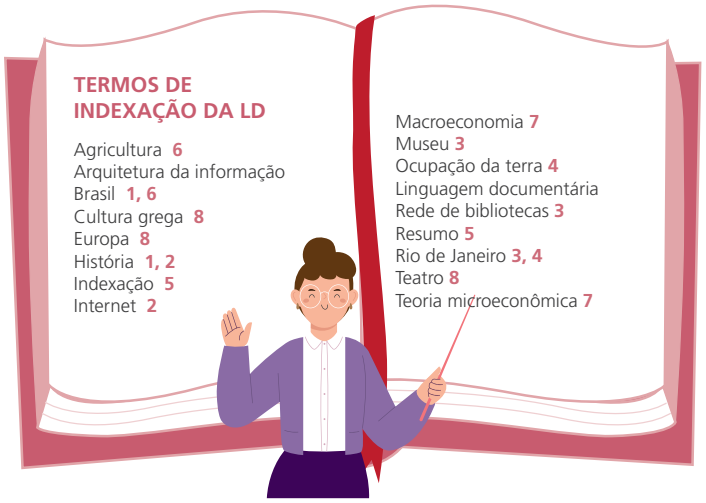
A

B



Resposta comentada

B



1. Logo de saída, o livro n. 1 só pôde ser contemplado com os assuntos "Brasil" e "história."No entanto, só com esses dois assuntos, você acha que ele ficou bem indexado? Onde foi parar a "arquitetura", que é o assunto principal? Será que o vocabulário não tinha esse termo? Ou será que o indexador se esqueceu mesmo de atribuí-lo? Se o vocabulário não tinha mesmo o termo "arquitetura", ele deveria ser incluído, já que é um assunto importantíssimo!
2. Se você indexou com os termos "história" e "internet", acertou! O vocabulário tinha os dois termos!
3. Alguns dos assuntos desse livro estão representados no vocabulário, não tem mistério. Se você colocou: "rede de bibliotecas", "museu" e "Rio de Janeiro", muito bom. Mas e o termo específico para designar o "controle de assuntos"? Certamente, esse tema tão importante teria ficado de fora e deveria ser criado.
4. Um bom indexador tem que ser curioso. Para indexar o tema "usucapião", é preciso, antes de mais nada, saber o que é isso, qual é o conceito do termo. É hora de ir ao dicionário para saber. Descobriu que tem a ver com posse de bens móveis e imóveis? Se você fez a correspondência com "ocupação da terra", fez bem, embora o vocabulário devesse ter o termo de entrada específico – "usucapião."Se você também indexou por "Rio de Janeiro", muito bem, acertou!

5. Esse é fácil! Você deve ter relacionado os termos "indexação" e "resumo."
6. Nesse caso, faltou ao vocabulário o termo específico "*agrobusiness*". O mais correto, com os termos disponíveis nesse vocabulário, seria indicar o termo mais genérico: "agricultura." E o outro: "Brasil."
7. Se você usou os termos, já disponíveis no vocabulário, "macroeconomia" e "teoria microeconômica", tudo bem; o usuário encontraria, certamente, o livro disponível no acervo (o de número 7 na figura). Porém, o ideal seria que o vocabulário tivesse, também, o termo "teoria macroeconômica." Assim, os dois tipos de economia estariam indexados pela forma mais correta, obedecendo à regra máxima da indexação, que é: indexe pelo termo mais específico.
8. Se você fez a correspondência com os termos: "cultura grega", "Europa" e "teatro", acertou! Mas a opção "teatro grego"(se estivesse incluída no vocabulário) + "Europa" seria a ideal, não acha?

1.6.5 Construindo um vocabulário controlado

Vários tipos de vocabulários têm sido usados em bibliotecas e outros serviços de informação: listas alfabéticas com controle de sinônimos e homógrafos; vocabulários contendo, também, definições para os termos e indicação das relações que eles guardam entre si; além de vocabulários mais sofisticados, que apresentam, entre outros artifícios, tradução dos termos para várias línguas ou indicação de outras listas que contêm termos com o mesmo significado.

Nos quadros a seguir, você pode visualizar exemplos dos dois primeiros tipos:

Quadro 1 - Relação de assuntos para cabeçalhos de fichas

RELAÇÃO DE ASSUNTOS PARA CABEÇALHOS DE FICHAS

Árvores
(Usar subdivisão geográfica)
(Os nomes de árvores não estão incluídos nessa lista, mas serão usados quando necessário. Ex.: Carvalho, Pinheiro etc.)
V. também Arboricultura, Floricultura, Madeiras, Plantas

Assistência
V. também Assistência social; Primeiros socorros

Catalogação
V. também Classificação

Costumes sociais – V. **Usos e costumes**

Mulher
(Subdividir por países, Ex: Mulher no Brasil)
V. também Casamento; Economia doméstica; Lar; Mães

Fonte: FERRAZ, Wanda. **Relação de assuntos para cabeçalhos de fichas.** 5. ed. rev. e aum. Rio de Janeiro: Freitas Bastos,1977.

Fonte: *Ferraz* (1997).

Quadro 2 - Lista alfabética com controle de sinônimos

LISTA ALFABÉTICA COM CONTROLE DE SINÔNIMOS

Abacaxi
DEFIN.: Fruta tropical do abacaxizeiro, da família das bromeliáceas.

Açaflor USE **AÇAFRÃO**

Banho-maria
DEFIN.: Processo utilizado para cozinhar ou aquecer o alimento sem contato direto com a fonte de calor.

Bolacha USE **BISCOITO**

Engenharia de alimentos
DEFIN.: Parte da engenharia que aplica seus princípios para a criação dos equipamentos necessários ao processamento de alimentos em grande escala.

Hortaliça
DEFIN.: Tipo de hortifrutícola comestível, de consistência e porte variado.

Raiz tuberosa USE **TUBÉRCULO**

Fonte: SCHMIDT, Wanda Lúcia (Org). **Microtesauro alimentos:** glossário. Brasília: SENAI/DN, 1999. v. 2.

Fonte: *Schmidt* (1999).

Basicamente, a construção de um vocabulário controlado, seja qual for o método de construção, deve ser desenvolvida dentro das seguintes etapas:

a) identificar as áreas de assunto a serem cobertas: quando mais de uma área, eleger uma delas, observando sua importância para o serviço de informação e sua clientela;

b) selecionar os termos apropriados para descrever essas áreas;

c) decidir sobre a forma dos termos, como devem aparecer no vocabulário (ex.: se no singular ou no plural, se em algarismos romanos ou arábicos, quando usar maiúsculas ou minúsculas);

d) organizar esses termos de uma forma útil (para o serviço de informação e seus usuários);

e) apresentar os termos em uma forma útil (para o serviço de informação e seus usuários).

Todas essas etapas serão vistas mais detalhadamente nas próximas Unidades, quando outras linguagens documentárias forem estudadas.

A elaboração de um vocabulário controlado merece todo o cuidado. A entrada dos termos, composta de expressões autorizadas e não autorizadas para uso (ex.: na entrada"Carretel USE BOBINA", "Carretel" é o termo não autorizado e "BOBINA" é o autorizado), precisa ser feita com base em princípios bem definidos, para garantir a coerência de seu uso e a precisão nos resultados das buscas.

Dependendo da forma como é elaborado, o vocabulário de entrada terá efeitos negativos ou positivos na recuperação da informação. Quer conhecer alguns exemplos? Eles são listados no Quadro 3:

Quadro 3 - Vocabulário de entrada e efeitos na recuperação da informação

PRÓS	CONTRAS
Uso de remissivas para termos equivalentes: o vocabulário de entrada indica que existem documentos sob determinado tópico, mesmo se remissivas forem usadas. Ex.: Mexerica USE **TANGERINA**. Ainda que o termo "mexerica" seja o não autorizado para uso, ele indica que na base de dados existem documentos sobre o assunto, indexados sob "tangerina", que foi o termo preferido para uso. Nesse caso, os dois termos foram considerados sinônimos.	**Não inclusão de termo no vocabulário**: o vocabulário de entrada que está sendo usado não inclui o termo específico necessário para indexar um documento. Ex.: uma pesquisa sobre "arquitetura de catedrais"; esse termo não existe no vocabulário. Solução: o indexador elabora uma estratégia de busca, na qual usa os seguintes termos: ARQUITETURA RELIGIOSA (que envolve não só catedrais, mas também outras espécies de templos), combinado com CATEDRAIS.
Redução do tempo gasto na indexação: espera-se que o tempo na indexação de cada documento seja reduzido, já que todos os termos autorizados para uso, assim como os não autorizados, já terão sido estabelecidos previamente.	**Inconsistência da indexação**. Ex.: em um vocabulário da área de Medicina, não há a expressão "ruptura prematura da membrana fetal." Consequentemente, os documentos sobre esse assunto foram indexados sob várias combinações, a partir dos termos disponíveis: MEMBRANA FETAL + RUPTURA; MEMBRANA FETAL + RUPTURA ESPONTÂNEA; MEMBRANA FETAL + COMPLICAÇÕES NO TRABALHO DE PARTO; MEMBRANA FETAL + COMPLICAÇÕES NA GRAVIDEZ; MEMBRANA FETAL + TRABALHO DE PARTO PREMATURO. Por certo, todas essas indexações levariam a documentos que satisfizessem à busca inicial. Porém, as informações estavam completamente dispersas na base de dados (20% dos documentos foram indexados sob cada um dos termos e nenhum dos pontos de acesso individualmente levaria à totalidade de informações disponíveis na base de dados).

Fonte: produção do próprio autor (2017).

Um vocabulário controlado é essencialmente um instrumento prático e seu principal requisito é que seja capaz de descrever assuntos encontrados nos documentos e nas perguntas feitas ao sistema pelos usuários. Ele deverá ser construído observando-se criteriosamente as características dos documentos da coleção a ser indexada e a linguagem usada pelos usuários. Seus termos devem refletir os assuntos contidos em pelo menos um documento existente na coleção. Em outras palavras, um bom voca-

bulário não deve incluir termos que, pressupostamente, poderão, um dia, ser usados. Finalmente, como vocabulários refletem a língua usada e ela é dinâmica, eles precisam ser revistos e atualizados periodicamente.

1.6.6 Atividade

Imagine que você acabou de assumir a coordenação de um SRI de uma das unidades da rede de bibliotecas de uma instituição nacional e verificou que a indexação vem sendo feita por meio da linguagem natural. Você quer convencer seu gerente sobre a necessidade de adoção de um vocabulário controlado para a representação e recuperação do conteúdo dos documentos.

Dentre os argumentos relacionados abaixo, assinale aqueles que justificam sua posição a favor da vantagem da adoção de um vocabulário controlado.

1. ( ) Reduz ambiguidades semânticas: sinônimos e homógrafos podem ser controlados.
2. ( ) Permite a coincidência entre a linguagem do indexador e a do pesquisador.
3. ( ) Propicia buscas genéricas.
4. ( ) Permite o acesso ao documento por qualquer termo que o pesquisador imaginar.
5. ( ) Evita a dispersão de assuntos com significados relacionados.
6. ( ) Padroniza a linguagem quando a indexação é realizada por uma rede de bibliotecas situadas fisicamente em lugares diferentes e com vários indexadores.
7. ( ) Dificulta a troca de informações entre as bibliotecas de uma rede.
8. ( ) Expressa linguagem em nível verbal.
9. ( ) Aumenta o tempo usado na indexação e na busca.

Resposta comentada

Se você assinalou o primeiro argumento, você acertou! Sinônimos e homógrafos podem ser facilmente controlados por meio das relações de equivalência: "Use/Usado por" ou "Ver."

Se você marcou o segundo argumento, acertou outra vez. Uma vez selecionados os termos para uso no vocabulário, somente esses termos poderão ser usados tanto pelo indexador, na indexação, quanto pelo usuário, na busca. Atencão deve ser dada, quando a inclusão de um novo termmo no vocabulário faz-se necessária. Porém, esta decisão só deverá ser tomada, quando a falta do termo for plenamente justificada.

O terceiro argumento também deve ser assinalado, pois é uma vantagem do vocabulário controlado. Em geral, tais vocabulários agregam os termos com significados relacionados, apontando a relação entre eles. Assim, propiciam a ampliação da busca a partir de um termo mais abrangente.

O quarto argumento é falso. A partir do momento em que o termo é escolhido para entrada no vocabulário, a busca deverá ser feita somente por ele. Caso contrário, o documento não será recuperado, a não ser que o vocabulário tenha uma rede de remissivas bem construída e que tenha incluído o termo imaginado pelo usuário.

O quinto argumento é verdadeiro. O vocabulário controlado dispõe de uma estrutura por meio da qual termos com significados próximos são reunidos por meio de relações de equivalência (para solucionar o caso de sinônimos e homógrafos), de gênero-espécie (para solucionar problemas de hierarquia conceitual), bem como relações de associação semântica.

O sexto argumento também é verdadeiro, pois imagina-se que, em uma rede, haja vários indexadores, cada um indexando de forma diferente. Daí a necessidade de padronização por meio do vocabulário controlado.

O sétimo argumento não está correto. Muito pelo contrário, o vocabulário, estando padronizado, facilita a troca de informações, já que todos utilizarão a mesma linguagem.

O oitavo argumento também está errado. Em vocabulários controlados, a unidade de representação deve ser sempre o conceito e não a palavra.

Por fim, o nono argumento está incorreto, pois o uso do vocabulário minimiza o tempo tanto do indexador quanto do usuário, já que o termo escolhido para representação do assunto foi determinado previamente.

RESUMO

Linguagem documentária (LD) foi um conceito introduzido na década de 1940 para designar uma linguagem artificial e dinâmica, em contraposição à linguagem natural – a falada no dia a dia. Destinada a ser instrumento auxiliar do processo de indexação/recuperação da informação, é artificial porque trata de uma linguagem construída, com regras pré-estabelecidas para sua aplicação; e é dinâmica porque os termos nela incluídos necessitam de atualização constante, para acompanhar as constantes mudanças da linguagem, que refletem, por sua vez, os desenvolvimentos científicos e tecnológicos.

O processo de indexação/recuperação da informação faz-se necessário para representar, em uma base de dados, os assuntos contidos nos documentos, a fim de que estes possam ser depois recuperados por usuários com necessidades de informação.

A LD garante maior precisão na busca da informação e seus principais objetivos são:

a) promover a consistência da representação de assuntos de documentos;

b) agregar termos com significados relacionados;

c) proporcionar a interoperabilidade de sistemas que utilizem um mesmo vocabulário para tratamento de suas coleções documentais.

As LD podem ser notacionais (ex.: *Colon Classification*, *CDD*, *CDU*) e verbais (ex.: lista de cabeçalhos de assunto, tesauro). As LD são compostas de vocabulário (léxico), de regras para aplicação deste e, às vezes, de sintaxe, para estabelecer a ordem de citação entre os elementos que formam os termos atribuídos ao documento.

O vocabulário de uma LD é de extrema importância, pois ele terá impacto direto sobre a qualidade da informação recuperada. Dentre os principais objetivos de um vocabulário, encontram-se:

a) promover a consistência da representação dos assuntos usados pelo indexador e da busca feita pelo usuário;

b) agrupar termos com significados relacionados;

c) proporcionar a interoperabilidade de serviços de informação automatizados que usem a mesma LD, garantindo, assim, a utilização de um vocabulário comum para os diversos SRI;

d) auxiliar o indexador a converter os assuntos tratados em um documento para o vocabulário do sistema;

e) auxiliar o usuário a converter sua necessidade de informação, expressa com suas palavras, nos termos usados no vocabulário;

f) reduzir ambiguidades semânticas por meio do controle de sinônimos e homógrafos;

g) permitir a coincidência entre a linguagem do indexador e a do pesquisador;

h) propiciar buscas genéricas;

i) evitar a dispersão de assuntos com significados relacionados;

j) restringir o número de termos de entrada do vocabulário usado para indexar;

k) minimizar o esforço intelectual na atribuição de termos de indexação.

Mesmo com todos os recursos da internet, onde as buscas são pautadas na linguagem livre usada pelos usuários, existe consenso sobre a necessidade de se controlar o vocabulário usado na indexação da informação disponibilizada na rede, já que a polissemia da linguagem natural é grande e interfere diretamente na qualidade da informação recuperada. Prova disso são os investimentos que têm sido feitos para o desenvolvimento da *web* semântica, em que os significados dos conceitos são a garantia para a recuperação mais precisa.

Sugestão de Leitura

COLLISON, Robert L. **Índices e indexação**. São Paulo: Polígono, [1972].

CONCEIÇÃO, V. P. da; PECEGUEIRO, Cláudia M. P. de A. Pressupostos teórico-metodológicos da linguagem documentária. **Cad. Pesq.**, São Luís, v. 13, n. 2, p. 51-64, jul./dez. 2002.

LANCASTER, F. W. **Vocabulary control for information retrieval.** 2nd ed. Virginia: IRP, 1986.

REFERÊNCIAS

CAMPOS, Maria Luiza de Almeida. **Linguagem documentária:** teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001.

FERRAZ, Wanda. **Relação de assuntos para cabeçalhos de fichas**. 5. ed. rev. e aum. Rio de Janeiro: Freitas Bastos, 1977.

HOUAISS, Antônio. **Dicionário Houaiss eletrônico:** versão 3.0. Rio de Janeiro: Editora Objetiva, 2009.

LANCASTER, F. W. **Indexação e resumos:** teoria e prática. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. **Vocabulary control for information retrieval**. Tradução de M. A. R. Piedade. Washington: IRP, 1972.

SCHMIDT, Wanda Lúcia (Org). **Microtesauro alimentos:** glossário. Brasília: SENAI/DN, 1999. v. 2.

UNIDADE 2

REFLEXÕES SOBRE O CABEÇALHO DE ASSUNTO

2.1 OBJETIVO GERAL

Apresentar o cabeçalho de assunto como instrumento na indexação e recuperação da informação, por meio da identificação de seu conceito, contextualização histórica, elaboração e avaliação de uso.

2.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

a) identificar a natureza, as características e funções do cabeçalho de assunto;

b) sistematizar os princípios utilizados na elaboração de cabeçalhos de assunto;

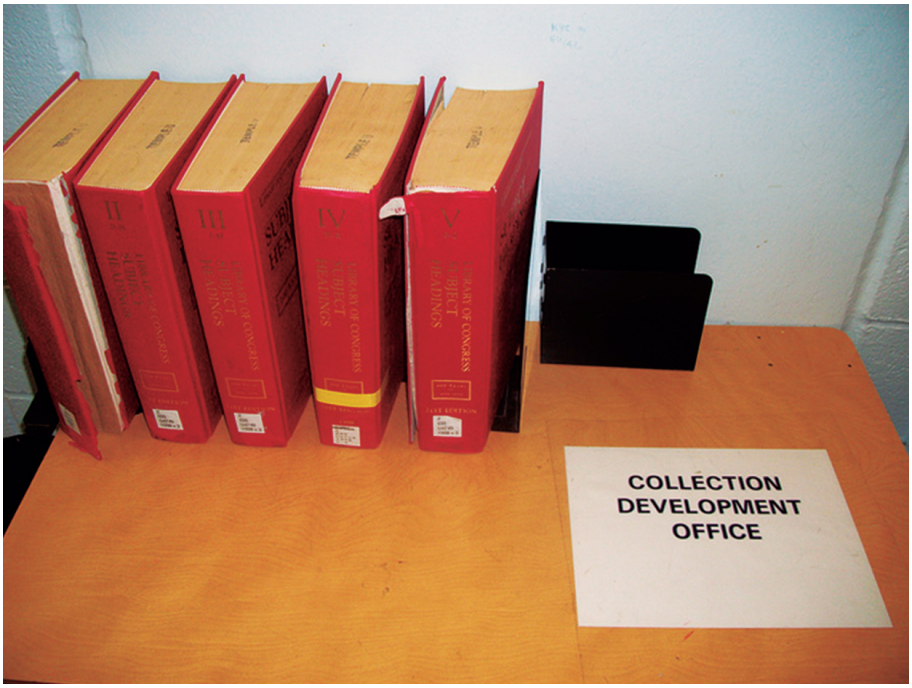
c) identificar limitações do uso de cabeçalhos de assunto pelos sistemas de recuperação da informação.

2.3 PRÉ-REQUISITO

Como a *Lista de Cabeçalhos de Assunto da Biblioteca do Congresso dos Estados Unidos*, conhecida como *Library of Congress Subject Headings (LCSH*), foi e ainda é bastante usada pelas bibliotecas brasileiras, sugerimos que você tenha em mãos um exemplar para acompanhar mais facilmente as considerações que faremos nesta Unidade.

2.4 LISTAS DE CABEÇALHOS DE ASSUNTO

Figura 11 - Listas de cabeçalhos de assunto da *Biblioteca do Congresso dos Estados Unidos*



Fonte: *Flickr* (20--?).[5]

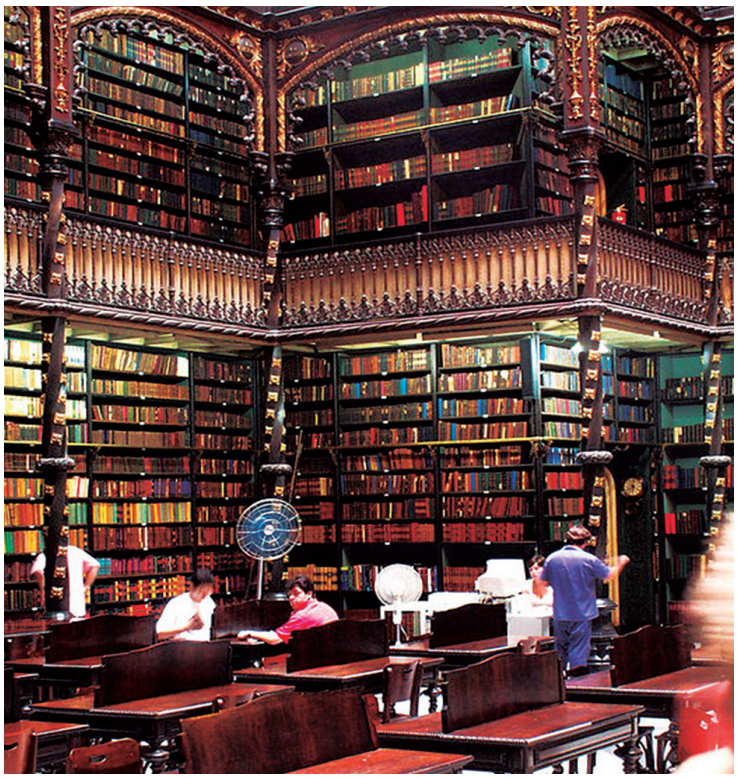
As listas de cabeçalhos de assunto têm gerado opiniões divergentes na área de informação. Dissemos anteriormente que elas não são consideradas, por muitos autores, como uma LD. Mas dissemos, também, que todas as LD destinam-se à facilitação dos processos de indexação e recuperação dos conteúdos temáticos dos documentos. Se considerarmos essa função, as listas de cabeçalhos de assunto não fogem à regra – elas são uma LD.

Quando se fala, porém, de maior ou menor eficiência das linguagens em cumprir tal função, isso fica por conta das características de cada uma delas. Vamos, então, saber da natureza das listas de cabeçalhos de assunto, da história que narra seu aparecimento e evolução, de suas características, por que é uma ferramenta tão questionada, quais são as razões que levam às restrições em seu uso e como essa ferramenta se insere no contexto informacional atual.

[5] Disponível em: <https://www.flickr.com/photos/tulpics/4882642683/in/dateposted/>.

2.5 TUDO TEM UMA HISTÓRIA...

Figura 12 - *Real Gabinete Português de Leitura* (RJ)



Fonte: *Wikimedia Commons* (2004).[6]

Figura 13 - Parte do egípcio *Livro dos Mortos*, escrito em papiro

Fonte: *Wikimedia Commons* (2005).[7]

[6] Disponível em: <https://commons.wikimedia.org/wiki/File:RealGabinetePortuguesLeitura1.jpg>.

[7] CC BY-SA 3.0. Disponível em: <https://commons.wikimedia.org/w/index.php?curid=127188>.

O domínio da escrita foi o caminho para a produção intelectual na Antiguidade, fato que logo despertou a preocupação com a guarda do material produzido e o aparecimento das bibliotecas. Antes mesmo da invenção do papel, fabricado pelos árabes, já havia registros feitos em tabletes de argila, nas chamadas "bibliotecas minerais." A estas, seguiram as "vegetais" e as "animais", constituídas de rolos de papiros e pergaminhos, respectivamente. Só depois surgiram as bibliotecas que guardavam papel e o livro propriamente dito. Mas é nas bibliotecas que a atividade de organização da informação surgiu, com a elaboração dos catálogos e das bibliografias, em ambos os casos, com a intenção de ordenação do material intelectual produzido.

Até o século XVIII, essa atividade de ordenação do conhecimento não tinha características tais como as conhecemos hoje. As primeiras bibliografias eram organizadas por cientistas e pesquisadores, de acordo com as classes de assuntos ou nomes de autores. A comercialização dos livros, anterior ao século XVIII, intensificou-se e trouxe a reboque as primeiras listas de assunto, tendo por base as palavras significativas dos títulos e dando origem a repertórios bibliográficos ordenados alfabeticamente por assunto. Foi então que surgiu a bibliografia nacional da Alemanha, dessa vez por iniciativa dos editores. A despeito dos avanços realizados para organizar o conhecimento, problemas persistiam.

Multimídia

No início era o verbo, depois...

Figura 14 - Obra do artista flamengo *Jan van Eyck* (século XV)

Fonte: *Pixabay* (2015).[8]

Quem hoje se deslumbra com a praticidade oferecida pelos *e-books* nem chega a imaginar o longo caminho percorrido pelos livros na história. Companheiros da escrita, eles tiveram grande importância para a realização de registros históricos, a compilação de leis e a divulgação de ideias. Atualmente, a produção de livros

[8] Disponível em: <https://pixabay.com/pt/jan-van-eyck-pintura-1027885>.

chegou a tal ponto que, por exemplo, o século XX foi responsável por uma literatura histórica superior à de todos os outros séculos somados juntos!

Essa história é fascinante! Você pode saber mais sobre ela. Visite o *site* a seguir e se apaixone pelo assunto: <http://www.historiadomundo.com.br/curiosidades/origem-dos-livros.htm>.

No século XIX, a atividade de ordenação do conhecimento se profissionalizou. As bibliotecas passaram a organizar as bibliografias e também seus catálogos. Isso, porém, não resolveu problemas relacionados à questão do estabelecimento da ordem para a representação e disseminação do que era produzido, como sintetizam *Cesarino* e *Pinto* (1978):

a) os títulos dos documentos não representavam adequadamente seus assuntos;

b) um mesmo documento passou a tratar de vários assuntos, tendo em vista a interdisciplinaridade das ciências, o que dificultou sua representação;

c) as subdivisões de assunto continuavam problemáticas pela ausência de regras para sua aplicação;

d) documentos que relacionavam os assuntos a lugares e épocas diferentes tinham representação inconsistente.

A necessidade de criar um novo instrumento para dar conta desses problemas era urgente. Na Europa, seguindo a tradição das bibliografias nacionais adotadas na Alemanha e nos catálogos das bibliotecas inglesas, a tendência era o uso dos catálogos sistemático e alfabético-por-classes. Neste último, se alguém estivesse interessado em documentos sobre "tuberculose", por exemplo, deveria procurar no termo de entrada correspondente ao assunto, mas numa ordenação superior, assim: como tuberculose é uma espécie de doença contagiosa, a busca deveria ser feita pela classe onde a tuberculose se insere e não pelo termo indicativo da espécie. Então, a entrada no catálogo seria: doença contagiosa – tuberculose. O exemplo explica porque as bibliotecas na Europa se destinavam ao público erudito, que, pressupunha-se, dominava as particularidades de cada ciência, ou seja, nesse caso, o usuário deveria saber de antemão que tuberculose é uma doença contagiosa, para poder fazer sua busca pela classe.

Porém, também no século XIX, os Estados Unidos da América (EUA) consideraram as bibliotecas instituições fundamentais à cultura do público em geral, devendo, pois, contribuir para o desenvolvimento educacional de cada indivíduo, e não mais somente para os eruditos.

Isso representou a democratização da informação. As bibliotecas passaram a oferecer novos produtos e serviços para facilitar o usuário em suas buscas. Os bibliotecários passaram a desempenhar novas funções e nova "vigilância" sobre o catálogo, peça fundamental para consulta nas bibliotecas, por permitir acesso à totalidade de suas coleções a partir de diversos pontos, ou seja, por tantos assuntos quantos o documento tratasse.

Foi nesse contexto ideológico que se deu a grande inovação da época: o catálogo-dicionário, surgido em 1876, em contraposição aos catálogos alfabético-por-classes e sistemático, usados nas bibliotecas inglesas e des-

tinados ao público erudito, como já dissemos aqui. O catálogo-dicionário será tratado adiante.

Mais tarde, *Ranganathan*, um dos expoentes da Biblioteconomia, ratificaria esse pensamento democrático em relação às bibliotecas, afirmando serem as mesmas para todos, quando enunciou, em uma de suas Leis da Biblioteconomia: "a cada leitor, seu livro".

Multimídia

Conhecimento nunca é demais!

Quadro 4 - As cinco leis da Biblioteconomia (*Ranganathan*)

As cinco leis da Biblioteconomia
(*Ranganathan*)

1. Livros são escritos para serem lidos.
2. A cada leitor, o seu livro.
3. A cada livro, o seu leitor.
4. Poupe o tempo do leitor.
5. A biblioteca é um organismo em crescimento.

Fonte: *Ranganathan* (1967).

Shiyali Ramamrita Ranganathan (1892-1972) é uma personalidade interessante, que revolucionou os caminhos da Biblioteconomia. Nascido na Índia, foi matemático e bibliotecário, mas, acima de tudo, um grande observador, com natureza altamente criativa.

Era também um profissional extremamente politizado e, como professor, preocupava-se com o ensino e a pesquisa em seu país. Graduou-se e pós-graduou-se em Matemática na *Universidade de Madras*, onde um dos requisitos para preencher o cargo de bibliotecário era estudar Biblioteconomia na Grã-Bretanha. Diante dessa exigência, ele ingressou na *Escola de Biblioteconomia na College University*, em Londres.

As cinco leis da Biblioteconomia, instituídas por *Ranganathan* em 1931, vigoram até os dias atuais.

Vale a pena saber mais sobre *Ranganathan*. Acesse o *site*: <http://www.conexaorio.com/biti> e consulte os artigos "A gênese do pensar de Ranganathan" e "Revisitando Ranganathan: a classificação na rede."

Mas você deve estar se perguntando: qual a diferença entre o catálogo-dicionário, o alfabético-por-classes e o sistemático? Veja a diferença, compare a definição dos três instrumentos e avalie qual é mais fácil para qualquer pessoa utilizar:

Catálogo sistemático

Catálogo cuja entrada é feita por uma notação previamente fixada a partir de um sistema de classificação, que representa o assunto do documento.

Ex.: 869.93 (na CDD, esse número representa o assunto "romance brasileiro").

Catálogo alfabético-por-classe

Catálogo cuja entrada é feita por um assunto representativo de uma classe, com subdivisões específicas de classes subordinadas.

Ex.: RELIGIÃO – BUDISMO (no catálogo-dicionário, a entrada seria direta por BUDISMO).

Catálogo-dicionário

Catálogo cuja entrada é feita diretamente pelo assunto tratado no documento, sem ser subordinado a outro termo representativo de uma classe mais ampla.

Ex.: ARGILA, e não PRODUTO DE ORIGEM MINERAL – ARGILA.

2.5.1 Cabeçalho de assunto: o que é?

É comum haver confusão entre os termos "cabeçalho de assunto" e "lista de cabeçalhos de assunto"; muitas vezes, um termo é tomado pelo outro. Por isso, nunca é demais esclarecer o significado de cada um.

Chamamos de "lista" o conjunto de cabeçalhos de assunto ordenados alfabeticamente, com regras para sua aplicação, constituindo-se em um instrumento de representação temática da informação.

E o cabeçalho de assunto, o que é? Veja, a seguir, seu conceito.

Atenção

Cabeçalho de assunto é uma palavra ou grupo de palavras que representam os conteúdos dos temas abordados num documento. Um conjunto de cabeçalhos de assunto forma uma lista de cabeçalhos de assunto.

2.5.2 Como surgiu o cabeçalho de assunto?

O sistema de cabeçalhos de assunto foi amplamente utilizado nas bibliotecas dos EUA a partir da segunda metade do século XIX e em outros países do mundo ocidental e oriental, desde então. A despeito de suas inconsistências, a *Lista de Cabeçalhos de Assunto da Biblioteca do*

Congresso dos Estados Unidos (*LCSH*)[9], ponto de partida para todos as demais existentes, continua a ser usada por muitas bibliotecas do mundo, inclusive no Brasil, haja vista a inexistência de outro sistema que dê conta do tratamento dos assuntos dos documentos de forma mais abrangente.

Todo sistema de indexação para recuperação por assuntos, seja ele pré ou pós-coordenado – conceitos já vistos mas que serão aprofundados mais adiante – é determinado pela tecnologia existente em sua época. No caso do cabeçalho de assunto, a tecnologia consistia nas fichas catalográficas e nos catálogos impressos (fichários). Os catálogos eram o repositório das fichas catalográficas (representação dos documentos de natureza monográfica existentes nas coleções das bibliotecas e já tratados), e a unidade de representação dos temas dos documentos era a palavra.

No cabeçalho de assunto, as palavras são coordenadas antes mesmo de serem utilizadas. Nele, a primeira palavra do cabeçalho colocado nas fichas catalográficas (principal e secundárias) revelou-se da máxima importância, já que a era o único ponto de acesso para a recuperação do documento. Por exemplo, uma monografia que tratasse de "acidentes aéreos", cujo cabeçalho de assunto atribuído a ela tivesse sido AVIAÇÃO – ACIDENTES, só seria recuperada pela palavra AVIAÇÃO, único ponto de acesso para se chegar até ela. O subcabeçalho ACIDENTES, pré-coordenado a AVIAÇÃO, ajudava a contextualizar esse assunto, mas não se constituía em uma entrada para busca.

2.5.3 A sistematização do cabeçalho de assunto

Dentro do espírito democrático da época, o norte-americano ***Charles Ammi Cutter*** tomou a iniciativa de sistematizar as entradas dos cabeçalhos de assunto usados na *Library of Congress (LC)*. *Cutter* não era simpático ao catálogo alfabético-por-classe, usado na Europa. Ele entendia que o usuário formulava suas buscas seguindo a ordem natural da linguagem, e não a partir de classes gerais com subdivisões. Ou seja, o usuário que quisesse obter informação sobre "ouro" não faria sua busca por MINERAIS PRECIOSOS – OURO, ele a faria diretamente pelo assunto OURO. Além disso, a subdivisão do assunto era a principal característica do catálogo alfabético-por-classes.[10]

Estes dois motivos somados – entradas no catálogo por classes e o catálogo caracterizado por subdivisões – levaram *Cutter* a abolir o princípio da subdivisão em seu trabalho sobre cabeçalhos de assunto, a não ser para casos em que a entrada fosse nome de lugar (país, cidade etc.), ex.: BRASIL–FOLCLORE. A grande característica e vantagem do catálogo-dicionário seria, então, a entrada direta pelo assunto tratado no documento, sem intermediação de outro assunto qualquer. Isso possibilitou o acesso ao catálogo por qualquer assunto específico desejável, e não de forma indireta, via classe.

O raciocínio de *Cutter* era lógico, afinal, quando nos expressamos por meio da linguagem, falamos em ordem direta, e não invertida. Apesar disso, as inversões e subdivisões existem por aí nas listas de cabeçalhos de assunto. Ex: MODA, DESVALORIZAÇÃO DA; ESTRADAS DE FERRO–ACIDENTES. Convenhamos, essa não é a forma comumente usada para nos comunicarmos! Se você ainda não teve oportunidade de consultar a *LCSH* ou outras listas de cabeçalhos de assunto, experimente consultar uma e constate o que foi dito anteriormente.

[9] Nesta disciplina, a Lista de *Cabeçalhos de Assunto da Biblioteca do Congresso dos Estados Unidos* será referida como *LCSH* ou, simplesmente, *Lista*.

[10] Disponível em: <https://commons.wikimedia.org/wiki/File:CharlesAmmiCutter_BostonAthenaeum.png>.

Charles Cutter

Figura 15 - *Charles Cutter*

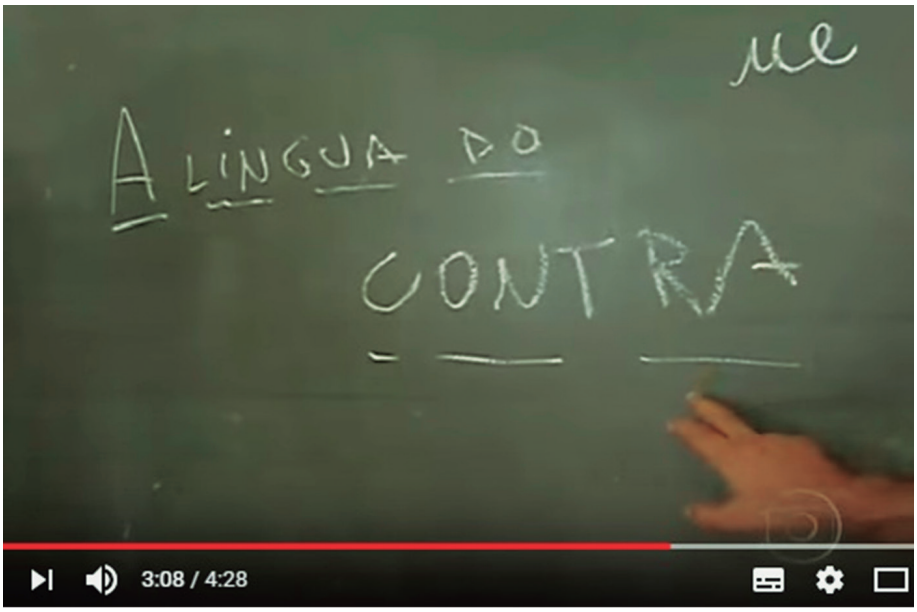
Fonte: *Wikimedia Commons* (1907).[10]

Charles Ammi Cutter nasceu em Boston, em 1837. Entrou na *Harvard Divinity School* em 1856, onde foi nomeado assistente bibliotecário, ainda como estudante. Graduou-se em 1859 e em 1860 já era bibliotecário catalogador. Em 1868, a *Biblioteca Athena*, em Boston, elegeu-o chefe bibliotecário, onde permaneceu por 25 anos. Em 1876, *Cutter* foi convidado pelo *Ministério da Educação* norte-americano para investigar a situação das bibliotecas estadunidenses. Concluiu esse trabalho com um extenso relatório, incluindo inovações que têm pleno vigor no século atual. Ainda em 1876, *Cutter* criou o catálogo-dicionário, baseado nos princípios da especificidade e da entrada direta, e apresentou o *Rules for a Dictionary Catalog* (*Regras para um Catálogo-Dicionário*). Foi editor da revista *Library Journal* de 1891 a 1893. Faleceu em 1903.

Multimídia

Falar em ordem invertida? Acredite se quiser!

Figura 16 - Cena da reportagem sobre Sabino - SP



Fonte: *Youtube* (2010).

Por incrível que pareça, existe uma cidade no Brasil em que as pessoas falam normalmente de trás para frente e conseguem se entender. Você seria capaz de se comunicar dessa maneira? Veja a reportagem veiculada no *link*: <www.youtube.com/watch?v=MUZiBe8E8z0>.

Para pensar...

Apesar de, em Sabino, as pessoas conseguirem se comunicar em uma linguagem falada de trás para frente, essa não é a forma usual de comunicação. Sendo assim, por que um vocabulário usado para representar e recuperar a informação deve se valer de inversões e subdivisões?

Em 1876, *Cutter* fez as primeiras tentativas de sistematização do catálogo-dicionário, novo instrumento de registro e acesso à informação, resultando no *Rules for a Dictionary Catalog (Regras para um Catálogo-Dicionário).* Ele estabeleceu diretrizes para a representação dos assuntos (o cabeçalho), mas não apresentou regras para construir o vocabulário: desde o início, a construção de novos cabeçalhos é prerrogativa da *LC.* Ao mesmo tempo, o cabeçalho buscava atuar – e ainda hoje é assim – como forma complementar (ou índice) da *Library of Congress Classification*, uma

vez que, em sua *Lista*, cada cabeçalho remete ao código de classificação correspondente.

Uma das primeiras medidas na construção dos cabeçalhos foi abandonar o modelo alfabético-por-classes: para o público geral, a entrada deveria ser direta, sem subdivisão.

Outras tentativas de sistematização (*Prévost*, em 1946; *Coates*, em 1960; *Schwartz*, em 1886) seguiram-se às de *Cutter* para resolver o problema da entrada direta dos cabeçalhos (GOMES; MARINHO, 1984). No entanto, nenhuma delas foi bem-sucedida, tendo sido rejeitadas ou ignoradas. Isso fez com que o número de subdivisões e inversões aumentasse, sacrificando os usuários e obrigando-os a percorrer o catálogo "catando" todas as entradas que julgassem úteis para suas buscas.

Multimídia

Figura 17 - Saber mais um pouco nunca é demais...

Fonte: *Pixabay* (2016).[11]

Você pode saber detalhes das tentativas feitas para a sistematização de entradas dos cabeçalhos de assunto, bem como sobre os critérios para o uso de subdivisões, consultando o trabalho "Introdução ao estudo do cabeçalho de assunto", de Hagar Espanha Gomes e Marcílio Teixeira Marinho, no *site*: <http://www.conexaorio.com/biti/cabecalho/cab_ass.htm>.

O fato de *Cutter* ter sido convidado pelo *Ministério da Educação* do governo norte-americano, em 1876, para investigar a situação das bibliotecas nos EUA certamente deu-lhe subsídios para o trabalho que desenvolveria naquele mesmo ano: estabelecer regras para a compilação/construção de catálogos alfabéticos de assunto. Antes disso, não havia critérios para a atribuição de cabeçalhos aos documentos, a indexação era feita com base no julgamento de cada indexador.

[11] Disponível em: <https://pixabay.com/pt/homem-leitura-livro-óculos-ponto-1424877>.

Cutter descartou esquemas de classificação hierárquica, enciclopédicos, e só manteve o princípio da pré-coordenação dos termos (que estabelece a conjugação entre termos para representar assuntos de interesse do leitor numa busca). O cabeçalho já vinha pronto para uso, por exemplo: *laboratory animals; women as authors.* Já falamos, mas não é demais enfatizar, dada sua relevância para a Biblioteconomia, que seu trabalho se intitulou: *Rules for a Dictionary Catalog* (*Regras para um Catálogo-Dicionário*).

É importante ressaltar que o cabeçalho de assunto foi um sistema desenvolvido especificamente para a *LC*, tomando por base a gramática da língua inglesa. Como decorrência disso, o cabeçalho de assunto teve seus três componentes baseados nessa língua: gramática, sintaxe (ordem dos elementos) e o vocabulário, este determinado pelo público por meio de suas buscas feitas nos sistemas. Como dito anteriormente, é importante salientar que a ferramenta foi desenvolvida para a tecnologia existente à época – o fichário, que só possibilitava a recuperação linear da informação. Um exemplo é o seguinte: se o cabeçalho de entrada estabelecido fosse "leis – interpretação", a busca só poderia ser feita dessa maneira. Assim, alguém que quisesse um documento sobre "interpretação de leis", teria que chegar a ele por meio daquele cabeçalho previamente estabelecido, sem chance de ser alterado na hora da busca.

Atualmente, quando pensamos em recuperar uma informação, pensamos em sistemas digitais. As novas tecnologias da informação possibilitam que se faça buscas por palavras, conjugando-as na hora da pesquisa – as chamadas buscas pós-coordenadas. Seguindo o exemplo dado, a busca num sistema automatizado poderia ser feita pelos dois termos conjugados, ou seja, "lei + interpretação" ou" interpretação + lei", não importa a ordem dos elementos.

Multimídia

Velhos tempos, novas-velhas ideias!

Figura 18 - Catálogo de fichas

Fonte: *Pixabay* (2012).[12]

[12] Disponível em: <https://pixabay.com/pt/catálogo-de-cartão-gavetas-madeira-194280>.

Há algum tempo, entende-se que a principal meta a ser atingida quando se desenvolvem sistemas de informação é fazer com que o usuário fique satisfeito com o resultado de suas buscas. É cada vez maior a noção de que o sistema deve ser usuário-orientado. A democratização da informação, no século XIX, fez com que o usuário se tornasse o foco de atenção das bibliotecas e outros serviços de informação. *Cutter*, em 1876, considerou essa ideia como ponto de partida para seu trabalho: observou o usuário comum e seus hábitos de pesquisa e desenvolveu uma nova tecnologia a partir do usuário e para ele. Você não achou *Cutter* inovador? Quer saber mais sobre ele? Visite o *site:* <https://prezi.com/tz-8l1lcb9hv/charles-ammi-Cutter/>.

2.5.4 Princípios de *Cutter* para cabeçalhos de assunto

Cutter achava que a indexação deveria ser feita de forma consistente, de acordo com regras pré-estabelecidas. Segundo ele, a inexistência de regras resultava na desorientação total, tanto do indexador quanto do usuário. Por um lado, o indexador indexava com o termo que melhor lhe aprouvesse. Por outro, o usuário buscava a informação desejada pelo termo que ele imaginava ser o mais conveniente para sua busca. Isso diminuía drasticamente a possibilidade de uma busca satisfazer sua necessidade de informação. Na ausência de regras, tanto fazia procurar sob um termo ou outro, à semelhança do que faziam os indexadores na hora da indexação, quando atribuíam termos aos documentos aleatoriamente.

Para sanar essa situação, melhor seria haver um instrumento com uma série de decisões já tomadas para as entradas dos cabeçalhos. Assim, indexadores teriam maior chance de decidir usar o cabeçalho mais adequado a cada situação e, havendo uma entrada já pré-estabelecida, os usuários não precisariam gastar seu tempo imaginando todas as formas possíveis de entrada. Por exemplo, em vez de imaginar se uma busca deveria ser feita por "tangerina", "bergamota" ou "mexerica", o vocabulário já indicaria a entrada por "tangerina", com remissiva para os outros termos, caso esse tivesse sido o escolhido para representar a fruta no vocabulário. Nesse caso, a entrada do cabeçalho ficaria assim: Bergamota USE TANGERINA; Mexerica USE TANGERINA.

Três princípios básicos criados por *Cutter* nortearam o estabelecimento das regras para a elaboração de cabeçalhos de assunto: o uso, a entrada direta e o princípio sindético, como veremos a seguir:

a) princípio do uso:

   Regras 167-175: Você se lembra que o *Catálogo-Dicionário* foi feito para a *Biblioteca do Congresso dos Estados Unidos*? Então, *Cutter* dizia que, além de o uso ser o supremo árbitro – nesse caso, não o uso do catalogador – mas o do público, ao falar de assuntos –, os cabeçalhos seriam aqueles em torno dos quais haveria consenso quando fossem procurados pela maioria dos americanos com formação escolar. Desse modo, a forma preferida para entrada dos cabeçalhos seria a usada pelo usuário.

   Cutter formulou esse princípio visando as bibliotecas americanas, mas ele pode e deve ser adotado quando se organiza o vocabulário de qualquer SRI. Contudo, o princípio não exclui a necessidade de

analisar fatores que podem influenciar a escolha de um cabeçalho, por exemplo:

- tipo de serviço de informação para o qual o vocabulário é desenvolvido (se para uma biblioteca geral ou especializada, se escolar ou empresarial);
- público a que se destina (cidadão comum ou especializado);
- região geográfica onde o serviço está localizado.

É claro que o vocabulário vai variar de acordo com cada uma dessas circunstâncias. No caso brasileiro, por exemplo, muita atenção deve ser dada aos regionalismos ao se estabelecerem vocabulários de entrada num sistema, como no caso do termo "abóbora". Se você consultar um dicionário, verá que existem vários outros termos considerados sinônimos: "abobra", "bobra", "jerimu", "jerimum".

Que decisão você tomaria nesse caso? Que termo escolheria para a entrada do cabeçalho, se todos são equivalentes? Na região em que moro, o legume é chamado "abóbora"; e onde você mora, como ele é chamado?;

b) princípio da entrada direta específica:

Regra n.º 161: Você deve lembrar que falamos que *Cutter* não simpatizava com o catálogo alfabético-por-classe usado na Europa. Pois bem, ele estabeleceu que um documento deveria ser indexado diretamente pelo assunto específico de que tratava, e não pela classe a que estivesse subordinado. Por exemplo, em um documento sobre "carência de potássio no organismo humano", um dos termos indexadores deveria ser POTÁSSIO, e não "metal alcalino"(termo correspondente à classe a que o elemento químico "potássio" pertence).

Porém, *Cutter* também disse que, quando não se pudesse representar o assunto por uma única palavra, o problema seria contornado de várias maneiras. Aqui começaram as inconsistências dos cabeçalhos de assunto, apesar das regras propostas!

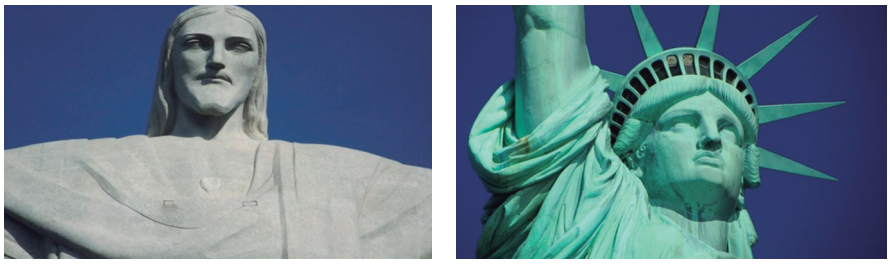
Cutter observou que alguns assuntos seriam representados por expressões formadas por mais de uma palavra e apresentou propostas de entradas diretas para os casos elencados abaixo, sem perder de vista que todo o trabalho de *Cutter* foi desenvolvido com base na gramática da língua inglesa:

- b.1) substantivo precedido de adjetivo. Exemplo: *Ancient history*; *Capital punishment*; *Moral philosophy*;
- b.2) substantivo ligado a outro por preposição. Exemplo: *Penalty of death; Fertilization of flowers;*
- b.3) substantivo precedido de outro substantivo usado como adjetivo. Exemplo: *Death penalty; Flower fertilization;*
- b.4) substantivo ligado a outro pela conjunção *and*. Exemplo: *Church and State;*
- b.5) frase ou sentença. Exemplo: *Women as authors*; *Insect as carriers of plant diseases*.

Explicativo

O vento que venta lá não venta cá!

Figura 19 - Línguas diferentes, sintaxes diferentes



Fonte: *Pixabay* (2013, 2014).[13]

Por que os exemplos apresentados estão em inglês? Porque as línguas têm sintaxes diferentes, ou seja, a ordem dos elementos é diferente em cada expressão linguística. Então, o que se aplica a uma não se aplica à outra, necessariamente. Em inglês, na maioria dos casos, o adjetivo precede o substantivo, caracterizando-o, por ex.: *Ancient history*. Já em português, os substantivos é que precedem os adjetivos, por exemplo: História antiga. Como estamos tratando do trabalho de *Cutter*, desenvolvido com base na língua inglesa, e queremos mostrar que sua tentativa de sistematização para a formulação de cabeçalhos deixou brechas para inconsistências, os exemplos devem ser dados em inglês. Você encontrará exemplos também inconsistentes em português, no estudo "Sistematização da sintaxe de cabeçalho de assunto", de *Lecy Maria Caldas Tôrres*, quando a autora fala da *Lista de Cabeçalho de Assunto do Sistema Bibliodata (LCASB).*

Visite o *site*: <http://www.conexaorio.com/biti/lecy/lecy.htm#5.1> e confira.

Como dito, as saídas dadas por *Cutter* geraram inconsistências. Volte às propostas b.2 e b.3 desta subseção e as compare. Notou algo estranho? Reparou que os cabeçalhos, embora com o mesmo significado, teriam entrada dupla: *Penalty of death* e *Death penalty*; *Fertilization of flowers* e *Flower fertilization*?

Por sua vez, a proposta b.4 sugere a ligação de dois assuntos que podem estar relacionados, mas que têm significados completamente diferentes. Por que juntá-los com a conjunção "e"?

Qual a explicação para casos como esses? As listas de cabeçalhos de assunto na língua portuguesa também estão repletas de inconsistências derivadas desse princípio, por exemplo, os cabeçalhos "Trabalho e trabalhadores", "Insetos úteis e noci-

[13] Primeira imagem: Cristo Redentor. Disponível em: <https://pixabay.com/pt/cristo-cristo-redentorcorcovado-518004>; segunda imagem: Estátua da Liberdade. Disponível em: <https://pixabay.com/pt/estátua-daliberdade-nova-iorque-ny-267948>.

vos", "Igrejas e catedrais", "Costumes e cerimônias", além de vários outros.

A proposta b.5 orienta que, numa frase ou sentença, a entrada deva ser feita pela palavra julgada mais significativa. Ora, o julgamento de significância é absolutamente pessoal, subjetivo. Você pode concluir que espécie de inconsistência essa proposta geraria? Se concluiu que é impossível haver consenso sobre o que é mais significativo e que a proposta é completamente arbitrária, acertou.

No exemplo dado na proposta b.5, a palavra "inseto" foi julgada mais significativa. Talvez sim, para um serviço de informação especializado em Zoologia. E se "doenças das plantas" for mais significativo para um serviço de informação em Botânica? Quem estaria com a razão? Nunca é demais lembrar que os sistemas devem existir em função de sua clientela! Como veremos na Unidade 3, conceitos variam de acordo com os contextos informacionais.

Além das regras relativas à sintaxe da língua inglesa, outras também deram origem a inconsistências.

A regra n.º 164, por exemplo, diz respeito à entrada para assuntos tratados em determinado local. Ela diz que, quando se tratar de tópico estudado com referência a um local, o único método satisfatório será uma entrada dupla: uma para o local, outra para o assunto científico. Ex.: um documento sobre o "ensino no Brasil" teria duas entradas: BRASIL – ENSINO; ENSINO – BRASIL.

Já a regra n.º 165 é relativa à subdivisão para assuntos tratados com referência a algum local. Essa regra estabelece que um assunto geral, estudado com referência especial a algum estado, país ou qualquer outro lugar, deveria entrar pelo local e os cabeçalhos constituiriam a subdivisão. É importante notar que a subdivisão foi permitida apenas nesses casos, quando a entrada for um local. Ex.: um documento sobre as "estátuas de Belo Horizonte" teria somente uma entrada: BELO HORIZONTE – ESTÁTUAS.

A regra n.º 175 tem a ver com a inversão de cabeçalho. Ela estabelece que, se um assunto for expresso por várias palavras, a entrada deve ser feita pela primeira, fazendo-se inversão da frase somente quando outra palavra for decididamente mais significativa, ou quando tal palavra tiver sido usada independentemente, com o mesmo significado do cabeçalho. Ex.: qual deveria ser a entrada para um documento sobre "Bandas de música"? Diretamente por essa expressão ou de forma invertida: MÚSICA, BANDAS DE, ou simplesmente MÚSICA – BANDAS? Reparou que caímos, outra vez, na questão da significância? O que é mais significativo para você pode não ser para mim, não é mesmo?

Ao permitir entradas diretas e entradas invertidas, *Cutter* deixou caminho aberto para inconsistências dos cabeçalhos de assunto e não conseguiu sanar os problemas da subdivisão;

c) princípio sindético:

A ordem alfabética do catálogo-dicionário compromete a reunião de assuntos com significados correlatosque, eventualmente, poderiam ser do interesse em uma busca, mas que não serão pesquisados por estarem dispersos no catálogo; a ordem alfabética não agrega termos com significados próximos. Como exemplo, podemos citar os assuntos "braçadeira", "jarreteira", "penacho alçado" e "tornozeleira" (todos relativos a adornos corporais indígenas), que, se dispostos em ordem alfabética, estariam completamente separados no catálogo.

Ao mesmo tempo, ela é capaz de reunir outros assuntos, cujos significados são absolutamente díspares. Como exemplo, pode-se citar os assuntos "magia negra", "magistério" e "magistrados", que estariam reunidos na letra M, um após o outro, mesmo possuindo significados diferentes.

Para resolver o problema da separação de termos com significados relacionados, *Cutter* propôs o desenvolvimento, nas listas de cabeçalhos de assunto, de um sistema de referências cruzadas, denominado estrutura sindética. Essa estrutura é tradicionalmente composta de referências dos seguintes tipos:

- c.1) "ver": destina-se a remeter um assunto não usado na lista de cabeçalhos de assunto para o adotado como entrada; o assunto que vier depois da notação "ver" será a forma adotada para uso. Essas referências normalmente são feitas em vários casos e os assuntos são considerados equivalentes, como você pode notar nos exemplos a seguir:
  - c.1.1) entre sinônimos (ex.: Abatimento VER DESCONTO; Lume VER FOGO; Insectologia VER ENTOMOLOGIA);
  - c.1.2) entre regionalismos (ex.: Culumim VER CURUMIM; Jabá VER CHARQUE);
  - c.1.3) entre o nome comercial e o genérico (ex.: Modess VER ABSORVENTE HIGIÊNICO; Gilete VER LÂMINA PARA BARBEAR);
  - c.1.4) entre uma sigla e seu nome por extenso (ex.: ONU VER ORGANIZAÇÃO DAS NAÇÕES UNIDAS);
  - c.1.5) entre palavras/siglas/expressões estrangeiras e suas traduções e usos já consagrados na língua vernácula (ex.: GMAW VER SOLDAGEM MIG MAG);
  - c.1.6) entre a segunda parte de um cabeçalho formado por duas palavras ligadas pela conjunção "e", remetendo para a primeira (ex.: Fronteiras VER LIMITES E FRONTEIRAS; Trabalhadores VER TRABALHO E TRABALHADORES)[14];
  - c.1.7) entre nomes científicos e os populares (ex.: Cloreto de sódio VER SAL DE COZINHA);
  - c.1.8) entre assuntos específicos e assuntos genéricos da mesma classe (ex.: Canoa VER EMBARCAÇÃO).

[14] As metodologias para a construção de Linguagens Documentárias baseadas em conceitos e não em cabeçalhos de assunto (ex.: tesauro) orientam para a inclusão de cada conceito em separado, por terem significados diferentes.

- c.2) "ver também": destina-se a agregar assuntos correlatos. Os assuntos mencionados após a notação "ver também" são assuntos sugeridos que, eventualmente, poderiam interessar ao usuário.

Ex.:

FUNCIONÁRIOS

Ver também: DEMISSÃO, DIARISTAS, EFETIVAÇÃO, ESTATUTO DOS FUNCIONÁRIOS, NOMEAÇÃO;

CAPITAL

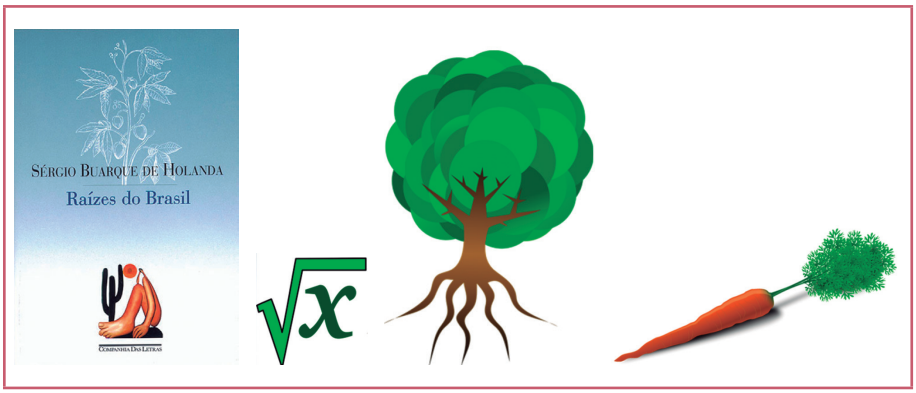
Ver também: BANCOS; CAPITALISMO, DINHEIRO, ECONOMIA, JUROS.

Explicativo

Que sinuca! Que assunto devo atribuir?

Atribuição de assuntos

Figura 20 - Tem tanta coisa que é raiz! Onde encaixar este livro?



Fonte: *Wikimedia Commons* (2008) e *Pixabay* (2012, 2012, 2013).[15]

Embora não fosse uma regra, *Cutter* ainda alertou para o erro de se atribuir um cabeçalho de assunto tomando por base somente o título, e afirmou: "o título governa o catálogo de títulos". Com isso, ele quis dizer que o título não se presta a um catálogo de assunto, porque nem sempre representa o assunto tratado no documento. *Cutter* tem toda a razão. É muito comum encontrar documentos em bases de dados indexados erradamente, e alguns desses erros devem-se, exatamente, à atribuição do assunto a partir da leitura só do título. Um caso célebre ocorrido numa biblioteca brasileira foi o livro *Raízes do Brasil*, do historiador brasileiro *Sérgio Buarque de Holanda*, que foi indexado sob o termo "Botânica". Fica claro que o indexador só se fixou na palavra "raízes"! E é claro, também, que quem procurou o livro sob os

[15] Primeira imagem: livro *Raízes do Brasil*. Disponível em: <https://commons.wikimedia.org/wiki/File:Ra%C3%ADzes_do_Brasil.jpg>; segunda imagem: Raiz quadrada matemática. Disponível em: <https://pixabay.com/pt/raizquadrada-matemática-verde-27895>; terceira imagem: Árvore. Disponível em: <https://pixabay.com/pt/árvore-ramos-raízesesqueleto-153964>; quarta imagem: Cenoura. Disponível em: <https://pixabay.com/pt/cenoura-raiz-vegetaisorange-33625>.

termos HISTÓRIA, CULTURA ou SOCIOLOGIA brasileiras – que são os assuntos de que o livro realmente trata – jamais o encontrou!

Banque o *Sherlock Holmes* e faça a Atividade a seguir!

Figura 21 - *Sherlock Holmes*

Fonte: *Pixabay* (2014).[16]

2.5.5 Atividade

Se você é um bom observador, já deve ter percebido que as propostas de *Cutter* não tiveram bem o final que ele desejou. Os resultados ficaram longe de seu propósito de facilitar as vidas do indexador e do usuário, eliminando entradas indiretas. Pelo contrário, elas criaram muito mais inconsistências do que soluções.

Antes de mais nada, releia as regras citadas no texto, observe bem cada uma delas e as compare. Agora, vamos ao trabalho e descubra:

Utilize o quadro a seguir para suas respostas:

1. Qual a inconsistência das propostas b.2 e b.3 da regra 161?	
2. Qual o paradoxo existente entre as propostas das regras 161 e 175?	

[16] Disponível em: <https://pixabay.com/pt/sherlock-holmes-detetive-462978>.

3. Em que as regras 164 e 165 se chocam?	

Resposta comentada

1. As propostas introduzem inconsistências por permitirem que um único assunto possa ter duas estruturas de cabeçalho, dando lugar a duas entradas diferentes no catálogo. Já vimos que isso acarretará que a informação sobre um mesmo assunto fique dispersa. Nesse caso, poderia ocorrer de 50% da informação sobre fertilização de flores, por exemplo, ficar indexada sob "*Fertilization of flowers*" e os outros 50% sob "*Flowers fertilization*." Embora os dois cabeçalhos signifiquem a mesma coisa, os assuntos estão representados de duas formas diferentes.

2. A proposta b.5 da regra 161 diz respeito a frases/sentenças, que, independentemente de quantas palavras as formem, deveriam sempre entrar de forma direta. Nos exemplos do texto, "*Women as authors*" (Mulheres como autoras) e "*insects as carriers of plant diseases*" (insetos como condutores de doenças das plantas), as entradas permaneceriam exatamente nessa ordem, ou seja, na ordem em que se fala. Contudo, a regra 175 estabelece para o mesmo caso que a entrada seja feita pela primeira palavra, fazendo inversão da frase somente quando outra palavra for decididamente mais significativa. Ora, no primeiro exemplo, o que é mais significativo, "mulher" ou "autora"? No segundo exemplo, qual a palavra mais significativa, "inseto", "condutores de doenças" ou "doenças das plantas"? O problema é que *Cutter* não explicou o que é uma palavra "mais significativa". Pergunta-se: mais significativa para que público? Para qual serviço de informação? Como *Cutter* aplicou em uma regra um conceito tão abstrato? Claro que as entradas ficariam ao sabor dos indexadores, afinal, o que é significativo para você pode não ser para mim.

3. A inconsistência das regras 164 e 165: as duas dizem respeito a assuntos estudados com referência a um local determinado (país, estado, cidade etc.). A diferença é a seguinte: a regra 164 estabelece que duas entradas sejam feitas, uma pelo assunto e outra pelo local a que o assunto se refere. Ex.: um documento sobre religiões brasileiras teria duas entradas no catálogo: 1- Religião – Brasil. 2- Brasil – Religião. Já a regra 165 prescreve para o mesmo caso uma única entrada, com subdivisão pelo local: Brasil – Religião. Como você viu, todas as inconsistências têm resultados insatisfatórios na recuperação da informação. Todas elas levam a sua dispersão, o que faz com que o usuário nunca recupere a totalidade da informação disponível na base de dados.

   Parabéns se acertou alguma das respostas!

2.5.6 Tipologia dos cabeçalhos de assunto

Quanto à forma dos cabeçalhos de assunto, as listas incluem, comumente:

a) cabeçalhos de uma só palavra (ex.: Diplomacia);

b) cabeçalhos de duas palavras, com assuntos muito relacionados ou opostos unidos pela conjunção "e" (ex.: Trabalho e capital; Caça e caçada);

c) frases (ex.: Terapia ocupacional para criança);

d) frases invertidas com suas palavras separadas por vírgula (ex.: Justiça criminal, Administração);

e) substantivos seguidos de adjetivos (ex.: Literatura brasileira);

f) nomes de indivíduos encabeçando trabalhos biográficos e crítica (ex.: Bilac, Olavo – Biografia);

g) palavras seguidas de uma explicação, entre parênteses [ex.: Composição (Musical)].

2.5.7 Características das listas de cabeçalhos de assunto

Figura 22 - Ordem alfabética: vantagem ou desvantagem na busca de informação?

Fonte: *Pixabay* (2013).[17]

Podemos dizer que as principais características das listas de cabeçalhos de assunto são:

a) linguagens alfabéticas: as listas de cabeçalhos de assunto são linguagens alfabéticas que usam termos da própria linguagem natural; incluem, também, mecanismos de controle de sinônimos, homógrafos e outros. Elas determinam uma ordenação alfabética para os arquivos.

   O que você acha de listagens em ordem alfabética? Considera vantagem ou desvantagem para quem faz uma busca para obter informação?

   Bem, vamos considerar que a ordem alfabética seja do conhecimento da maioria das pessoas; que alfabetar é uma tarefa que aprendemos

[17] Disponível em: <https://pixabay.com/pt/alfabeto-letras-deixar-75669>.

desde cedo e que, portanto, estamos familiarizados com essa ordem, que é de fácil manuseio. Vista por esse ângulo, podemos concluir que a ordem alfabética é, sim, uma vantagem.

Porém, quando se trata de uma lista de cabeçalhos de assunto, que se propõe a facilitar o uso tanto para quem trata o documento (indexador), como para quem o procura (usuário), outra forma de apresentar os termos que compõem essa lista seria necessária. Essa alternativa teria a função de tornar possível a visualização ampla e completa do universo dos cabeçalhos incluídos, classificados e organizados segundo um ou mais critérios. Estamos falando de um arranjo sistemático, previsto na apresentação de outras LD, como na de um tesauro, por exemplo.

Outra desvantagem do arranjo alfabético, já abordada anteriormente, quando falamos das funções de um vocabulário controlado (Unidade 1, seção 1.6), é que a simples ordenação alfabética dispersa termos próximos de significado, empobrecendo as buscas;

b) linguagens pré-coordenadas: no cabeçalho de assunto, as palavras/expressões que formam os cabeçalhos são predeterminadas e combinadas antes de seu uso na entrada e saída de dados no sistema. A síntese dos assuntos é representada pelos subcabeçalhos (palavras ou expressões acrescentadas às entradas, geralmente por meio de traços), pelo uso de conectivos (conjunções, preposições) ou por vírgula (no caso de cabeçalho invertido).

Os seguintes cabeçalhos, citados por *Ferraz* (1977) e pela *LC* (2015), servem como exemplos:

- Química médica e farmacêutica;
- Literatura – Estudo e ensino;
- Rio de Janeiro – Vida social e costumes;
- Pontes e viadutos;
- *Government corporations – Finance*;
- Terra Indígena Wedezé (*Brazil*);
- *Brazil* – Terra Indígena Wedezé;
- *Refuse and refuse disposal*;
- *Recreation areas – Washington* (*State*);
- *Washington* (*State*) – *Fishtrap Recreation Area*;
- *Automobiles – Motors – Carburators*;
- *Philosophy, French*.

Tanto quanto a ordenação alfabética das listas de cabeçalhos de assunto, a pré-coordenação dos cabeçalhos também tem prós e contras.

Sua vantagem é o fato de os cabeçalhos de assunto serem formados por elementos pré-conjugados, fato que determina maior detalhamento do assunto buscado, ocasionando maior precisão na busca. Por exemplo, se a pesquisa é sobre "crime contra a administração pública" e o vocabulário tem exatamente esse termo, sorte do pesquisador!

Reparou que nesse cabeçalho os assuntos "crime" e "administração pública" foram coordenados previamente por meio da preposição "contra" e do artigo "a", formando uma frase que não deixa margem a dúvidas quanto ao sentido do cabeçalho? Esse nível de pré-coordenação é que assegura a maior precisão na busca. Nesse caso, se o usuário estivesse buscando documentos sobre aquele assunto, a chance de ele ter sua necessidade de informação satisfeita seria grande.

c) linguagens prescritivas: a pré-coordenação torna os cabeçalhos de assunto linguagens prescritivas, o que poderia ser considerado uma de suas desvantagens. O que isso quer dizer? Nelas, os cabeçalhos já estão escolhidos, pre- determinados e estão ali para serem usados exatamente como da forma indicada, sem possibilidade de modificação. Vamos aproveitar o exemplo dado anteriormente: "Literatura – Estudo e ensino". Suponhamos que determinado serviço de informação tivesse um documento sobre "ensino de literatura nas escolas do estado do Rio de Janeiro". Um dos cabeçalhos a serem escolhi- dos para representar esse documento teria que ser exatamente o constante da lista de cabeçalhos de assunto: "Literatura – Estudo e ensino". Você acha que esse cabeçalho seria satisfatório, mesmo o documento não abordando nada específico relativo a "estudo"? Pare e pense um pouquinho sobre isso.

   Agora, faço outra pergunta a você: por que o termo "estudo" foi atrelado a "ensino" pela conjunção "e", se são dois termos com significados diferentes? E o que me diz sobre "pontes e viadutos", outro exemplo dado anteriormente? Por que esses termos não são entradas separadas? Não são casos intrigantes? Assim como esses, as listas de cabeçalhos de assunto estão cheias!

   Que tal você mesmo verificar isso? Vamos, consulte um dicionário e veja o que os termos significam! Veja se descobre alguma razão para termos com significados tão diferentes serem colocados juntos, pré-coordenados. Aproveite e descubra outros casos semelhantes na *LCSH*;

d) redução das possibilidades de busca: o fato de os cabeçalhos serem linguagens prescritivas reduz as chances de busca do usuário, não lhe dando autonomia para manejar sua busca como quiser. Ele tem que usar aquele e só aquele cabeçalho, que já vem atrelado, em grande parte das vezes, a outro. Como afirma *Cesarino* (1978, p. 276-277):

> A pré-coordenação reduz as chances de o usuário modular sua pesquisa, tornando os cabeçalhos de assunto sistemas fechados de recuperação da informação.

   Vamos observar melhor o cabeçalho "Química médica e farmacêutica". E se o usuário quiser pesquisar somente sobre "química farmacêutica"? Ele não tem opção de separar os dois assuntos na hora da busca. Terá que pesquisar tudo o que estiver indexado com aquele cabeçalho e separar somente os documentos relativos a "Química farmacêutica". Mais uma vez, assuntos diferentes foram postos juntos. Esse é mais um exemplo sobre os efeitos não desejados da pré-coordenação;

e) hierarquia pobre e assistemática: essa é mais uma característica dos cabeçalhos de assunto. Dissemos anteriormente que as listas de cabeçalhos de assunto enumeravam os cabeçalhos alfabeticamente, sem possuir uma base sistemática sólida para a construção e apresentação das relações entre palavras que têm significado próximo – e que poderiam ser alocadas em uma mesma categoria (ex.:"peças para vestuário" e "jaqueta"; "aquarela" e "tinta a óleo"). Se você consultar uma edição das listas da *LCSH* anterior aos anos 1980 e comparar com uma atual, poderá observar que as relações de proximidade entre as palavras eram indicadas, até 1979, pelo símbolo "SA", representando a expressão *"see also"* (ver também) e, a partir daí, pelo símbolo "RT", representando a expressão *"related term"* (termo relacionado)[18]. Porém, tais relações são construídas sem referencial teórico da Ciência da Informação que as justifique, como acontece em outras ferramentas de representação temática do conhecimento, por exemplo,nos tesauros.

2.5.8 Funções do cabeçalho de assunto

As funções do cabeçalho de assunto são aquelas comuns a todas as linguagens documentárias, respeitadas as características peculiares de cada uma, e já foram citadas antes, na parte consagrada às linguagens documentárias em geral. Se você está lembrado, as LD, inclusive os cabeçalhos de assunto, têm como objetivo final as seguintes funções:

a) facilitação dos processos de indexação;

b) recuperação dos conteúdos temáticos dos documentos.

2.5.9 Atividade

Nem tanto ao mar, nem tanto à terra

Apesar das críticas negativas associadas ao uso de cabeçalhos de assunto para a indexação e recuperação da informação, devemos considerar um aspecto: porque será que eles ainda são muito usados nas bibliotecas do mundo inteiro? Será que eles realmente facilitam o processo de indexação? Certamente você observou que há aspectos positivos e negativos em sua utilização.

O que esperamos de você agora? Esperamos que você analise cada uma das afirmações relacionadas a seguir, marque as que acha que estão certas com um C e as que estão erradas com um E. Se precisar, volte ao texto.

1. ( ) O cabeçalho de assunto é baseado na pré-coordenação de assuntos.
2. ( ) O cabeçalho de assunto continua a ser usado porque é um sistema eficaz para a indexação/recuperação da informação.
3. ( ) O usuário de cabeçalho de assunto tem plena liberdade de conjugar os assuntos que ele quer durante uma busca.

[18] Segundo *El-Hoshy* (2001), esse e outros símbolos ("x" e "xx" – usados para remissivas) foram incorporados às listas da *LCSH* desde 1948, tornando-se padrão de suas edições e suplementos, até serem substituídos por designações usadas em tesauros, em 1980: BT (termo superordenado, mais abrangente); NT (termo subordinado, mais específico); RT (termo relacionado) e UF (termo sinônimo)

4. ( ) No cabeçalho de assunto, o termo de entrada é sempre usado tal como falado na linguagem do dia a dia.

5. ( ) As listas de cabeçalhos de assunto são apresentadas, geralmente, só em ordem alfabética.

6. ( ) O cabeçalho de assunto representa os assuntos de um documento.

Resposta comentada

Se você marcou a primeira afirmação com um C, acertou. O cabeçalho de assunto é uma linguagem na qual os assuntos apresentam-se coordenados a priori. Isso quer dizer que eles devem ser usados exatamente como previstos e indicados.

Se marcou a segunda com um E, também acertou. Sua eficácia é bastante questionável. O cabeçalho de assunto continua a ser usado porque ele é a base da *LCSH* e as bibliotecas, na maioria das vezes, possuem essa lista e a usam, na falta de outros instrumentos de controle vocabular.

Não. O usuário deve utilizar os cabeçalhos que tiverem sido indicados, nos quais os assuntos já vêm pré-conjugados, prontos para uso. Você deve marcar a letra E na terceira afirmação.

A letra da vez é a E. Na quarta afirmação, embora o cabeçalho de assunto tenha sido pensado para suas entradas serem diretas, esse princípio nem sempre tem sido observado. As entradas às vezes são diretas, conforme nos expressamos comumente (ex.:"- Doenças respiratórias infantis"), e, em outras, o cabeçalho aparece subdividido (ex.: Paisagens - Proteção) ou na ordem invertida (ex.: Crianças, Desenvolvimento das).

Sim! As listas de cabeçalhos de assunto são comumente apresentadas em ordem alfabética. Nesse caso, a letra a ser marcada na quinta afirmação é a C.

Por fim, na sexta afirmação, a resposta é a letra C. Cabeçalho de assunto é uma palavra ou grupo de palavras que representam os conteúdos dos temas abordados em um documento.

A sequência da Atividade fica, então: C - E - E - E - C - C.

2.6 CABEÇALHOS DE ASSUNTO NA PRÁTICA

Na Unidade 1, seção 1.6, tivemos a oportunidade de abordar alguns aspectos da elaboração de vocabulários controlados de maneira geral, quando vimos como gerar um vocabulário controlado. Agora, vamos ver como fica a elaboração dos cabeçalhos de assunto, especificamente.

2.6.1 Elaboração de cabeçalhos de assunto

Se você está lembrado, um conjunto de cabeçalhos de assunto forma uma Lista de Cabeçalhos de Assunto – que é um dos tipos de LD. Daí procurarmos saber como são construídos.

Poucas são as iniciativas de que se tem notícia sobre princípios para geração de cabeçalhos de assunto. Não existem regras internacionais que orientem sua construção, como acontece com a construção de tesauros. Para estes, existem regras específicas, que serão vistas na Unidade 3.

A ausência de normas voltadas para a construção de cabeçalhos de assunto pode ser atribuída a vários motivos. Talvez o mais influente seja a existência da *LCSH*. A *LC* determina os cabeçalhos a serem usados, sem, contudo, enunciar as regras que regem sua elaboração.

Como já dito aqui, a *LCSH* continua a ser usada pela maioria das bibliotecas do mundo. Apesar de suas inconsistências, a *Lista* e suas atualizações permanentes podem ser adquiridas por qualquer serviço de informação. Isso faz com que os profissionais encarregados do tratamento dos documentos sintam-se, digamos, confortáveis. Pelo menos, o marketing da *LC* para a *Lista* mostra-se bem eficiente para seduzir os bibliotecários ao redor do mundo! Basta olhar a página do Serviço de Distribuição de Catalogação da *LCSH* para constatar isso, quando anunciam:

> [a Lista] de Cabeçalhos de Assunto da Biblioteca do Congresso (LCSH) é a única lista de cabeçalhos de assunto aceita no mundo todo como padrão. A LCSH é a lista de cabeçalhos de assunto mais completa do mundo. É a única ferramenta de que bibliotecário algum pode prescindir[...] A LCSH é a melhor forma de manter seus cabeçalhos de assunto em dia – a chave certa para catalogação e recuperação temática. (LC, 2015, *on-line*, tradução nossa).

No caso brasileiro, a tradição da *LCSH* é grande e sua utilização é feita sem observância de suas inconsistências. Não são poucas as críticas aos cabeçalhos de assunto em geral e, particularmente, à *LCSH*, na tentativa de mostrar sua fragilidade como ferramenta para a representação e recuperação da informação. Mas isso será visto na próxima seção, quando fizermos uma avaliação desse instrumento.

Se os cabeçalhos de assunto mostram-se frágeis em listas tradicionalmente utilizadas, imagine a situação de bibliotecas que não possuem qualquer lista para controlar seu vocabulário na indexação/recuperação da informação! O que fazer nesse caso? Não há outro jeito, a não ser usar palavras da linguagem natural, que é livre e isenta de qualquer controle! Talvez você já tenha enfrentado uma situação dessas. Talvez, até, você trabalhe em uma biblioteca assim, sem um instrumento auxiliar de indexação!

Também é comum encontrarmos bibliotecários às voltas com a elaboração de vocabulários controlados sem que nenhuma orientação lhes seja dada! Como construí-los, então?

Vimos, anteriormente, que *Cutter* formulou regras para cabeçalhos de assunto e que algumas propostas geraram inconsistências nos cabeçalhos. Porém, os princípios de *Cutter* – uso, entrada direta e princípio sindético (estrutura de remissivas) – podem orientar aqueles que se vejam envolvidos com o desenvolvimento de algum vocabulário controlado. É

só conferir. Volte aos "Princípios de *Cutter* para cabeçalhos de assunto", citados anteriormente nesta seção.

Até onde se sabe, as orientações para a construção de cabeçalhos de assunto são iniciativas isoladas de instituições que necessitam de um mínimo de padronização do vocabulário tradicionalmente usado. O que acontece normalmente é que os vocabulários vão sendo construídos ao longo do tempo para suprir as necessidades de tratamento dos documentos, porém sem regras explícitas que norteiem sua construção.

Multimídia

Construção de cabeçalhos de assunto: vale conferir!

Figura 23 - Página inicial do *site* da *Biblioteca Digital do Senado Federal*

Fonte: Portal do *Senado Federal* (2017).[19]

Se você quiser conhecer iniciativas sobre a construção de cabeçalhos de assunto, basta consultar um dos seguintes estudos:

a) *Cabeçalhos de assunto: manual para estudantes de José Carlos Abreu Teixeira,* editado pela *Universidade Federal Fluminense (UFF)* em 1979;

b) *Análise e representação de assunto: diretrizes para a Rede Virtual de Bibliotecas – Congresso Nacional – RVBI*, elaborado pelo *Senado Federal* em 2007. Disponível em: <http://www2.senado.leg.br/)>;

c) *Diretrizes para a elaboração de um cabeçalho de assunto*, elaboradas como subsídio aos trabalhos desenvolvidos pela biblioteca da *Universidade Federal do Rio Grande do Sul* (UFRGS). Disponível em: <http://rogeriopetrinialmeida.blogspot.com>;

d) *Manual para elaboração do vocabulário controlado da Biblioteca do Tribunal de Justiça do Estado do Rio de Janeiro*, de 2012, que tem por objetivo estabelecer critérios para a construção do vocabulário controlado usado na

[19] Disponível em: <http://www2.senado.leg.br/bdsf/>.

representação dos assuntos no que se refere às tarefas de indexação legislativa, análise de artigos de periódicos e classificação dos livros, folhetos e acervo digital. Disponível em: <http://www.tjrj.jus.br/documents/10136/18813/MAN-DGCON-038-01-REV-0.pdf>. Acesso em: 16 jul. 2017.

Multimídia

Curiosidades sobre a *LCSH*

Figura 24 - Manual da *LCSH*

Fonte: *Library of Congress* (20--?).[20]

Os cabeçalhos de assunto começaram a ser estabelecidos em 1898, mas a primeira edição da *Lista* data do período entre 1909 e 1914, quando ela foi impressa em partes. Seu nome era, então, *Subject Heads Used in Dictionary Catalogues of the Library of Congress.*

A cada ano, milhares de novos cabeçalhos são incluídos.

Atualmente, ela apresenta: uma rede de remissivas (USE/UF – use e usado por), remetendo termos não usados para usados e vice-versa; outra de referências (SA – *see also*), indicando o uso de termos afins; uma estrutura de relações hierárquicas, mantendo próximos termos que pertencem a uma mesma classe temática); e uma estrutura de relações associativas (mantendo próximos termos que tenham significados afins).

[20] Disponível em: <http://www.loc.gov/aba/publications/FreeSHM/freeshm.html>.

Todas as informações necessárias para o uso correto da *LCSH*, você poderá obter em seu *Manual*, no endereço: <http://www.loc.gov/aba/publications/FreeSHM/freeshm.html>.

2.6.2 Panorama dos cabeçalhos na *LCSH*

Falamos do uso ainda intenso da *LCSH* nas bibliotecas brasileiras. E da grande probabilidade de você, já como profissional, deparar-se com esse instrumento para desenvolver seu trabalho.

Nada mais natural, então, do que lhe dar algumas informações básicas sobre como funcionam os cabeçalhos na *Lista*[21]. Apresentaremos uma síntese feita a partir das folhas da Introdução da 37ª edição da *LCSH*, mostrando uma tipologia dos cabeçalhos, a forma como eles se relacionam na *Lista* e os tipos de suas subdivisões:

a) tipologia dos cabeçalhos

- a.1) os cabeçalhos de entrada são grafados em negrito:

  Ex.: ***Alphabet***

  Life on other planets

  Nuclear physics

- a.2) um cabeçalho pode vir acompanhado da indicação *(May Subd Geog)*, mostrando que pode ser subdividido por nomes de lugares;
- a.3) inclui notas de escopo: orientações sobre o significado ou aplicação do cabeçalho, usadas para assegurar consistência em seu uso. Contextuam o significado da palavra tal como aplicado nos catálogos da *LC*;
- a.4) os cabeçalhos consistem em uma ou mais palavras. Um cabeçalho de uma só palavra é normalmente um substantivo. Ex.: *Dogs; Schools; Viscosity;*
- a.5) os conceitos são normalmente mencionados no singular; os objetos, no plural, embora exceções sejam encontradas. Ex.:*Viscosity*; *Songs;*
- a.6) cabeçalhos de duas palavras contêm um adjetivo e um substantivo. Ex.: *Nuclear physics; Local taxation*;
- a.7) cabeçalhos invertidos[22]:
  - a.7.1) com duas palavras: comumente usados com adjetivos, quando descrevem línguas, nacionalidades ou etnias, ou aqueles qualificados por período temporal. Ex.: *Lullabies, Urdu; Songs, French; Art, American; Drawing, Australian*; *Art, Medieval.*

[21] Os cabeçalhos serão mantidos na língua original da *LCSH*, tendo em vista que alguns dos exemplos não fazem sentido quando traduzidos para o português, já que a gramática e a sintaxe das duas línguas são diferentes. Exemplo: o termo "*databases*", atualmente em uso na *LCSH*, já teve uma forma antiga na lista: "data bases", havendo uma remissiva da forma antiga para a atual. Na língua portuguesa, o termo é "base de dados" e não haveria necessidade de uma remissa desse tipo, porque a palavra só existe escrita dessa forma, separadamente.

[22] Apesar das inversões, é dito que a política da *LCSH* é manter as entradas em ordem direta.

- a.7.2) independentemente dos casos citados em a.7.1, há inversões encontradas com o mesmo objetivo: manter o substantivo como palavra de entrada. Ex.: *Love, Maternal*; *Injections, Intramuscular*.

- a.8) cabeçalhos com subdivisão de tópicos, com o nome da classe como entrada. Ex.:*Photography - Studios and dark rooms*; *Railroads - Timetables* (cabeçalhos invertido por hífen); *Art, Byzantine* (cabeçalho invertido por vírgula); *Cook (Fish)* (cabeçalho invertido por parênteses);

- a.9) nomes de acidentes geográficos invertidos, com a parte significativa do nome do acidente na frente, separada do nome genérico do acidente por vírgula. Ex.:*Amazon, River*; *Erie, Lake*;

- a.10) cabeçalhos formados por frases com conjunções e preposições. Ex.:*Technology and civilization* (expressando conjugação de duas ideias); *Criminal justice, Administration*(inversão da frase); *Photography of birds*; *Occupational therapy for children* (ordem direta na frase);

- a.11) vários cabeçalhos contêm números de classificação da *LC*. Ex.:*Hydraulic presses* [TJI 460]; *Norwegian language* [PD-2571-PD2699];

b) relações entre os cabeçalhos

- b.1) referências USE e UF: essas referências são feitas de um termo não autorizado pela *Lista*/não preferido (UF) para uso para um autorizado/preferido para uso (USE).

  Ex.: *Cars* (*Automobiles*)
  USE *Automobiles*
  Automobiles
  UF *Cars (Automobiles)*

  Os códigos USE (use, do verbo usar) e UF (*used for*/ usado para) funcionam como recíprocos. Dessa forma, cada referência USE gera, automaticamente, uma recíproca UF, como mostra o exemplo anterior.

  As referências USE/UF são feitas para sinônimos, variações ortográficas, formas variadas de expressão, formas de construção alternadas, formas antigas de cabeçalhos etc.

  Ex.: *Raw foods*
  UF Food, Raw [Former heading]
  Uncooked food

  A referência que representa uma forma antiga de cabeçalho é seguida pela legenda [*Former heading*].

  Ex.: *Databases*
  UF Data bases [Former heading].

- b.2) relações hierárquicas: os cabeçalhos de assunto são ligados a outros por meio das referências expressas por:

- b.2.1) *broader terms* (BT), indicando termos mais abrangentes. O código BT precede o cabeçalho, representando a classe da qual ele é membro.

  Ex.: *Exterior lighting*
  BT *Lighting*;

  Vehicles
  BT *Transportation*.

- b.2.2) *narrower terms* (NT), indicando termos mais específicos. O NT precede o cabeçalho de assunto, representando, na maioria dos casos, um membro da classe representada pelo cabeçalho sob o qual o NT aparece.

  Ex.: *Lighting*
  NT *Exterior lighting*;

  Transportation
  NT *Vehicles*.

  Como se pode observar nos exemplos, na estrutura hierárquica, os códigos BT e NT estão sempre numa relação de superordenação e subordinação, respectivamente, ambos funcionando como recíprocos.

- b.3) relações associativas: essa relação, representada pelo código RT (*related term*), liga cabeçalhos associados de alguma forma que não seja a hierárquica.

  Ex.: *Birds*

  RT *Ornitology*.

- b.4) referências genéricas: são feitas a um grupo inteiro de cabeçalhos e não a cabeçalhos específicos. Até a presente 37ª edição, essas referências vem sendo expressas por meio do símbolo SA (*see also*/ ver também).

  Ex.: *Woodworking industries*

  SA names of specific industries, ex.: *Furniture industry and trade*.

  - b.4.1) as referências genéricas também são feitas de um cabeçalho genérico para um grupo de cabeçalhos, todos começando pelo mesmo nome.

    Ex.: *Chemistry*

    SA headings beginning with the word Chemical, ex.: *Chemical survey*, *Chemical analysis*.

  - b.4.2) outras referências genéricas podem levar à subdivisão.

    Ex.: *Economic History*

    SA *subdivision Economic conditions under names ofcountries, cities etc under classes of persons and ethnic groups*, ex.: *Brazil - Economic history.*

- b.5) referências genéricas do tipo USE.

  Ex.: *Access control*

  USE *subdivision Access control under types of archives, records, computers, computer networks, and statistical and data-gathering services, e.g., Computers - Access control; Psychiatric records - Access control.*

c) subdivisões: as aplicações dos cabeçalhos da *LCSH* requerem uso extensivo de subdivisões de assunto, como meio de combinar um número de conceitos em um só cabeçalho. Ex.: *Massachussets-History-New Plymouth, 1620-1691.*

- c.1) categorias de subdivisões: em geral, quatro categorias de subdivisões são reconhecidas na *LCSH*:

  - c.1.1) por tópicos: as subdivisões por tópicos são usadas sob os cabeçalhos principais ou sob outras subdivisões para limitar o conceito expresso pelo cabeçalho de determinado subtópico.

    Ex.: *Corn - Harvesting*;

    Women - Employment;

    Automobiles - Motors - Carburators;

  - c.1.2) por tempo/cronologia: as subdivisões cronológicas são usadas para limitar um cabeçalho ou um subcabeçalho a um período de tempo determinado.

    Ex.: *Philosophy, French - 18th century*;

    Art, Chinese - To 221 B.C.;

  - c.1.3) por local/geográfica: a designação (*May Subd Geog*), depois de um cabeçalho de assunto ou subdivisão, indica que eles podem ser seguidos pelo local geográfico. O contrário é indicado pela expressão (*Not Subd Geogr*).

    O cabeçalho *Labour supply* (*May Subd Geog*) indica que o cabeçalho pode ser dividido por qualquer local.

    Ex.: *Labor supply - France* **ou**

    Labor supply - France-Paris **ou**

    Labor supply - France-Paris-Marseille;

  - c.1.4) por forma: as subdivisões de forma são usadas para indicar a forma na qual o material sobre determinado assunto é organizado e apresentado, ex.: congressos, dicionários, periódicos. Como tais, são postas como o último elemento do cabeçalho. Podem ser adicionadas praticamente a qualquer tópico e são raramente impressas na *LCSH*. Essas subdivisões tornaram-se *free-floatings* a partir de 1974, como será visto adiante.

    Ex.: *United States - History - Periodicals*

A maioria dos subcabeçalhos de forma são indicados na *Lista* por uma referência genérica SA (*see also* – ver também), sob o cabeçalho representado como um todo.

Ex: *Periodicals*

> *SA sudivision Periodicals under specific subjects, e.g., Engineering - Periodicals*;

- c.1.5) subdivisões *free-floating*: são combinações de cabeçalhos já estabelecidos, criadas com base em regras e não em autorizações específicas da *LCSH*.

  As *free-floatings* podem ser subdivisões de forma; de classes de pessoas e grupos étnicos; de instituições corporativas, pessoas e famílias; de lugares. (LIBRARY OF CONGRESS, 2014);

- c.1.6) subdivisões *Pattern heads*: válidas para classes das quais nem todas as subdivisões são impressas na *LCSH*. Por exemplo, o cabeçalho "*Joints – Biopsy*" não é impresso na *LCSH*. Entretanto, é um cabeçalho válido, porque a subdivisão *"biopsy"* aparece sob "*heart*", cabeçalho que pode ser considerado da mesma classe de*"joints."* "*Heart*" (coração) e *"joints"* (juntas, articulações) são ambos partes do corpo de seres do reino animal.

2.6.3 Situação dos cabeçalhos de assunto no Brasil

Figura 25 - Fachada da *Biblioteca Nacional do Rio de Janeiro*

Fonte: *Flickr* (20--?).[23]

A *Biblioteca Nacional do Rio de Janeiro (BN)* (Figura 25) baseia seus cabeçalhos de assunto na *LCSH*.

[23] Autor: André Melo. Disponível em: <https://www.flickr.com/photos/ministeriodacultura/with/5142787117/>.

Atenção

No século passado, precisamente na década de 1980, a *BN* iniciou um trabalho de revisão de seus cabeçalhos de assunto, numa tentativa de sistematizá-los. Tal iniciativa se deu no âmbito do Projeto Cabeçalho de Assunto Unificado (CAU) e se ocupou, primeiramente, das divisões de forma dos cabeçalhos. O trabalho resultou em duas edições, de 1983 e 1984. A primeira delas consistiu na tradução, com leves modificações, das subdivisões da *LCSH* chamadas *free floatings*, já citadas anteriormente (c1.5). A segunda ocupou-se das **divisões de forma** existentes no catálogo da *BN*, além de acrescentar cabeçalhos considerados absolutamente necessários:

a) **subdivisão de forma:** qualquer palavra que possa ser combinada a outra para representar temas/assuntos constantes em documentos, que representem a manifestação física de uma obra. Ex.: Economia – *Dicionários*; Engenharia – *Abreviaturas*; Automóveis – *Desenhos*; Crianças – *Poemas*; Vargas, Getúlio, 1883-1954 – *Poemas*; Língua catalã – *Mapas;*

b) **subdivisão *free-floating*:**

> Termo que se refere a uma forma ou subdivisão de assunto que pode ser usada sob outros assuntos indicados [...], [porém] sem que um registro oficial tenha sido criado para cada combinação de cabeçalho/subdivisão que pudesse ser necessária" (LC, 2014, *on-line*, tradução nossa).

A segunda edição traz um conjunto de formas julgadas mais usuais ou expressivas, devidamente conceituadas, instruções de como usá-las e exemplos de aplicação prática. Dentre as várias opções de divisões de forma incluídas nessa edição, podem-se citar: abreviatura, anuário, atlas, bibliografia, biografia, catálogos, desenhos, dicionários, discografia, edições críticas, estatísticas, filmografia, guias, normas, peças teatrais, periódicos, poemas, programas de computador, retratos, sermões, tabelas e muitas outras. Ex.: *Assis, Machado de*, 1839-1908 - *Bibliografia*; Insetos - *Catálogos;* Engenharia de tráfego - *Programas de computador*; Marés - *Tabelas.*

Multimídia

Divisões de forma

Figura 26 - Quem procura sempre encontra!

Fonte: *Pixabay* (2013).[24]

Certamente você já se deparou com divisões de forma em um catálogo de biblioteca e nem sabia que se chamavam assim. Se quiser ficar mais familiarizado, visite o *site* <http://www.conexaorio.com/biti/> e entre no *link* do trabalho realizado para a *BN* intitulado *Divisões de forma: instruções para seu uso,* coordenado pela professora *Hagar Espanha Gomes*.

2.7 AVALIAÇÃO DOS CABEÇALHOS DE ASSUNTO DA *LCSH*

Já dissemos que, apesar de seus problemas, cabeçalhos de assunto continuam a ser usados em várias bibliotecas do mundo, inclusive nas brasileiras. No Brasil, o uso da *LCSH* ainda é alto, por tradição ou porque não se tem mesmo outro instrumento auxiliar mais adequado para o tratamento da informação.

[24] Disponível em: <https://pixabay.com/pt/inspetor-homem-detetive-masculino-160143>.

Muitas são as críticas que os cabeçalhos de assunto recebem. Vários autores mencionam a falta de base científica para sua elaboração; as inconsistências das regras, nunca resolvidas; os problemas na recuperação da informação, devidos, principalmente, à linguagem pré-coordenada, que impossibilita a combinação de assuntos que não tenham sido estabelecidos previamente nos cabeçalhos, entre outras. (EL-HOSHY, 2001; FISCHER, 2005; KIRTLAND; COCHRANE, 1982; WELLISCH, 1995; YI; CHAN, 2010).

Especificamente quanto à *LCSH*, esses autores expressam suas opiniões da seguinte forma:

a) se a *LCSH* não mudar, mais cedo ou mais tarde ela será abandonada;

b) se ela quiser se manter como uma ferramenta útil na era do computador, terá que fazer alguma coisa que envolva uma grande revisão;

c) a *LCSH* peca pela falta de terminologia adequada e pormenorizada para as necessidades de certas áreas do conhecimento;

d) ela foi uma solução para os problemas do século XIX, portanto já é hora de repensar toda a questão dos sistemas de acesso por meio de vocabulário controlado e aplicar o novo pensamento para possibilitar acesso à *web*. É fato comprovado que a *LCSH* é um sistema que precisa de muito trabalho para seu desenvolvimento;

e) o sistema *LCSH* precisa de muitos aperfeiçoamentos que facilitem seu uso.

Outro motivo bem convincente para as críticas à *LCSH* pode ser atribuído a sua inadequação à era digital. Como já dito, enfaticamente, não podemos esquecer que o cabeçalho de assunto foi criação do século XIX e que, na época, foi uma inovação louvável. Porém, cabe destacar que ele foi a solução para o uso em catálogos de fichas. Atualmente, sua estrutura super-rígida, em virtude da pré-coordenação, já não faz sentido em recuperação da informação *on-line*.

A existência de novas tecnologias concede mais autonomia aos usuários na hora da busca através da pós-coordenação de termos, escapando da linearidade da pré-coordenação, onde é imposta ao usuário uma única forma de procurar o que deseja, sem a liberdade de combinar os assuntos de que ele realmente necessita e da forma como quiser.

Exemplificando: atualmente, se um usuário deseja informação sobre "preços de pousadas em Natal", ele tem três pontos de acesso para iniciar a busca, não interessando em que ordem procure. Ele pode combinar "Natal" com "pousadas" e "preços","preços" com "Natal" e "pousadas" ou, ainda, "Natal" com "preços" e "pousadas". Acha que ele não conseguiria essa informação? Você também pode fazer uma tentativa!

Quem está habituado a utilizar a internet sabe que "todos os caminhos levam a Roma"! Basta digitar, em um *site* de busca, uma palavra ou palavras sobre as quais se queira informação e é quase certo que se chegue a ela, ainda que seja necessária uma filtragem na resposta para extrair o que realmente interessa. Isso para não falar dos avanços da *web* semântica, que privilegiam o significado das palavras e expressões (ver Unidade 1, seção 1.5 e Unidade 4).

Resumindo, pode-se dizer que as críticas à *LCSH* são de duas ordens:

a) relativas aos problemas de sua própria natureza, colocando em jogo sua eficácia como ferramenta de orientação para a indexação e recuperação da informação;

b) relativas à existência de tecnologias mais modernas, que facilitam a vida do usuário, já que a competição de mecanismos para busca automatizada no texto é cada vez maior.

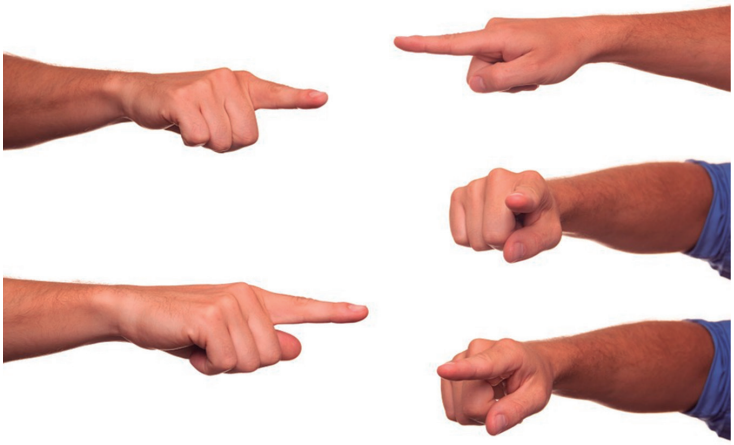
Percebe-se também que essas críticas têm um ponto em comum quando grande parte dos autores que se dedicam ao assunto afirmam que a *LCSH* precisa mudar.

Contudo, em meio aos problemas, vez por outra a literatura deixa transparecer certo desapontamento caso a *LCSH* venha a terminar. Alguns autores afirmam que, apesar das falhas, sem ela, o futuro do acesso aos catálogos e coleções seria grandemente empobrecido. Outros entendem que é muito mais eficiente o uso de cabeçalhos de assunto, defendendo que são mais informativos, por serem sistemas pré-coordenados e, dessa forma, contribuírem para a maior precisão na recuperação da informação.

Multimídia

Críticas à *LCSH*

Figura 27 - Crítica pela crítica, NÃO!



Fonte: *Pixabay* (2014).[25]

Criticar por criticar não leva a nada. Nos artigos sugeridos a seguir, você encontrará um bom repertório das críticas feitas à *LCSH*, os argumentos que as sustentam, avaliações e tentativas de melhorá-la:

> SVENONIUS, Elaine; MCGARRY, Dorothy. Objectivity in evaluating subject heading assignment. **Cataloging & Classification Quarterly**, [S.l.], v. 16, n. 2, p. 5-40, 1993;

[25] Disponível em: <https://pixabay.com/pt/designar-picando-dedo-indicar-427537>.

KIRTLAND, Monika; COCHRANE, Pauline. Critical views of LCSH-Library of Congress Subject Headings: a bibliographic and bibliometric essay. **Cataloging & Classification Quarterly**, [S.l.], v. 1, n. 2-3, p. 71-94, 1982;

YI, Kwan; CHAN, Lois Mai. 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**, [S.l.], v. 61, n. 4, p. 677-687, 2010.

Se você tem facilidade em ler na língua inglesa, explore também este estudo elaborado pela *Biblioteca da Universidade de Princeton* (EUA):

SHAVA, Tsering Wangyal**. Critical analysis of Library of Congress Subject Headings**. New Jersey: Princeton University, 1998. Disponível em: <http://www.princeton.edu/~shavatw/LCSH.html>. Acesso em: 6 ago. 2015.

2.8 O QUE EXISTE ALÉM DA *LCSH*?

Como dito anteriormente, os sistemas que adotam cabeçalhos de assunto baseiam-se na *LCSH* e, portanto, originam-se nos princípios adotados para o catálogo-dicionário idealizado por *Cutter* (ver subseção 2.5.4). Dentre as listas de cabeçalhos de assunto existentes, algumas das mais conhecidas são:

a) *LCSH*: a mais difundida nos EUA e em grande parte dos países do mundo; reflete o acervo da *LC*;

b) *Sears List of Subject Headings*: adaptação da *LCSH* para bibliotecas pequenas e médias norte-americanas. Foi mais aceita nas bibliotecas britânicas.

No Brasil, as mais difundidas foram:

a) *Lista de encabezamientos de materias para bibliotecas* (de *C. Rovira* e *J. Aguayo*, 1967. 3v.): adaptação da *LCSH* e da *Sears* para aplicação nas bibliotecas latinoamericanas;

b) *Relação de assuntos para cabeçalhos de fichas*: aplicável a pequenas coleções, com cabeçalhos gerais (FERRAZ, 1977);

c) *Lista compilada de cabeçalhos de assunto da Rede Bibliodata* (IBICT, 2013): utiliza linguagem pré-coordenada, fundamentada na *LCSH*, respeitando particularidades da língua portuguesa.

Todas as listas de cabeçalhos de assunto têm falhas que variam de uma para a outra, mas pode-se dizer, resumidamente, que todas têm pontos críticos em comum. Confira:

a) não são eficazes o bastante para representar assuntos específicos, numa época em que eles se tornam cada vez mais complexos;

b) os cabeçalhos de assunto não são construídos sistematicamente;

c) os sistemas de referências cruzadas também não são construídos sistematicamente;

d) as regras atualmente existentes para subdivisões dos cabeçalhos continuam gerando inconsistências;

e) as listas são produto de uma tecnologia que representou avanços para a época em que foi idealizada (século XIX), mas que, no momento, encontra-se em desalinho com tecnologias mais avançadas, precisando de mudanças substanciais.

Será que agora você já tem sua própria opinião sobre o uso das listas de cabeçalhos de assunto nos SRI?

Você está convidado a participar da atividade a seguir e, se for curioso, vai descobrir coisas interessantes, que vão ajudá-lo a usar as listas de forma mais crítica. Mãos à obra!

2.8.1 **Atividade**

Começando a praticar...

Talvez você seja "convidado", um dia, a usar uma lista de cabeçalhos de assunto para indexar o material de um SRI. Nesse caso, é bom estar preparado para usá-la criticamente, portanto, da forma mais apropriada. A seguir, estão cabeçalhos retirados de uma lista bastante difundida no Brasil há algum tempo. Agora é com você: o que deve fazer? Analise cada conjunto e comente, no espaço a seguir, por que são inconsistentes. Se houver mais de um cabeçalho no conjunto, compare-os.

O primeiro já está feito. Continue a partir do número 2.

Cabeçalhos:

(1)

Água - Purificação
Água - Taxa de consumo
Código de águas

(2)

Catolicismo **Ver Igreja católica**
Budismo
Protestantismo

(3)

Crianças - Doenças cardíacas
Doenças respiratórias infantis
Doentes mentais

(4)

Crianças, Desenvolvimento das
Educação de crianças

(5)

Organização do trabalho
Organização escolar
Bibliotecas - Organização

(6)

Simbolismo
Impressonismo (Arte)
Arte moderna

(7)

Insetos úteis e nocivos

1	Os dois primeiros cabeçalhos têm subdivisões de assunto e o terceiro, não, sua entrada é direta. Por que não entram diretamente como o terceiro, desta forma: "Purificação da água" e "Taxa de consumo da água"? Não seria assim que o usuário procuraria?
2	
3	
4	
5	
6	
7	

Resposta comentada

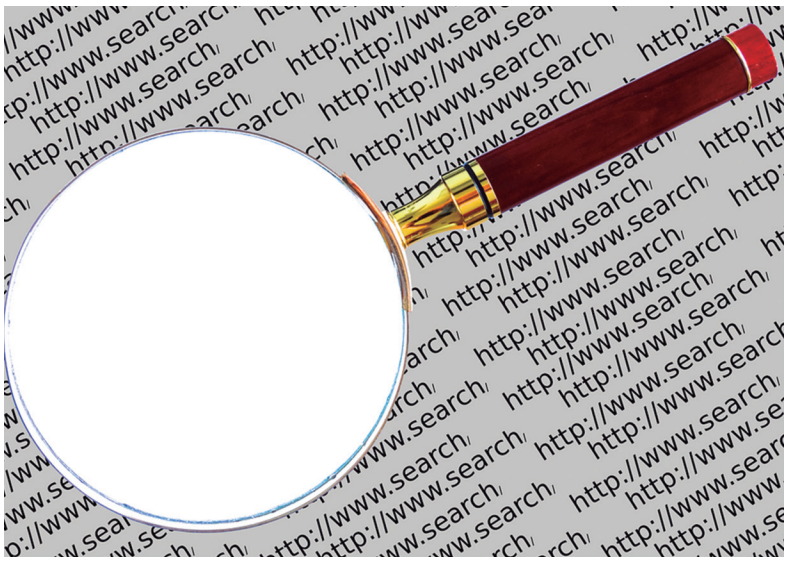
2. Por que "protestantismo" e "budismo" entram direto por seus nomes e "catolicismo" não?
3. Por que não "doenças cardíacas infantis" e "doenças mentais", para ficar condizente com "doenças respiratórias infantis"? Assim, todas as entradas seriam por tipos de doenças. Como saber, numa busca, tudo o que existe disponível sobre "doenças"?
4. Por que não usar o mesmo critério de entrada de "educação de crianças" e entrar por "desenvolvimento de crianças",

diretamente? Por que usar a inversão em "crianças-desenvolvimento das"?

5. Por que os dois primeiros cabeçalhos entram por "organização" e o terceiro entra por "biblioteca"? Não seria a ordem natural das coisas todos entrarem de forma consistente por "organização"? Qual terá sido a razão dessa decisão?
6. Repararam que "simbolismo" e "impressionismo" estão entrando especificamente pelos nomes dos movimentos artísticos? E por que "arte moderna" também não entrou por "Modernismo (Arte)"? Aliás, porque somente "impressionismo" está qualificado com "(Arte)"? Isso não deveria ser regra para todos? Não deveria ser também "Simbolismo (Arte)" e "Impressionismo (Arte)"?
7. Existem insetos nocivos que sejam úteis? Pode ser...! De qualquer jeito, "útil" é um conceito e "nocivo" é outro completamente diferente. Aliás, são conceitos antagônicos! Por que teriam que ser juntados num só cabeçalho? E a busca do usuário que quiser só material sobre "animais nocivos", como fica?

2.9 OS CABEÇALHOS DE ASSUNTO EM TEMPOS DE *WEB*, COMO ESTÃO?

Figura 28 - Pesquisando assuntos na *web*



Fonte: *Pixabay* (2014).[26]

[26] Disponível em: <https://pixabay.com/pt/motor-de-busca-pesquisa-463398>.

As iniciativas de cabeçalhos de assunto na *web* não trouxeram muitas novidades. Apesar da nova tecnologia, os *softwares* dos catálogos disponíveis na rede não sofreram mudanças substanciais. A questão da recuperação da informação por assuntos continua pré-coordenada em muitos sistemas *on-line*, sem permitir a combinação de assuntos na hora da busca, ou seja, sem permitir a pós-coordenação.

Vamos dar exemplo de dois catálogos *on-line*: um, o de cabeçalhos de assunto da própria *LCSH*; o outro, baseado neles:

a) *LCSH*: permite a busca por assuntos preexistentes na *Lista*, direcionando-a para assuntos mais genéricos, mais específicos e correlatos. Ex.: uma busca por *"protestantism"* traz a seguinte estrutura:

   Protestantism

   BT Christianity

   NT Arminianism

   Angelicalism

   Fundamentalism

   Holiness movement

   RT Protestant churches

   Reformation

   Como citado antes, a partir dos anos 1980, a estrutura de relacionamentos entre os termos da *LCSH* (BT, NT e RT) passou a refletir a estrutura originalmente adotada nos tesauros (outra ferramenta para indexação/recuperação da informação que será vista na Unidade a seguir).

As relações de equivalência (sinônimos e quase sinônimos) também são permitidas na *Lista*. Referências são feitas para variantes semânticas e lexicais, de termos e frases equivalentes para cabeçalhos válidos. Ex.: "*parabolic antennas*" USE *SATELLITE DISH ANTENNAS*; "*heart attack*" USE *MYOCARDIAL INFARCTION*. A *LCSH on-line* ainda inclui termos equivalentes de outras listas de cabeçalhos de assunto;

b) *Handbook of Latin America Studies (HLAS on-line*): bibliografia multidisciplinar sobre a América latina editada pela *LC*, publicada desde 1936. Permite a busca por palavras-chave, localizando trabalhos sobre o assunto pesquisado, além de conduzir a assuntos correlatos de outros trabalhos. Ex.: uma busca feita por "doença mental" não só recupera trabalhos sobre esse assunto, como também pelos assuntos relacionados, como"prisioneiros políticos – Argentina", por exemplo.

Multimídia

Não somos *São Tomé*, mas vale ver para crer!

Figura 29 - Cabeçalhos de assunto na internet

Fonte: *Pixabay* (2016).[27]

Você mesmo pode se distrair simulando várias buscas nos *sites* desses catálogos. Pesquise em:

a) <http://id.loc.gov/search> (em inglês);

b) <http://lcweb2.loc.gov/hlas/portugues/hlashome.html>.

RESUMO

O cabeçalho de assunto é um componente de uma LD formado por palavras/expressões que representam o conteúdo dos assuntos dos documentos. Ao conjunto de cabeçalhos de assunto dá-se o nome de lista de cabeçalhos de assunto, cuja proposta é facilitar o processo de indexação/recuperação da informação. Como toda linguagem, ela é composta de vocabulário e gramática. Algumas iniciativas envolvendo o cabeçalho de assunto têm, também, uma sintaxe.

O cabeçalho de assunto tem como função controlar o vocabulário empregado na indexação/recuperação da informação, padronizando-o e, desse modo, aproximar a linguagem usada pelo indexador da linguagem usada pelo usuário em suas buscas.

[27] Disponível em: <https://pixabay.com/pt/menino-negócios-desenhos-animados-1300226>.

As listas de cabeçalhos de assunto são consideradas LD, tendo em vista seu papel na indexação/recuperação da informação. As LD são linguagens documentárias artificiais (construídas), cuja principal característica é ser pré-coordenada. Isso significa que as palavras/expressões que formam os cabeçalhos são deliberadamente conjugadas antes mesmo de o indexador ou o usuário fazerem uso deles.

Em 1876, *Charles Ammi Cutter* fez a primeira tentativa de sistematização dos cabeçalhos usados na *LC*, introduzindo o catálogo-dicionário. Esse catálogo foi uma inovação para a época, pois *Cutter* pretendia que todas as entradas dos cabeçalhos fossem diretas pelos assuntos, sem inversões de palavras nem interposições de classes, seguindo a ordem usada na linguagem natural.

Mas as regras apresentadas para formação e aplicação dos cabeçalhos geraram muitas inconsistências, que persistem até hoje na *Library of Congress Subject Headings* (*LCSH)*.

Historicamente, essa linguagem vem sendo usada há muito tempo pelas bibliotecas e, apesar de suas inconsistências, a *LCSH* continua a ser empregada em várias bibliotecas do mundo, inclusive no Brasil. O sistema de cabeçalhos de assunto vem sendo constantemente criticado, quer pela falta de princípios científicos que validem sua estrutura – o que compromete sua eficácia como ferramenta de auxílio na indexação/recuperação da informação, quer pela maior eficiência de novas tecnologias desenvolvidas para o mesmo fim.

INFORMAÇÕES SOBRE A PRÓXIMA UNIDADE

Na próxima Unidade, você começará a estudar outra linguagem documentária, desenvolvida a partir dos anos 1960 do século XX – os tesauros. Até lá!

PRÉ-REQUISITOS

Seria ótimo se você acessasse a internet e fizesse uma busca pela palavra "tesauro". Você também pode procurar por *thesauri*. A grafia pode variar. Experimente começar pelo *Tesauro de Folclore e Cultura Popular Brasileira*, passeie por ele. Clique, inicialmente, na "parte alfabética". Você vai se divertir com as novidades a respeito de nosso folclore. O *site* é: <http://www.cnfcp.gov.br/tesauro/index_m.htm>.

Sugestão de Leitura

MAC EVAN, Andrew. Working with LCSH: the cost of cooperation and the achievement of access: a perspective from the British Library. In: CONFERÊNCIA GERAL DA IFLA, 64., 1998, Amsterdam. **Proceedings...** Amsterdam: IFLA, 1998.

REFERÊNCIAS

ALMEIDA, Rogerio C.P. de. **Diretrizes para elaboração de um cabeçalho de assuntos**. Porto Alegre: Ed. UFRS, 2011. Disponível em: <rogeriopetrinialmeida.blogspot.com>. Acesso em: 20 ago. 2015.

BRASIL. Senado Federal. **Análise e representação de assunto**: diretrizes para a Rede Virtual de Bibliotecas: Congresso Nacional: RVBI. Brasília: Senado Federal, 2007. Disponível em: <www2.senado.leg.br/>. Acesso em: 23 ago. 2015.

CESARINO, M. A. da N.; PINTO,M.C.M. F. Cabeçalho de assunto como linguagem de indexação. **R. Esc. UFMG**, Belo Horizonte, v.7, n. 2, p. 268-288, set. 1978.

EL-HOSHY, Lynn M. Relationships in Library of Congress Subject Headings. In: BEAN, Carol A.; GREEN, Rebecca (Ed.). **Relationships in the organization of knowledge**. Dordrecht: Kluwer Academic Publishers, 2001. ch.9.

FERRAZ, Wanda. **Relação de assuntos para cabeçalhos de fichas**. 3. ed. Rio de Janeiro: Freitas Bastos, 1977.

FISCHER, Karen S. Critical views of LCSH, 1990-2001: the third bibliographic essay. **Cataloging & Classification Quarterly**, [S.l.], v. 41, n. 1, p. 63-109, 2005.

GOMES, Hagar E.; MARINHO, Marcílio.T. Introdução ao estudo do cabeçalho de assunto. **Biblioteconomia**, **Informação & Tecnologia da Informação**, Rio de Janeiro, 1984.Disponível em: <http://www.conexaorio.com/biti/>. Acesso em: 16 nov. 2014.

KIRTLAND, Monika; COCHRANE, Pauline. Critical views of LCSH - Library of Congress Subject Headings: a bibliographic and bibliometric essay. **Cataloging & Classification Quarterly**, [S.l.], v. 1, n. 2-3, p. 71-94, 1982.

LIBRARY OF CONGRESS. LCSH Cataloguing Distribution Service. **Library of Congress**, Washington, [201-?]. Disponível em: <http://www.loc.gov/cds/products/>. Acesso em: 20 ago. 2015.

LIBRARY OF CONGRESS. **Library of Congress Subject Headings:** monthly List 07. Washington: **Library of Congress**, 2015. Disponível em: <http://classificationweb.net/approved-subjects/1507.html>. Acesso em: 18 ago. 2015.

LIBRARY OF CONGRESS. **Library of Congress Subject Headings (LCSH)**. Washington: Library of Congress, 2014. Disponível em: <http://www.db.dk/bh/lifeboat_ko>. Acesso em: 18 mar. 2010.

LIBRARY OF CONGRESS. Library of Congress Subject Headings (LCSH 37): introduction. **Library of Congress**, Washington, [201-?]. Disponível em: <http://www.loc.gov/aba/publications/FreeLCSH/LCSHintro.pdf>. Acesso em: 25 ago. 2015.

RIO DE JANEIRO (Estado). Poder Judiciário. **Manual para elaboração do vocabulário controlado da Biblioteca do Tribunal de Justiça do Estado do Rio de Janeiro**. Rio de Janeiro: Tribunal de Justiça do Estado do Rio de Janeiro, 2012. Disponível em: <http://www.tjrj.jus.br/documents/1013...>. Acesso em: 22 ago. 2015.

SHAVA, Tsering Wangyal. **Critical analysis of Library of Congress Subject Headings**. New Jersey: Princeton University, 1998. Disponível em: <http://www.princeton.edu/~shavatw/LCSH.html>. Acesso em: 6 ago. 2015.

SVENONIUS, Elaine; MCGARRY, Dorothy. Objectivity in evaluating subject heading assignment. **Cataloging & Classification Quarterly**, [S.l.], v. 16, n. 2, p. 5-40, 1993.

TEIXEIRA, José Carlos Abreu. **Cabeçalhos de assunto:** manual para estudantes. Niterói: Ed. UFF, 1979.

TÔRRES, L. M. C. **Inconsistências de cabeçalhos de assunto:** propostas de sistematização.1992. Dissertação (Mestrado em Ciência da Informação) – UFRJ/ ECO/ IBICT, Rio de Janeiro, 1992.

WELLISCH, Hans H. et al. **Indexing from A to Z**. 2nd ed. New York: HW Wilson, 1995.

YI, Kwan; CHAN, Lois Mai. 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**, [S.l.], v. 61, n. 4, p. 677-687, 2010.

UNIDADE 3

TESAURO

3.1 OBJETIVO GERAL

Apresentar o tesauro como linguagem documentária e instrumento auxiliar do processo de indexação e recuperação da informação, reconhecendo sua natureza, características e tipos, bem como os princípios, normas e técnicas que regem sua elaboração para aplicação prática, além de identificar os mecanismos para sua avaliação.

3.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

a) reconhecer o tesauro como uma linguagem documentária, sua natureza, evolução histórica, características, tipologia e funções;

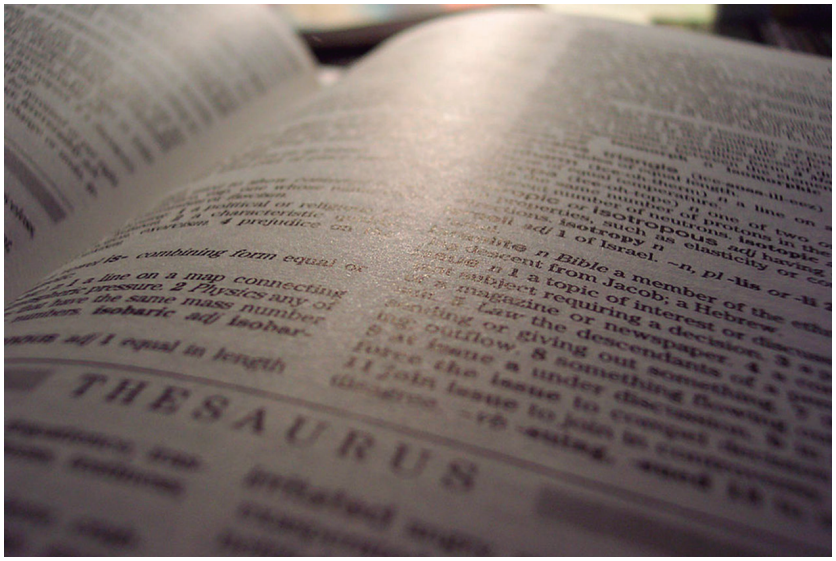
b) sistematizar princípios utilizados na elaboração de tesauros;

c) identificar critérios para avaliação de tesauros;

d) identificar critérios para avaliação de *softwares* para a geração de tesauros.

3.3 VOCÊ JÁ OUVIU FALAR EM TESAURO?

Figura 30 - Tesauro



Fonte: indisponível.

Como você vai constatar no decorrer desta Unidade, nem sempre o que se chamou de tesauro é o que conhecemos hoje por esse nome. O tesauro já foi um simples dicionário com palavras em ordem alfabética; já foi um dicionário organizado em classes nas quais as palavras se associavam pela ideia. Depois, transformou-se em um instrumento destinado à indexação temática de documentos. Enfim, em seu processo evolutivo, o tesauro passou por diversas mudanças. Nesta Unidade você o conhecerá de perto.

3.4 AFINAL, DE QUE SE TRATA?

Atualmente, considerando-se as normas internacionais para a construção de tesauros e os últimos desenvolvimentos teóricos nessa área, podemos definir um **tesauro** da seguinte maneira:

Tesauro

É uma LD dinâmica, estruturada, composta por termos (descritores) que se inter-relacionam pela hierarquia e pelo significado. O tesauro abrange uma área específica do conhecimento e é usado na indexação e na recuperação da informação.

3.4.1 Origens do tesauro: um pouco de história

Muita gente acha que a palavra correta é tesouro. Será mesmo que ela foi criada nesse sentido? Tesauro ou tesouro? Descubra!

Figura 31 - Tesouro ou tesauro?

Fonte: *Pixabay* (2017).[28]

A origem etimológica da palavra "tesauro" remonta à Antiguidade. Ela provém do vocábulo latino *thesaurus,* que, por sua vez, deriva do grego *thesaurós* e foi usada durante vários séculos com o significado de "tesouro ou armazém/repositório de palavras".

Esse termo foi usado nesse sentido pela primeira vez por *Bruno Latini* (1220-1294) para designar uma enciclopédia sistematizada, à qual deu o nome de *Os livros do tesouro.* Em 1572, *Thierry* e *R. Etienne* publicaram seu *Dicionário ou tesouro do idioma italiano*, com arranjo alfabético. Em 1572, *A. Etienne* publicou um dicionário, o *Thesaurus linguae graecae*.

Entre os dicionários publicados até hoje, o *Shorter Oxford Dictionary* registrou, pela primeira vez em 1736, o uso da expressão inglesa *"treasury or store house of knowledge"*, definindo-o como "tesouro ou armazém de conhecimento, similar a um dicionário ou a uma enciclopédia". Já o dicionário *Webster's* define o termo como um livro de palavras ou de informação sobre determinado campo, ou ainda um conjunto de conceitos, especificamente um dicionário de sinônimos.

Se você for bom observador, notará que, com a evolução do tempo, as próprias definições deixam transparecer diferenças que definitivamente contribuiriam para o atual conceito de tesauro: de dicionários gerais de uma língua, passa-se a considerar a informação sobre determinado campo do conhecimento; de palavras, passa-se a considerar também conceitos.

Essas obras eram arranjadas alfabeticamente e, como em todo dicionário alfabético, parte-se da palavra para saber seus vários significados. Tem-se notícia, porém, de dicionários ideológicos, nos quais parte-se do significado para saber qual a melhor palavra a ser empregada para expressar uma ideia.

A literatura da área aponta, como o dicionário ideológico da Antiguidade mais conhecido, o *Onomástikon*, no qual as palavras do idioma

[28] Disponível em: <https://pixabay.com/pt/piratas-telescópio-mapa-do-tesouro-2014558>.

grego (incluindo todos os sinônimos) estão agrupadas em dez classes: deuses, homem (partes do corpo), genealogias, ciência e arte, caça, alimentos, ofícios, leis, organização urbana, utensílios.

Foi com o dicionário ideológico de *Peter Mark Roget*, em Londres, 1850, que o termo "tesauro" se popularizou. Esse dicionário – *o Thesaurus of English Words and Phrases* – é definido pelo próprio autor como uma coleção arranjada não em ordem alfabética, como num dicionário, mas segundo as ideias que as palavras e frases representam.

Multimídia

Usar a palavra imprópria às vezes cria um problema!

Figura 32 - A palavra certa tem seu lugar!

Fonte: *Pixabay* (2012).[29]

"Quero usar uma palavra que signifique 'arranjo'. Mas não é bem essa palavra que eu gostaria de usar..."

Se você recorrer ao *Dicionário Analógico da Língua Portuguesa: ideias afins/thesaurus*, de *F. F. dos Santos Azevedo*, e localizar a palavra "arranjo", você verá que existem muitas outras palavras com esse significado: "preparação", "composição", "arrumação", "ordenação", "combinação", "organização", "esquema", "estrutura","classificação" e muitas outras. Tente! Faça a busca e se encante com a pluralidade de palavras com significados equivalentes em nossa língua!

[29] Disponível em: <https://pixabay.com/pt/pensador-pensamento-pessoa-idéia-28741>.

3.4.2 Dicionários ideológicos: a analogia entre as palavras

Figura 33 - Peter *Mark Roget* criou o dicionário de ideias afins

Fonte: *Wikipédia* (1833).[30]

Os dicionários ideológicos nasceram na segunda metade do século XIX e o médico inglês *Peter Mark Roget* é considerado o iniciador desse tipo de repertório. Esse trabalho de *Roget*, que ele chamou de "tesauro"– *Thesaurus of English Words and Phrases*, 1852 –, foi considerado o primeiro tesauro moderno e levou cerca de 50 anos para ser editado. Nele, as entradas são arranjadas conceitualmente e não alfabeticamente. A busca por palavras dá-se sempre pelo que elas expressam, por seu significado. Por isso, foi a primeira publicação que influenciou os tesauros de nossa época, nos quais a unidade de representação dos assuntos é o conceito. O tesauro de *Roget* nunca deixou de ser editado e é muito usado no mundo todo, principalmente pelas pessoas que se esmeram na arte de escrever.

A obra tem duas características que a distinguem dos demais tesauros elaborados até então: seu objetivo e sua forma.

O objetivo do tesauro de *Roget* é ajudar o usuário a se mover de uma ideia para a palavra que ele pode usar a fim de expressar essa ideia num texto escrito. O subtítulo do dicionário expressa bem seu objetivo: "palavras classificadas e arranjadas para facilitar a expressão das ideias e para ajudar na composição literária".

Quanto à forma, o tesauro agrupa palavras, tipicamente sinônimos, além de termos genéricos, relacionados e antônimos. Como diz o próprio *Roget*, na introdução de seu trabalho, "palavras arranjadas de acordo com as ideias que elas expressam [...] estritamente de acordo com seu significado".

No Brasil, cabe destacar o *Dicionário Analógico da Língua Portuguesa – Ideias Afins/Thesaurus*, editado pela primeira vez em 1950, por *Francisco Ferreira dos Santos Azevedo*, que entendia que as palavras, além de apresentarem sinônimos, também estão relacionadas a outras ou a gru-

[30] Disponível em: <https://en.wikipedia.org/wiki/Peter_Mark_Roget#/media/File:Roget_P_M.jpg>.

pos de outras pelas ideias que representam, ou seja, por seu significado. Trata-se de um dicionário temático, no qual as entradas são organizadas por temas, como "semelhança", "pureza" e "convergência", e não como uma lista alfabética de palavras. Dois tipos de busca são possíveis no dicionário de *Azevedo*, para achar uma palavra específica: buscar o tema ao qual a palavra está relacionada ou recorrer ao índice. Esse *Dicionário Analógico* registra e classifica, em língua portuguesa, as palavras e suas correlações, abstratas ou concretas, tangíveis e intangíveis.

3.4.3 A trajetória do tesauro e as descobertas científicas nas guerras

Ao mesmo tempo que causaram uma enorme destruição no mundo, as duas grandes guerras trouxeram consigo o avanço científico e tecnológico. O homem fez, nesse período, o que não havia feito em 19 séculos.

Esse fato influenciou diretamente a informação. Os relatos feitos pelos cientistas e publicados nos veículos de divulgação científica da época eram relatos de novas descobertas que necessitaram ser "batizadas" com novos nomes para serem identificadas. Uma nova terminologia surge, então.

A literatura da época passou a exigir um novo tipo de representação nos catálogos dos serviços de informação, já que o cabeçalho de assuntos, que era a tecnologia de informação usada até então para a representação dos assuntos, não dava mais conta de representar assuntos tão complexos como os que passaram a surgir no pós-guerra, formados por várias e novas palavras. O que você acha que aconteceu?

Curiosidade

E depois da guerra?

Figura 34 - Invenções pós-Segunda Guerra Mundial

Fonte: *Wikimedia Commons* (2012) e *Wikipédia* (2010).[31]

O século XX traz o fardo das guerras. Mas, de 1914 a 1945, em precisamente 31 anos, a humanidade presenciou também a maior revolução científica jamais vista. É muito antiga a ligação entre as ati-

[31] Primeira imagem: Televisão. Disponível em: <https://commons.wikimedia.org/w/index.php?curid=22419876>; segunda imagem: M&M'S. Disponível em: <https://en.wikipedia.org/wiki/M%26M%27s#/media/File:Plain-M%26Ms-Pile.jpg>.

vidades bélicas e a ciência, ou, mais exatamente, a tecnologia derivada das guerras. Já na Antiguidade, as propriedades elásticas dos materiais e a estabilização de projéteis foram largamente empregadas na construção de arcos e flechas, e, posteriormente na fabricação de máquinas de guerra capazes de arremessar pedras a distâncias consideráveis. Você pode saber mais sobre esse assunto no *link*: <http://www.comciencia.br/reportagens/guerra/guerra08.htm>.

3.5 A BUSCA POR NOVAS FERRAMENTAS PARA A RECUPERAÇÃO DA INFORMAÇÃO

Figura 35 - Buscando e recuperando informações

Fonte: *Pixaba*y (2014).[32]

Se você está lembrado, na Unidade anterior falamos das características da recuperação por meio de cabeçalhos de assunto, sendo uma delas a linearidade. Vimos que essa característica restringe as buscas à primeira

[32] Disponível em: <https://pixabay.com/pt/bússola-definição-palavra-390903>.

palavra do cabeçalho de assunto, fazendo com que outras palavras que compõem os cabeçalhos não possam ser encontradas. Vimos, ainda, que isso limita a possibilidade de o usuário "montar" sua própria estratégia de busca, obtendo uma recuperação mais precisa.

Foi mostrado que as descobertas do pós-guerra dificultaram a representação dos assuntos nos catálogos dos serviços de informação, tendo a tecnologia de informação da época se tornado insuficiente. Em sua opinião, o que acontece quando uma tecnologia deixa de ser funcional? Pense...! Parece que a resposta é óbvia: novas tecnologias são descobertas. Foi isso que você imaginou? E foi exatamente o que aconteceu com os sistemas de representação dos assuntos de um documento.

Não existe consenso na literatura de Biblioteconomia e Ciência da Informação a respeito do surgimento do primeiro tesauro de recuperação da informação, mas, definitivamente, o tesauro surge como nova tecnologia em relação às listas de cabeçalhos de assunto, para representação temática da informação. As inovações introduzidas por essa nova ferramenta serão vistas no decorrer desta Unidade.

Há fontes que indicam que, durante os anos 1950, o termo "*thesaurus*" foi utilizado pela primeira vez por *Peter Luhn*, da *International Business Machines* (*IBM*), nos EUA. *Luhn* achou que, no sistema por ele desenvolvido, uma listagem alfabética não seria suficiente para localizar a palavra mais adequada à recuperação e que alguma relação entre as palavras deveria ser estabelecida.

Nomeando seu sistema de *"thesaurus"*, um novo tipo de linguagem documentária foi lançado. Dessa vez, a facilitação da recuperação da informação passou a ser uma meta importante a ser atingida. Surgia o tesauro de recuperação da informação.

A influência de *Roget* sobre *Luhn* foi clara. *Roget* preocupava-se em ajudar as pessoas a encontrar a palavra mais adequada para uma composição literária; a preocupação de *Luhn* também era a localização da palavra mais adequada, só que para a recuperação de uma informação desejada. O foco de ambos era a recuperação da informação mais precisa.

Embora *Luhn* tenha nomeado seu sistema de "*thesaurus*", *Lancaster* (1986) declara que o primeiro tesauro destinado ao controle do vocabulário na recuperação da informação parece ter sido desenvolvido pelo Centro de Informação de Engenharia da empresa norte-americana *Du Pont de Nemours*, por volta de 1959.

Em 1951, *Mortimer Taube* introduziu um novo sistema que tinha por característica principal a representação de assuntos por meio da utilização de uma só palavra, daí seu nome: Sistema Unitermo. Esse sistema possibilitava juntar vários assuntos na hora da busca e, por isso, foi chamado, inicialmente, de sistema coordenado, e, mais tarde, pós-coordenado. Porém, verificou-se que, às vezes, mais de uma palavra eram necessárias para que o assunto fosse representado em sua plenitude.

O Sistema Unitermo não durou muito tempo. Seu primeiro recuo foi a aceitação de "termos ligados", ou seja, termos que apareciam sempre juntos em determinado vocabulário, por exemplo, "válvula átrio ventricular esquerda." Isso descaracterizou o Unitermo, que passou a permitir o uso de mais de uma palavra para representar os assuntos. Vários sistemas que adotaram o Unitermo começaram a usar, cada vez mais, esse tipo de pré-coordenação – com termos ligados – e o princípio da indexação ficou sendo conhecido como coordenação de conceitos.

3.6 TESAURO DOCUMENTÁRIO

A grande fragilidade do Unitermo foi que ele não possuía controle algum do vocabulário, principalmente nos sistemas que ficaram fiéis a seu princípio e só usavam assuntos com uma só palavra, exigindo manipulação de grande número de termos. Percebeu-se que, em qualquer sistema, fosse ele pré ou pós-coordenado, melhores resultados poderiam ser obtidos na recuperação da informação se tivessem um vocabulário controlado.

Isso deu espaço para o surgimento do tesauro documentário, nova proposta que surgiu, com o objetivo de controlar rigorosamente o vocabulário, entre outras finalidades.

De acordo com Campos (2001, cap.4), os anos 1960 testemunharam o aparecimento de vários tesauros, agora já aperfeiçoados. Os primeiros tesauros largamente divulgados foram os seguintes:

a) *Thesaurus of ASTIA Descriptors* (do *Ministério da Defesa dos Estados Unidos*, anteriormente *Armed Services Technical Information Agency*, em 1960);

b) *Chemical Engineering Thesaurus* (do *American Institute of Chemical Engineers,* em 1961), baseado no tesauro da *Du Pont* e publicado em 1964 pelo *Engineers Joint Council* (*EJC*) para cobrir toda a área de Engenharia;

c) *Thesaurus of Engineering and Scientific Terms* (TEST) – iniciativa no âmbito do Projeto Lex, do *Department of Defense* dos Estados Unidos, em colaboração com o *Engineers Joint Council* (EJC), publicado em 1967, contendo mais de 20 mil termos. O Projeto Lex teve dois objetivos:

   - reunir os dois tesauros da área de Engenharia, criando um único instrumento para as duas instituições;
   - estabelecer princípios comuns de construção e uso, o que deu origem a um manual de construção de tesauros. Esse manual foi recomendado como fonte para a construção de tesauros, tendo sido a base para as diretrizes e normas produzidas pela *American National Standardization* (*ANSI*, 1981) e pela *Organização das Nações Unidas para a Educação, a Ciência e a Cultura* (*UNESCO*) (1973). As normas da *UNESCO* originaram as normas nacionais e internacionais para a construção de tesauros.

3.6.1 O desenvolvimento dos tesauros nos continentes

Vocês se lembram de que, no início deste Curso, falamos que os catálogos usados na Europa eram os alfabéticos-por-classes, nos quais as entradas dos termos eram dadas pelas classes de objetos, e não pelo objeto específico que se queria procurar? E que, nos EUA, o catálogo alfabético tinha por objetivo indicar entradas diretas pelo termo que o usuário queria pesquisar?

Pois bem, essas realidades influenciaram diretamente os tesauros. E, na prática, o que se observou foi que, na América do Norte, principal-

mente nos EUA, os tesauros foram inspirados no sistema Unitermo, que usava uma só palavra retirada da literatura, sem critério algum e sem controle desse vocabulário. Na Europa, porém, os tesauros sofreram influência da Teoria da Classificação Facetada, de *Ranganathan*, como veremos a seguir (ver também Unidade 1, seção 1.4), desenvolvidos com a inclusão de uma parte sistemática. Atualmente, a elaboração de tesauros é regida por normas internacionais e a tendência é que eles sejam desenvolvidos observando os princípios não só da Classificação Facetada, mas de outras teorias, como será visto a seguir.

3.6.2 Teorias que orientam a construção de tesauros

Ainda que várias iniciativas atuais de linguagens para a representação temática da informação sejam nomeadas de tesauros, nem todas acompanham as orientações das três teorias indicadas para a construção de tesauro com base em conceito, a mais nova tendência.[33]

Quadro 5 - Teorias para a construção de tesauro com base em conceito

Teoria da Classificação Facetada (RANGANATHAN,1967)	Fornece princípios sobre a alocação do termo numa categoria, juntamente com outros de significado próximo, explicitando relações de equivalência (sinônimos e quase sinônimos, ex.: pipa = papagaio de papel = pandorga), de gênero/espécie (ex.: flauta e flauta transversa) e outras relações associadas pelo significado (ex.: pintura e tinta).
Teoria do Conceito (DAHLBERG, 1978)	Define "conceito" no âmbito da Ciência da Informação, fornecendo princípios para a fixação do conteúdo do conceito e sua representação por meio de um termo. O objetivo é "delimitar o uso do termo em um discurso", assim como criar princípios para possibilitar a colocação do termo em determinada área de assunto.
Teoria Geral da Terminologia (WÜSTER,1981)	Estabelece princípios para a nomeação dos conceitos de forma mais adequada (fornecendo, desse modo, bases mais seguras para seu relacionamento com outros termos), assim como para a criação de novos conceitos.

Fonte: produção do próprio autor (2017).

Até o final dos anos 1950 e início dos anos 1960, os tesauros eram apresentados somente em ordem alfabética, o que impossibilitava uma boa representação das relações entre os termos. Se você quiser relembrar a questão da fragilidade da ordenação alfabética em sistemas de recuperação da informação, basta voltar à Unidade 2, seção 2.5.7, e rever o item que fala sobre as características das listas de cabeçalhos de assunto.

A partir da metade dos anos 1960, sentiu-se a necessidade de usar princípios classificatórios nos sistemas de representação temática da informação. Então, os tesauros evoluíram e passaram a usar uma técnica mais aprimorada de construção, baseada na já mencionada Classificação Facetada, que possibilita o posicionamento do conceito num sistema de conceitos, revelando uma área de assunto.

[33] As informações apresentadas nesta Unidade sobre a natureza, a estrutura ou a construção de tesauros consideraram as orientações das Teorias da Classificação Facetada, Geral da Terminologia e do Conceito.

Dentro dessa nova orientação, cabe citar o *Thesaurofacet: A Thesaurus and Faceted Classification for Engineering and Related Subjects,* nome dado ao tesauro e classificação facetada para Engenharia e assuntos correlatos, de autoria de *Aitchison* (1970). Esse tesauro compõe-se de duas partes que se complementam: a Tabela de Classificação Facetada e o Tesauro Alfabético.

O *Thesaurofacet* foi apresentado, pela própria *Aitchison*, como um "novo conceito em esquemas de recuperação de assuntos". A pesquisadora percebeu, depois, que sua proposta não era tão nova assim, já que outros tesauros vinham adotando a classificação facetada como base para sua construção.

No início desta Unidade, definimos tesauro como o concebemos hoje. No entanto, as primeiras definições dadas, ainda nos anos 1960, referiam-se a esse termo como uma lista alfabética de palavras, onde cada uma delas era seguida de outras que a ela se relacionavam.

No início dos anos 1970, a *UNESCO* definiu o **tesauro** segundo dois aspectos, como pode ser visto a seguir:

a) **tesauro, segundo a estrutura:** "vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento."(UNESCO, 1973, p. 6);

b) **tesauro, segundo a função:**

> Dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema (linguagem de documentação, linguagem de informação) mais restrita. (UNESCO, 1973, p. 6).

A definição dada pela *UNESCO* mostra alguns avanços, especificando os tipos de relação que o tesauro tem (semânticas e hierárquicas). Porém, mesmo que essa definição fale em "termos", nos primeiros tesauros, o que representava os assuntos tratados nos documentos era a palavra e, muitas vezes, uma só palavra. Atualmente, nos tesauros documentários é que a unidade de representação deve ser o conceito, que pode ser formado por uma ou várias palavras, como: "automóvel", "pé-de-moleque" (termos formados por uma só palavra), "infecção oportunista" (termo formado por duas palavras), "machado de pedra semilunar" (termo formado por três palavras).

3.6.3 Componentes de um tesauro documentário

Como toda LD, o tesauro é composto de:

a) vocabulário: reúne um conjunto de conceitos de determinada área do conhecimento;

b) regras para a utilização da linguagem.

Além desses dois componentes, há tesauros que ainda incluem:

a) notas de escopo ou definições dos termos, visando à aplicação mais precisa do conceito;

b) indicação da categoria à qual o conceito pertence;

c) notas de aplicação, para orientar o uso do termo na hora da indexação.

3.6.4 E as características de um tesauro, quais são?

Dentre suas características, pode-se destacar que:

a) atua como instrumento de controle terminológico na indexação: controla sinônimos, homógrafos, formas de entrada do termo, além do relacionamento entre os termos;

b) cobre uma área específica do conhecimento ou áreas afins de um assunto;

c) é uma linguagem especializada com fins documentários;

d) constitui uma base de conhecimento sobre a terminologia mais apropriada de uma área (sua estrutura possibilita o mapeamento da área coberta);

e) não comporta exceções para a aplicação de suas regras, que são explícitas;

f) trabalha com o conceito para representar os assuntos, e não com a palavra, ou seja, em nível conceitual, e não verbal;

g) apresenta estrutura sistemática, na qual os termos são organizados em relações de dependência e definidos uns em relação aos outros, sendo, portanto, impossível compreendê-los de forma isolada.

   Ex.: DIADEMA ROTIFORME

   Definição: Adorno corporal indígena em que as penas ornamentais acompanham a forma arqueada do suporte

   Usado por: Diadema de arco irradiante

   Termo genérico

      Diadema

   Termo específico

      Diadema rotiforme alçado

      Diadema rotiforme occipital

      Diadema rotiforme vertical

h) sintetiza as informações convertidas da linguagem natural dos textos para a linguagem artificial, por meio dos termos. Ex.: um documento que trate da "fragilidade da economia global e mecanismos para sua recuperação" poderá ter esse assunto sintetizado no termo "economia mundial";

i) é um sistema essencialmente pós-coordenado, ou seja, apresenta o termo de forma isolada para ser conjugado a outros no momento da busca. Ex.: o termo "inflação" pode ser conjugado com "Brasil" e, ainda, com "século XX", quando se quer procurar documentos sobre a "inflação brasileira no século XX";

j) por ser uma linguagem construída, pode ser considerada uma metalinguagem, ou seja, uma linguagem que pressupõe a existência de outra;

k) é uma linguagem dinâmica, que, como qualquer outra, reflete a prática social e, por isso, necessita de constantes atualizações;

l) agrupa termos relacionados pelo significado (ver exemplo da alínea g);

m) possibilita buscas genéricas, compreensivas, já que, a partir de um termo, pode-se chegar a todos os outros a ele relacionados, tornando a busca mais completa;

n) tem, minimamente, duas formas de apresentação: alfabética e sistemática.

3.6.5 Quais as funções de um tesauro?

O tesauro é um importante componente de um sistema de recuperação da informação e, como tal, deve cumprir prioritariamente as seguintes funções:

a) representar adequadamente e de forma sintética uma área do conhecimento, reunindo seus conceitos mais importantes;

b) controlar o vocabulário dessa área, de maneira que haja perfeita coincidência entre conceitos e termos, formando entre eles uma relação unívoca (cada conceito refere-se a um só termo e cada termo, a um só conceito);

c) auxiliar o indexador na busca do termo, ou termos, mais apropriado para expressar o assunto abordado no documento;

d) auxiliar na busca de informações, servindo de intermediário entre os documentos e os usuários, e permitindo que o usuário amplie ou restrinja sua busca por meio da navegação na estrutura do tesauro.

3.6.6 Que tipos de tesauro existem?

Os tesauros podem ser vistos sob vários aspectos. Dependendo deles, tais instrumentos poderão ser classificados em várias categorias:

a) de acordo com a língua do texto: monolíngue (texto em um só idioma) ou multilíngue [o tesauro original em uma língua e versões para outra(s)];

b) de acordo com o(s) assunto(s) coberto(s) pelo tesauro: multidisciplinar (voltado para um assunto e para áreas do conhecimento relativas a ele) ou especializado (voltado para uma área específica do conhecimento);

c) de acordo com o nível de especificidade: macrotesauro (descritores com conceitos em nível genérico) ou microtesauro (descritores com maior nível de especificidade).

3.6.7 Atividade

Imagine que você entra numa biblioteca de Matemática a fim de buscar informações para uma aula de Geometria que tem de apresentar. A bibliotecária lhe encaminha ao catálogo da biblioteca, que usa o sistema do conhecido cabeçalho de assunto, pois, para organizar seus assuntos, a biblioteca emprega como base a *Lista de Cabeçalhos de Assunto da Biblioteca do Congresso dos Estados Unidos*. Iniciando sua busca, você procura pelo termo "geometria" e encontra a seguinte informação na ficha do catálogo de assuntos:

GEOMETRIA

V. também Desenho geométrico: Desenho linear; Trigonometria; e cabeçalhos que comecem pela palavra 'geometria'.

Rem. de Geometria plana; Geometria sólida

Não muito satisfeito, no dia seguinte você parte para outra biblioteca de Matemática. Lá, a bibliotecária lhe informa sobre a existência de um tesauro *on-line* para a recuperação da informação. Você acessa o tesauro, lê seu resumo na introdução e as instruções de como usá-lo. Busca novamente pelo termo "geometria" e encontra a seguinte estrutura:

GEOMETRIA

Definição: Ramo da matemática cujo objeto é o estudo do espaço e das figuras que podem ocupá-lo; trata de medidas, prioridades e relações entre ponto, retas, ângulos, superfícies e sólidos, havendo vários tipos: geometria analítica, dedutiva, descritiva, do espaço, experimental, plana.

Termo do todo: MATEMÁTICA

Termo específico: GEOMETRIA ANALÍTICA
GEOMETRIA DEDUTIVA
GEOMETRIA DESCRITIVA
GEOMETRIA DO ESPAÇO
GEOMETRIA EXPERIMENTAL
GEOMETRIA PLANA

Termo associado: MEDIÇÃO
TEOREMA DE PITÁGORAS
TOPOLOGIA (MATEMÁTICA)

O que foi diferente para você nas duas buscas? Por que foi mais difícil achar o que precisava em uma biblioteca do que em outra?

Pense em, pelo menos, três características que diferenciem a busca num sistema de cabeçalhos de assunto e num tesauro. Se precisar, volte aos textos já estudados. Anote suas impressões a seguir:

Cabeçalho de assunto

1. _____
_____

2. _____
_____

3. _____
_____

Tesauro

1. _____

_____

2. _____

_____

3. _____

_____

Resposta comentada

Você deve ter ficado um pouco perdido quando fez a busca no sistema de cabeçalhos de assunto. Você queria saber os tipos de geometria existentes, talvez quisesse pesquisar sobre um deles. A orientação dada era procurar os cabeçalhos que começassem pela palavra "geometria", mas você não se lembrava de todos os tipos. Também não conseguiu saber o que queria dizer a indicação **"Rem. de"** (não havia explicação no catálogo sobre os símbolos usados). Além do mais, você também estava interessado em aspectos históricos da "geometria". E o cabeçalho de assunto não lhe permitia juntar um assunto ao outro; você teria que consultar o assunto "história" para saber se teria algum documento sobre a "história da geometria".

Essa é apenas uma das características que distinguem o cabeçalho de assunto do tesauro. Enquanto o tesauro permite que você junte assuntos ou aspectos de um assunto, propiciando a busca pós-coordenada (na hora da busca), o cabeçalho de assunto já indica o cabeçalho que terá que ser usado, muitas vezes de forma pré-coordenada, ex.: Geometria – Aspectos históricos, com direito a acessar esse assunto somente pelo termo de entrada, que é "geometria". Você terá que percorrer todos os assuntos iniciados por "geometria" para saber se o que lhe interessa existe naquela coleção.

Você também não deve ter ficado satisfeito com a forma de apresentação dos termos no cabeçalho de assunto, que dá somente a indicação de ver (**V.**) e **Rem. de**, a primeira para consultar outros termos que possam ser de interesse e a segunda para mostrar os termos que não são usados no vocabulário (no exemplo dado, são "geometria plana" e "geometria sólida").

O tesauro, por sua vez, indica, de forma estruturada, as relações entre os termos, explicitando cada uma delas (de equivalência, gênero/espécie, de todo/parte, associações por significado). Graças a essa estrutura, você decidiu planejar sua aula sobre "geometria do espaço", que era exatamente o assunto que lhe interessava, e ainda resolveu começá-la pelo "teorema de Pitágoras", dois dos assuntos indicados pelo tesauro.

Se você se lembrou de outras diferenças entre os dois instrumentos, talvez tenham sido algumas dessas: no cabeçalho de assunto, os assuntos são representados por palavras; no tesauro documentário, por conceitos. Cabeçalhos de assunto são dispostos alfabe-

ticamente; tesauros pretendem ter duas formas de apresentação: alfabética e sistemática; às vezes, até planigráfica. Cabeçalhos de assunto prestam-se a cobrir o universo do conhecimento; tesauros são linguagens de especialidade, cobrem determinada área do conhecimento ou áreas afins de um mesmo assunto. Existem outras diferenças, é só voltar ao texto e relembrá-las. Sucesso!

3.7 ELABORAÇÃO DE UM TESAURO DOCUMENTÁRIO: BOTANDO A MÃO NA MASSA!

Nesta seção, vamos falar dos princípios que regem a elaboração de um tesauro documentário. Os mesmos princípios também são válidos para a construção de outras ferramentas que tenham como objetivo a recuperação da informação.

Multimídia

Navegar é preciso!

Figura 36 - Navegando em rede

Fonte: *Pixabay*.[34]

[34] Pessoa navegando na internet. Disponível em: <https://pixabay.com/photos/student-woman-startup-business-849821/>.

Quer aprender de uma forma super *light*? Então consulte o *site* <www.conexaorio.com/biti/tesauro/> e navegue pelo texto, que contém as instruções práticas e necessárias, apresentadas detalhadamente, sobre a elaboração de um tesauro. Você verá que é fácil aprender quando o texto é feito com *hiperlinks*, permitindo a navegação de maneira leve e divertida.

É comum encontrar tesauros construídos de várias maneiras. Porém, vimos anteriormente que o *Thesaurofacet*, organizado por *Jean Aitchison*, formalizou a linha teórico-metodológica para a elaboração de tesauros – a metodologia facetada. A partir de então, outros tesauros foram construídos dentro da mesma perspectiva, tornando-os diferentes dos demais.

Essa metodologia provou-se mais adequada, principalmente por ser flexível e adaptável a mudanças, características fundamentais a um instrumento que se destina ao controle terminológico e, por isso, necessita acompanhar a nova linguagem usada para expressar as inovações técnico-científicas.

Por essa razão, este Curso tomará por referência o método facetado, visando à prática da construção de tesauros.

3.7.1 Normas para a elaboração de tesauros

A partir de 1970, a construção de tesauros passou a ser regida por normas nacionais e internacionais. Veja algumas das normas que foram editadas desde aquela época:

a) UNITED NATIONS FOR EDUCATION, SCIENCE AND CULTURE ORGANIZATION. **Guidelines for the establishment and development of monolingual thesauri**. Paris: UNESCO, 1973;

b) INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2788:** guidelines for the establishment and development of monolingual thesauri. Suíça: ISO, 1974. Revista em 1986;

c) INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 5964:** guidelines for the establishment and development of multilingual thesauri. Suíça: ISO, 1985;

d) AMERICAN NATIONAL STANDARDS INSTITUTE. **ANSI Z39.19:1980:** American National Standard Guidelines for Thesaurus Structure, Construction, and Use. Bethesda: Md. NISO, 1980. 20 p.;

e) AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19:** guidelines for the construction, format and management of monolingual controlled vocabularies. Bethesda: Md. NISO, 2005. 176 p.

A norma mais recente com relação ao assunto foi divulgada em 2011 pela *International Standard Organization (ISO),* como pode ser conferido abaixo:

a) INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval. Suíça: ISO, 2011;

b) INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-2:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 2: Interoperability with other vocabularies. Suíça: ISO, 2013.

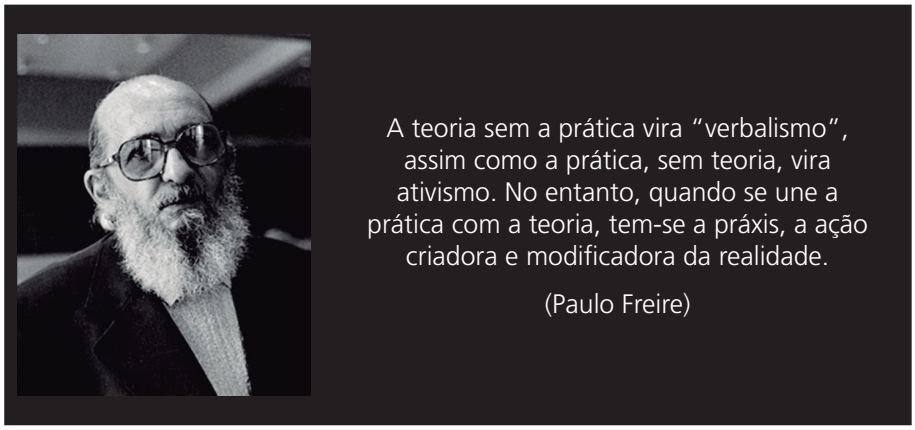
No Quadro 6, sistematizamos a descrição de alguns tópicos citados na norma *ISO 25964*, partes 1 e 2 – aplicável a vocabulários utilizados para recuperar informações sobre todos os tipos de recursos de informação (tesauros e outras linguagens):

Quadro 6 - *Norma ISO 25964*

- É aplicável a tesauros mono e multilíngues;
- sugere recomendações para o desenvolvimento e manutenção de dicionários destinados a aplicações em busca de informação;
- é aplicável a vocabulários utilizados para recuperar informações sobre todos os tipos de recursos de informação, independentemente do meio utilizado (texto, som, imagem fixa ou em movimento, objeto ou multimídia física), incluindo bases de conhecimento e portais, bases de dados bibliográficos, textos, museu ou multimídia, coleções, e os itens dentro deles.

Fonte: adaptado da norma *ISO 25964*.

Figura 37 - Teoria e prática completam-se



Fonte: *Nepo* (20--?) e *Wikimedia Commons* (1977).[35]

3.7.2 Teorias que fundamentam a elaboração de um tesauro documentário

Falamos anteriormente que a nova tendência dos tesauros é observar e incorporar os princípios de três teorias: Teoria da Classificação Facetada (RANGANATHAN, 1967), Teoria do Conceito (DAHLBERG, 1978) e Teoria Geral da Terminologia (WÜSTER, 1981). Cada uma delas deu sua contribuição para que o tesauro, construído a partir dessa base teórica, evoluísse e se tornasse um instrumento confiável para auxílio na indexação e recupera-

[35] Primeira imagem: Retrato de Paulo Freire. Disponível em: <https://commons.wikimedia.org/wiki/File:Paulo_Freire.jpg>; segunda imagem: Frase de Paulo Freire. Disponível em: <www.nepo.com.br>.

ção da informação. Vamos ver um pouco de cada um dos métodos básicos dessas teorias, utilizados para a elaboração de tesauros documentários.

3.7.2.1 *Teoria da Classificação Facetada*

Todo assunto pode ser dividido de acordo com a maneira pela qual nós o "enxergamos". Dependendo de como o conhecimento esteja organizado, a divisão resultará em categorias, que, se subdivididas, resultarão em outras, e assim por diante.

Ranganathan, com sua Teoria da Classificação Facetada, desenvolvida nos anos 1930, procurou organizar ideias, conceitos. Ele baseou sua teoria no postulado das Categorias Fundamentais, que poderia ser interpretado assim: todas as coisas da natureza podem ser enquadradas em categorias abrangentes, capazes de acomodar todos os conceitos pertinentes a determinado assunto.

Nas classificações da Antiguidade, *Ranganathan* (1967) já encontrou as categorias coisa, propriedade e ação, às quais acrescentou outras duas: espaço e tempo. Assim, ele estabeleceu as seguintes categorias: **p**ersonalidade, **m**atéria, **e**nergia, e**s**paço, **t**empo (conhecidas pela sigla PMEST).

Com base na lógica, *Ranganathan* (1967) estabelece princípios e postulados para agrupamento dos conceitos, formando **cadeias** e **renques**. A seguir, vamos ver a definição desses conceitos:

Atenção

Cadeia

Série vertical de conceitos: agrupamento de conceitos em hierarquias, onde existe um conceito genérico ao qual vários outros, com características semelhantes (da mesma espécie), são subordinados.

Ex.:

Instrumento

Instrumento musical

Instrumento de sopro

Flauta

Renque

Série horizontal de conceitos. É constituído de conceitos subordinados a um mesmo conceito, ou seja, são conceitos coordenados entre si, "irmãos". Por exemplo, "política comercial", "política econômica" e "política fiscal" são todos termos subordinados a "política". Logo, são do mesmo nível de coordenação, formando um renque.

O tesauro, já sabemos, é um instrumento em que o conhecimento é estruturado em **categorias** e **facetas**, e sua estruturação parte dessas duas noções, que podem ser assim definidas:

Atenção

Categoria

Conjunto de seres (pessoas, coisas, fatos) que possuem características comuns e podem ser abrangidas ou referidas por um conceito ou concepção genérica. Ex.: produto vegetal, artefato indígena, meio de transporte. Por vezes, os termos "categoria" e "classe" são usados indistintamente, por serem equivalentes.

Faceta

Diferentes aspectos pelos quais uma categoria pode ser vista, se revela. Ex.: a categoria "artefato indígena" poderia ter o seguinte desdobramento, de acordo com o aspecto a ser observado, podendo ser subdividida em várias facetas: "adorno corporal", "equipamento mágico-ritual", "equipamento para trabalho", "equipamento para caça/pesca/ataque/defesa", "equipamento lú- dico infantil", "equipamento doméstico". Resumindo, facetas nada mais são do que subcategorias

Cada área de assunto pode possuir tantas categorias quantas forem necessárias para acomodar os assuntos que nela aparecerão para ser agrupados. Embora seus nomes possam variar de acordo com a área de assunto estudada, essas categorias sempre corresponderão às categorias fundamentais de *Ranganathan* (1967), depois retrabalhadas e ampliadas pelo *Classification Research Group (CRG)*, como veremos mais adiante. Verifique isso nos exemplos a seguir, que apresenta algumas áreas de assunto e respecticas categorias que se encaixariam perfeitamente nas categorias fundamentais de *Ranganathan* (citadas entre parênteses):

a) na Engenharia Mecânica: máquinas operatrizes (**p**ersonalidade); seus componentes (**p**artes); esmerilhamento, polimento, frezamento (**p**rocessos); profissionais envolvidos (agentes/**p**ersonalidade) e assim por diante;

b) na Economia: teorias, política econômica (instrumentos/agentes/**p**ersonalidade); planejamento (operações/processo/**e**nergia) e assim por diante;

c) na Construção Civil: casas, edifícios, teatros, praças (**p**ersonalidade); técnicas de edificação (técnicas/**e**nergia); cimento, pedra, cerâmica, vidro (materiais/**m**atéria); profissionais envolvidos (agentes/**p**ersonalidade) e assim por diante;

d) na Biblioteconomia: instituições e organizações, documentos, usuários, suporte documental (**p**ersonalidade); tratamento documentário, recuperação da informação, aquisição (**p**rocessos); linguagens documentárias; profissionais envolvidos (agentes dos processos/**p**ersonalidades) e assim por diante (CAMPOS; GOMES; MOTTA, 2004).

A Teoria da Classificação Facetada foi desenvolvida nos anos 1930 para organização de material bibliográfico. A novidade introduzida por *Ranganathan* na criação da tabela de sua *Colon Classification* (também chamada de Classificação dos Dois Pontos) representou um rompimento com o modo como os esquemas de classificação bibliográfica elaboravam suas tabelas – descrevendo os assuntos já conhecidos conforme dispostos nas disciplinas e nos grandes tratados. Tais classificações não previam espaço para novos assuntos que surgissem com a expansão do conhecimento.

O esquema de classificação produzido segundo o Método de Faceta é importantíssimo, pois permite mapear o assunto do jeito que ele se encontra estruturado na literatura, seja ele já estabelecido ou novo. Esse método resolve o problema de onde enquadrar determinado assunto. Existe um assunto? Então o método de faceta vai acomodá-lo.

Como visto anteriormente, *Ranganathan* procurou sistematizar conceitos (ideias). Tomou por base, então, o postulado das categorias fundamentais e facetas. Com base na lógica, ele estabeleceu princípios e postulados para agrupar os conceitos em hierarquias – formando cadeias – e, dentro delas, para agrupar os conceitos de mesmo nível – formando renques.

O que *Ranganathan* fez, nada mais foi do que aplicar o processo de categorização a conceitos. Vimos, antes, que categoria é o conjunto de pessoas, coisas ou fatos mais geral dentro de uma área de assunto. O interessante é que cada área de assunto pode ser vista sob um ângulo diferente, de acordo com o interesse de quem a está estudando.

Por exemplo, se falo de "informática", posso querer tratar esse assunto do ponto de vista das tecnologias desenvolvidas, mas também posso querer falar do ponto de vista de suas consequências sociais, ou, ainda, dos aspectos morais envolvidos. Isso seriam manifestações de uma categoria ou facetas. Verifique no exemplo do Quadro 7, a seguir:

Quadro 7 - Categorização do assunto "artesanato"

Assunto: ARTESANATO

CATEGORIAS E RESPECTIVAS FACETAS:

- Técnica artesanal (bordado, cerâmica, colagem, xilogravura, encadernação e assim por diante)
- Matéria-prima (barro, cola, couro, gesso, latão, linha, madeira e assim por diante)
- Equipamento (arquinho, balata, gravador, giradeira, pincel, raspadeira e assim por diante)
- Agente (barrista, encadernador, escultor, louceira, marceneiro, tecelão e assim por diante)
- Produto (brinco, broche, colar, tapete, mobiliário, louça, cesta, vatapi e assim por diante).

Fonte: produção do próprio autor (2017).

Ranganathan deu o primeiro passo na modernização do processo classificatório com as categorias fundamentais. Partindo de seus estudos, o *CRG*, da Inglaterra, desdobrou e expandiu as categorias fundamentais e identificou outras tantas, capazes de facilitar a análise de qualquer área de assunto

ou documento, possibilitando seu mapeamento mais perfeito. A seguir, no Quadro 8, você pode conferir as categorias estabelecidas pelo *CRG*:

Quadro 8 - Categorias do *Classification Research Group*

Coisas, substâncias, entidades
que ocorrem naturalmente
produtos
instrumentos
constructos mentais
Suas partes
constituintes
órgãos
Sistemas de coisas
Atributos de coisas
qualidades, propriedades, incluindo
estruturas
medidas
processos, comportamentos
Objeto da ação (paciente)
Relações entre coisas, interações
efeitos
reações
Operações sobre coisas
experimentos, ensaios
operações mentais
Propriedades de atributos, relações e operações
Lugar, condição
Tempo

Fonte: *Campos*, *Gomes* e *Mottas* (2004).

Lembre-se de que os nomes das categorias fundamentais e das facetas podem variar de acordo com a área de assuntoassunto, cabendo ao profissional que está desenvolvendo a classificação nomeá-las adequadamente. Por exemplo, na categoria "coisa" do *CRG* ("personalidade", para *Ranganathan*), certamente caberá uma faceta chamada "inseticida"; na categoria "partes das coisas", poderá ter uma faceta chamada "componente periférico"; na categoria "ação/processo", deverá haver uma faceta chamada "instrumento musical", e assim por diante.

3.7.2.2 *Teoria Geral da Terminologia (TGT)*

Também nos anos 1930, *Wüster*, engenheiro austríaco, desenvolveu a Teoria Geral da Terminologia. Com base na lógica, ele estabeleceu princípios voltados para a identificação do conceito e para a organização de sistemas de conceitos, que levariam à definição de conceito e, por meio desta, o sistema ficaria evidente.

Este trabalho, porém, não forneceu elementos para orientar a reunião dos conceitos de mesma natureza em ordem hierárquica, a ponto de revelar um sistema. Tal operação poderia ser realizada com a aplicação do Método de Faceta, disponibilizado pela Teoria da Classificação Facetada. Neste ponto é que as duas teorias – a da Classificação Facetada e a da Terminologia – convergem, uma complementando a outra, ambas com um objetivo comum: trabalhar com os conceitos e com as relações existentes entre eles, disponibilizando as bases para a identificação e sistematização dos conceitos/termos de uma área de assunto.

3.7.2.3 *Teoria do Conceito*

Para completar a base de sustentação do tesauro documentário, só falta abordar a Teoria do Conceito. Mas não poderíamos falar dessa teoria sem falar do próprio conceito.

Figura 38 - Qual o conceito de "manga"?

Fonte: *Pixabay* (2016, 2017, 2017).[36]

Para a Filosofia, o conceito é uma abstração. Cada um pode imaginar o que quiser quando é perguntado: "o que você entende por...?" Só para dar um exemplo, o termo "manga" pode ter vários significados, conforme você vê na Figura 38. Várias pessoas diferentes imaginarão coisas diferentes ao ouvir a palavra "manga": pode ser uma fruta, pode ser uma parte de peça de vestuário, pode ser um componente de uma lamparina, pode ser muita coisa! Não quer consultar um dicionário e verificar?

O **conceito** é, portanto, um objeto formal, uma construção mental. Cada um tem o direito de imaginar seja lá o que for. E, por ser uma construção mental, pode-se ter até um conceito que não se refira a algo com existência real, por exemplo, um herói de história em quadrinhos ou uma fada. O que queremos dizer é que, se cada indivíduo entende um termo de determinada forma, isso é muito subjetivo e as chances de um entendimento consensual sobre o termo serão mínimas.

Já imaginou isso acontecendo na hora da estruturação de um vocabulário, num SRI? Nessa hora, as pessoas que estão desenvolvendo o vocabulário não podem agir a partir de um livre pensar; há que haver concordância entre as partes que estão definindo o termo, para que só um significado seja atribuído a ele.

Percebendo esse impasse, *Ingetraut Dahlberg* (1978) dá sua contribuição com a Teoria Analítica do Conceito Voltada para o Referente (aqui chamada apenas de Teoria do Conceito). E define conceito de outra forma, no contexto da Biblioteconomia e da Ciência da Informação, visan-

[36] Primeira imagem: Manga (fruta). Disponível em: <https://pixabay.com/pt/manga-frutas-tropicaissuculenta-1982330>; segunda imagem: Camisa. Disponível em: <https://pixabay.com/en/shirt-dress-white-clothesclothing-2345417>; terceira imagem: Lamparina. Disponível em: <https://pixabay.com/pt/lamparina-luz-uvaslâmpada-1804670>.

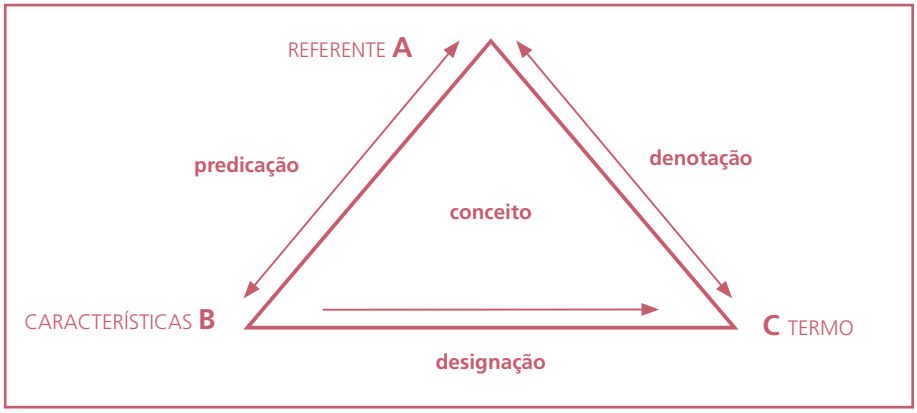
do à representação/recuperação da informação. *Dahlberg* entende que, quando se fala de ciência e tecnologia, o significado dos termos têm que ser fixados, o que se dá por meio de definição. Então, a cientista define conceito como uma **unidade de conhecimento**.

Se você reparar bem, *Dahlberg* (1978) contrapõe essa definição à definição filosófica, que, como vimos antes, diz que o conceito é uma abstração, uma unidade do pensamento. Dessa forma, a pesquisadora restringe sua definição ao âmbito da Ciência da Informação, particularmente à área de Organização do Conhecimento.

Quando *Dahlberg* (1978) se refere à "unidade do conhecimento", fica claro que o objeto a ser definido – o referente – deve ser conceituado de acordo com tudo o que se conhece do objeto até a época em que a definição for feita, ou seja, de acordo com o conhecimento técnico-científico que se tem do objeto, registrado em textos, manuais, glossários, dicionários especializados e outras fontes de referência.

A Figura 39, a seguir, refere-se ao triângulo conceitual, que explica o que é um conceito:

Figura 39 - Triângulo conceitual



Fonte: *Dahlberg* (1978).

Nesse triângulo, o referente pode ser um objeto, um evento, uma experiência, uma produção científica, coisas imaginárias (ex.: fada, duende, fantasma), um processo e outras coisas que são do conhecimento do ser humano; por sua vez, característica refere-se às propriedades do referente; já termo é o signo linguístico, em forma de palavras ou expressões, que expressa o conjunto das características e nomeia o referente.

Em outras palavras, se o conceito for algo conhecido (unidade do conhecimento) e se enunciarmos exatamente suas propriedades (características), então poderemos saber de que se trata.

Vamos fazer um exercício prático. Vou enunciar algumas características de uma coisa, sem citar seu nome; detenha-se no enunciado e imagine que você está de olhos fechados ouvindo a minha voz: "**mamífero** de **grande porte** que respira o **ar atmosférico** e se adaptou à vida na água, sendo, portanto, um **mamífero aquático**. Seus membros anteriores se modificaram em **nadadeiras** e os posteriores desapareceram".

Reparou que as características estão destacadas em negrito no texto?

Conseguiu adivinhar sobre o que falei? Veja se acertou!

Figura 40 - O conhecimento do objeto e suas propriedades permitiu a definição de baleia

Fonte: *Pixabay* (2010).[37]

O enunciado sobre as características das coisas acaba por defini-las. No exemplo dado, a partir do momento em que acabei de enunciar as características de uma baleia, sua definição estava dada. A definição é a soma das características enunciadas e seu papel é individualizar os objetos, de modo que só ele possa ter aquele nome.

O que podemos concluir do triângulo conceitual mostrado antes? Que é possível enunciar as propriedades do referente e expressá-las por meio de um termo. Então, o conceito é tudo isto: o referente, as características e o termo. Sem um dos vértices, o triângulo deixa de existir; da mesma forma, sem um dos elementos, o conceito não existe.

3.7.3 Características dos conceitos e sua classificação

Todas as coisas (tudo quanto existe ou possa existir, de natureza corpórea ou incorpórea) possuem características, por meio das quais as coisas podem ser identificadas. Tais características podem ser classificadas em: extrínsecas e intrínsecas, dependentes e independentes ou características de divisão, como veremos a seguir.

> Características extrínsecas e intrínsecas: convido você a observar a Figura 41, a seguir. Repare bem o tema que deu origem a ela. Em seguida, fique atento aos detalhes.

[37] Disponível em: <https://pixabay.com/pt/baleia-assassina-orca-violação-1945421>.

Figura 41 - Características extrínsecas e intrínsecas dos conceitos

Fonte: produção do próprio autor (2017).

Qual o tema dessa figura? Que se trata de uma figura de bombons, ninguém pode negar. O bombom pode ser caracterizado em si mesmo, sem relação com os outros bombons, não precisa ser comparado a nenhum outro bombom para ser definido da seguinte forma: bombom = confeito, geralmente de chocolate, por vezes com cobertura de glacê ou caramelado geralmente recheado. Esse tipo de característica, que descreve a natureza dos objetos, é chamado de "**intrínseca**" – aquela que caracteriza um referente em si mesmo, sem que tenha relação com os outros. O fato de os bombons serem feitos geralmente de chocolate é outra característica intrínseca; quando se pensa em bombom, pensa-se logo em chocolate, não é verdade? Então, até aqui você viu dois aspectos que caracterizam um bombom: ser um confeito/doce (natureza do objeto definido; no exemplo dado: bombom) e ser de chocolate (matéria de que é feito).

Existem outros aspectos que podem caracterizar um referente em si mesmo, sem que ele necessite ser comparado a outros, por exemplo, forma (o triângulo é definido como polígono de três lados; trilátero); cor (o sangue é definido como líquido vermelho viscoso); tamanho (o anão é definido como um homem de tamanho pequeno).

Recapitulando:

> Característica intrínseca de um referente é a que descreve sua natureza, é a que caracteriza um referente em si mesmo, sem que ele tenha que ser comparado a outros objetos.

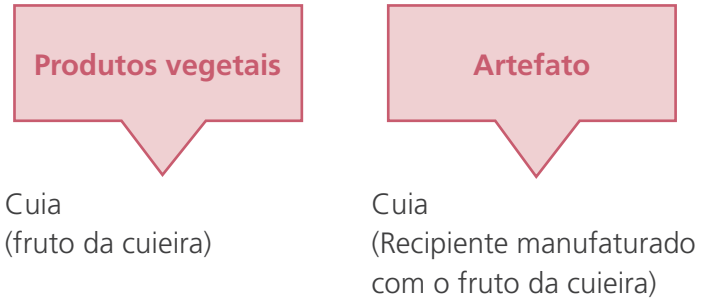
Vamos continuar observando a Figura 41. Basta olhá-la para afirmarmos que os bombons são diferentes entre si. A princípio, o que se pode reparar é que cada um está embalado em um papel de cor diferente: há embalagem vermelha, verde, prata e preta. A "cor", nesse contexto, portanto, é uma característica que os identifica e os diferencia. Só que, desta vez, trata-se de uma **característica extrínseca**, porque o objeto precisa ser comparado a outros para que seja individualizado.

Você notou alguma outra característica além da cor da embalagem? Preste bastante atenção à descrição de cada bombom. Que todos são feitos geralmente de chocolate, isso já foi dito anteriormente (ser de chocolate é uma característica intrínseca a um bombom, inerente a ele). Mas e o tipo de chocolate? E os recheios? Eu reparei que alguns dos chocolates são ao leite e outros de chocolate negro. E reparei também que cada bombom tem um recheio diferente. Se não tinha reparado isso, volte à Figura 41. Pois bem, essas são outras características extrínsecas do bombom. Então, vamos à definição:

> Característica extrínseca de um referente é aquela que o caracteriza por ter uma função/finalidade que outro não tem, ou por ter componentes diferentes de outros referentes semelhantes, ou seja, elas são identificadas num objeto sempre em relação a outro.

Características dependentes e independentes: uma **característica é dependente** de outra se esta tiver que ser definida para que a primeira possa ser compreendida. Por exemplo, no conceito "instrumento de sopro de palheta", o conceito de "palheta" terá que ser definido antes, do contrário, não se compreenderá o que "instrumento de sopro de palheta" significa. A mesma coisa se passa com o conceito "tecido de malha por urdume". Se "urdume" não for conceituado previamente, não se poderá compreender o que é "tecido de malha por urdume".

As **características independentes** são aquelas que produzem conceitos que podem pertencer a mais de uma hierarquia. Por exemplo, num tesauro de cultura material indígena, o termo "cuia" pode ser o fruto arredondado da cuieira, ou o recipiente manufaturado com o fruto da cuieira depois de esvaziado. O termo participa, então, de dois conjuntos, como você pode ver a seguir:



Para que essas noções fiquem bem sedimentadas, não é demais repeti-las:

Característica dependente

Aquela que depende da definição prévia de outra para, então, ser definida.

Característica independente

Aquela que produz conceitos que podem pertencer a mais de uma hierarquia.

Características de divisão: **característica de divisão** não é um tipo de característica, mas uma função que uma característica assume na formação de um renque (termos coordenados, "irmãos", conforme explicado quando vimos Teoria da Classificação Facetada, nesta Unidade, subseção 3.7.2.1). Características de mesma natureza determinam o agrupamento natural dos conceitos em categorias ou subcategorias. O agrupamento se dá por uma característica de divisão.

Como vimos antes, qualquer assunto pode ser visto sob vários aspectos. Por exemplo, suponhamos que você vá comprar um carro. Certamente, você vai considerá-lo segundo sua marca, país de procedência, cor, preço e outros aspectos. Esses aspectos, que de fato são características dos automóveis, determinarão o agrupamento dos conceitos em categorias ou subcategorias, cumprindo a função de organizar, de dividir os conceitos segundo os diferentes pontos de vista. Essas características são conhecidas como características de divisão e vêm indicadas nos tesauros, entre parênteses.

No exemplo citado (carro), a organização ficaria da seguinte maneira: "carro" seria o termo superordenado, com os diversos tipos de carro a ele subordinados – formando uma cadeia; as características de divisão organizariam os renques de termos (termos "irmãos"). Observe o Quadro 9, a seguir:

Quadro 9 - Características de divisão

(termos)	**(relacionamentos)**
Carro	(termo superordenado)
(segundo a função)	*(característica de divisão)*
Carro blindado Carro de boi Carro de corrida Carro de passeio	(renque: termos coordenados/irmãos)
(segundo a marca)	*(característica de divisão)*
Chevrolet Fiat Mitsubishi Renault Volkswagen	(renque: termos coordenados/irmãos)
(segundo a procedência)	*(característica de divisão)*
Carro nacional Carro importado	(renque: termos coordenados/irmãos)

Fonte: produção do próprio autor (2017).

Cada característica de divisão poderá ser desdobrada em outras, tantas quantas forem necessárias para expressar os aspectos representados nos assuntos dos documentos, agrupando conceitos de mesma natureza. Tais desdobramentos devem ser feitos sempre de acordo com o objetivo do sistema a ser desenvolvido, considerando-se os objetos a serem representados no sistema e o usuário. Quanto maior o número de características de divisão, maior será o nível de especificidade do sistema e mais precisa será a busca. Será também maior a visibilidade da coleção existente em todos os seus aspectos.

No exemplo anterior, cada marca de carro poderia ser ainda desdobrada segundo o modelo, assim:

Chevrolet
(segundo o modelo)
Chevrolet Onix
Chevrolet Cruze
Fiat
(segundo o modelo)
Fiat Palio
Fiat 500

Quando abordamos as bases teóricas do tesauro, vimos que a Teoria do Conceito foi descrita por *Dahlberg* (1978). Segundo a autora, as características do conceito podem ser resumidas assim:

a) são obtidas por meio dos predicados (enunciados);
b) os conceitos possuem elementos que são suas respectivas características;
c) um conjunto de características determina o conceito;
d) os conceitos são unidades de conhecimento constituídas dos objetos, associadas a elementos linguísticos.

A Professora *Hagar Espanha Gomes*, referência de estudos elaborados na área de Organização do Conhecimento, que incluem vários trabalhos sobre tesauros, define **característica de divisão** da seguinte maneira: não se trata de um tipo de característica, mas de uma função que dada característica assume na formação de um renque (série horizontal de conceitos que refletem relações lógicas). Características de mesma natureza determinam o agrupamento natural dos conceitos em classes ou subclasses. O agrupamento se dá por uma característica de nível de abstração mais elevado conhecida como característica de divisão, sendo indicada entre parênteses.(CAMPOS; GOMES; MOTTA, 2004).

3.7.4 Relações entre conceitos

Um dos princípios básicos do tesauro-com-base-em-conceitos é que os termos nele incluídos se inter-relacionam, definindo-se uns em relação aos outros e formando conjuntos de significado próximo. No tesauro não existe, então, a possibilidade de haver um termo sozinho.

As relações entre conceitos são obtidas a partir da análise dos termos, consequentemente, dos conceitos.

No tesauro, essas relações são de dois tipos:

a) lógicas: são relações de semelhança que reúnem conceitos com características em comum. Vão produzir relações hierárquicas entre os termos. Ex.: sei que o conceito de "mamífero" se relaciona com o conceito de "gato", porque este é uma espécie de mamífero. A relação gerada será entre um termo que representa um gênero, termo genérico (TG), e outro que representa uma espécie, termo específico (TE). Outros animais que mamam também estariam subordinados a "mamífero." E se o assunto em foco for "animais", o conceito de "mamífero" estará, ainda, subordinado a outro: "vertebrado".

   Nos tesauros, essas relações têm sido representadas por TG e TE. No exemplo dado, a representação ficaria assim:

   TG Vertebrado
   TE Mamífero
      TE Gato

b) ônticas: essas relações incluem duas outras – as partitivas e as associativas, como a seguir:

   - b.1) relações partitivas: como o nome indica, o conceito abrangente se refere ao todo, enquanto os partitivos se referem às partes. Por exemplo, biela é uma peça do motor, é parte dele.

     Nos tesauros, essas relações têm sido representadas por termo total (TT) e termo partitivo (TP). No exemplo dado, a representação ficaria assim:

     TT Motor
     TP Biela

   - b.2) relações associativas: são estabelecidas a partir da relação que os objetos guardam entre si no mundo real, associando-se. Por exemplo, existe uma relação de associação entre "tecido", "tinta" e o "processo de tecelagem". Nos tesauros, essas relações têm sido representadas por termo associado (TA) ou termo relacionado (TR). Preferimos optar pelo primeiro símbolo, por entendermos que todas as relações num tesauro dão-se por relacionamentos, enquanto, nas relações associativas, especificamente, elas se dão por associação. No exemplo dado, a representação ficaria assim:

     Tecelagem
     TA Tecido
     TA Tinta

3.7.5 Etapas da elaboração de um tesauro

Multimídia

Do leite ao queijo

Figura 42 - Cena de vídeo sobre a produção do queijo minas

Fonte: *Youtube* (2011).

Se tiver chance, assista a esse vídeo, que mostra as etapas de produção do queijo tipo minas, vale a pena! Embora ele quase não tenha narrativa em palavras, as imagens revelam claramente a demarcação entre uma etapa e outra do processo de fabricação de um produto comestível.

Você pode perguntar: o que isso tem a ver com o tesauro? E a resposta é simples: o que eles têm em comum é que tanto um tesauro quanto um produto comestível passam por processos para sua produção, em que cada etapa deve ser seguida rigorosamente para que o resultado seja satisfatório. O *link* para o vídeo é: <www.youtube.com/watch?v=r-1q844fp2vE>. Confira!

Os tesauros-com-base-em-conceito podem ser elaborados de acordo com as seguintes etapas:

a) planejamento: um tesauro consistente deve ser elaborado a partir de um cuidadoso planejamento, que inclui várias etapas, descritas no *Manual de elaboração de tesauro – tutorial* (CAMPOS; GOMES; MOTTA, 2004). Elas foram aqui sintetizadas da seguinte forma:

Quadro 10 - Etapas do planejamento de um tesauro

- Delimitação da área: que assunto(s) o tesauro vai abranger?
- Forma de criação: será desenvolvido por uma biblioteca de uma só instituição ou por uma rede de bibliotecas?
- Público-alvo: a quem o tesauro vai servir?
- Classificação: que categorias o esquema vai ter?
- Levantamento das fontes: de onde os termos serão coletados?
- Forma de apresentação: o tesauro será apresentado em que partes (alfabética, sistemática, planigráfica, algumas delas ou todas elas?)
- Período de atualização: a terminologia da área varia pouco ou muito? O número de termos novos sugeridos requer uma atualização em menor prazo?
- Divulgação: o tesauro será disponibilizado em cópia papel, em CD, na internet?
- Seleção do *software*: qual o *software* mais adequado ao tipo de tesauro que será desenvolvido? Está disponível no mercado? Há recursos para sua aquisição?
- Manutenção: a melhor forma de manter o tesauro atualizado e funcional é fazer sua manutenção frequentemente, a intervalos pequenos. Não se esqueça de desenvolver mecanismos auxiliares que ajudem a fazer o controle dessa manutenção (data de entrada do termo no sistema, data de alteração, data de exclusão, fontes pesquisadas para retirada e validação do termo, outros mecanismos). Cada termo admitido no tesauro deve ter um registro individual.

Fonte: *Campos*, *Gomes* e *Motta* (2004).

b) levantamento do vocabulário: a coleta de termos que formará o vocabulário do tesauro é muito importante. Dependendo da área (ou áreas) de assunto a ser coberta, você terá algumas opções. Há situações em que o próprio vocabulário já usado na indexação dos documentos de um SRI (com ou sem controle terminológico) servirá de base para a seleção dos termos. Outras situações exigirão a coleta a partir de fontes especializadas: dicionários especializados, glossários, documentos normativos, fontes confiáveis em meio eletrônico, consulta a técnicos e outras fontes disponíveis e cabíveis. A eficácia do tesauro dependerá em grande parte dessa seleção, pois, se o vocabulário não estiver adequado aos interesses dos usuários, ele será um grande motivo para insatisfações.

A estrutura dos sistemas de controle terminológico torna-se mais confiável se os termos de seu vocabulário forem bem definidos. E a explicação é simples: a base para o relacionamento entre os termos está na definição.

Como dito antes, a soma das características do conceito é igual ao conceito e, como essas características também são conceitos, a indicação do que terá que ser relacionado quase sempre é clara.

Por exemplo, vamos reparar na definição do termo "floresta amazônica": "tipo de **floresta tropical** situada na porção norte da **América do Sul**, que possui a maior **diversidade** e complexidade desse **ecossistema**". A definição já nos indica bastante coisa sobre os relacionamentos que serão feitos. Mas, de saída, já podemos saber que as palavras destacadas em negrito certamente irão se relacionar ao termo "floresta amazônica."

Multimídia

Se quiser saber mais sobre esse assunto, vou dar uma dica: consulte o trabalho *Método relacional como nova abordagem para a construção de tesauros* (MOTTA, 1987). Lá você encontrará muitos termos definidos da área de Economia e poderá ver como se deram as relações. Basta visitar o *site* <http://www.conexaorio.com>.

c) organização dos conceitos: deverá seguir as seguintes etapas:

- c.1) categorização da área de assunto, momento em que serão estabelecidas as categorias gerais de assunto que irão agrupar os termos com características em comum (elaboração do esquema de facetas: aplicação dos princípios do "Método de Facetas/Categorização", compreendendo estudo preliminar em fontes especializadas da área para conhecimento de sua abrangência);
- c.2) fragmentação do campo de assunto em facetas (princípio de análise em facetas);
- c.3) levantamento de subfacetas; classificação dos termos nas categorias;
- c.4) indicação de relações entre termos (lógicas e semânticas).

d) apresentação: uma vez os termos analisados e devidamente relacionados, chegou a hora de providenciar as formas de apresentação final do tesauro.

Os termos e suas inter-relações podem ser apresentados de várias maneiras:

- d.1) apresentação alfabética, com notas (definições, notas de aplicação) e relações entre os termos indicados em cada um deles;
- d.2) apresentação sistemática, auxiliada por um índice alfabético;
- d.3) apresentação gráfica, com uma seção alfabética.

Existem tesauros com uma única forma de apresentação, como os alfabéticos. Outros apresentam duas formas associadas, a alfabética e a sistemática. Outros, ainda incluem uma parte gráfica.

Multimídia

Cultura material indígena mais perto de você

Figura 43 - *Tesauro de Cultura Material dos Índios no Brasil*

Fonte: produção do próprio autor (2017).

O *Tesauro de Cultura Material dos Índios no Brasil*, de autoria de Dilza F. da Motta, possui as três formas de apresentação, além de um suplemento fotográfico. Se você não dispuser de um exemplar, solicite um à *Biblioteca Marechal Rondon*, do *Museu do Índio*, no Rio de Janeiro. Você também poderá consultar outro tesauro bastante interessante no *link* <www.cnfcp.gov.br/tesauro/>, para observar o tesauro de cultura popular do *Centro Nacional de Folclore e Cultura Popular*. Repare em sua apresentação.

e) elaboração de instruções de uso: um tesauro será de muito mais utilidade se contiver uma parte preliminar com instruções para seu uso. Nas instruções devem constar esclarecimento sobre os seguintes elementos:

- e.1) âmbito temático;
- e.2) número de termos;
- e.3) língua ou línguas usadas (texto original e versões, se houver);
- e.4) objetivo;
- e.5) termos: fontes usadas para o levantamento/coleta; forma de apresentação; uso de singular e plural; nível de pré-coordenação; apresentação tipográfica, tipos de inter-relacionamentos dos termos, notas referentes aos termos e convenções adotadas;
- e.6) categorias e subcategorias incluídas e seus conteúdos temáticos;
- e.7) apresentação do tesauro;
- e.8) prazo previsto para atualização do tesauro;
- e.9) critérios para avaliação de tesauros.[38]

[38] Esta parte será apresentada em separado na seção 3.6 desta Unidade.

3.7.6 **Atividade**

Aprendendo a facetar

Vamos fazer um passeio por uma loja de departamentos. Esse nome já diz tudo: se são departamentos, é porque as mercadorias estão departamentalizadas, ou seja, colocadas em seções. Pense nessas seções como se fossem nossas "categorias/facetas." Como as mercadorias estão arrumadas? Elas estão organizadas de acordo com vários aspectos pelos quais podem ser vistas: material de que são feitas (objetos de plástico, vidro, cerâmica), função (objetos para cozinha, artigos de cama e mesa, artigos de higiene), roupas classificadas por sexo (*lingeries*, cuecas) ou por faixa etária (para bebês, crianças, jovens e adultos), entre outros critérios.

Assim são todas as coisas da natureza: podem ser enquadradas em categorias/facetas. E locais como esses são ótimos para exercitarmos nossa imaginação e categorizarmos itens. Em qualquer ambiente você pode repetir esse exercício, seja em uma loja comercial, em uma exposição, em um clube, em uma empresa. Tente!

Mas agora, mãos à obra! Imagine que você foi incumbida/o de organizar um vocabulário de música . A seguir, você vai encontrar vários termos dessa área que você deverá categorizar, todos já conceituados, permitindo que você os categorize adequadamente. Depois, você também vai encontrar um quadro com sugestão de uma categorização (facetação da área de música), com alguns termos já enquadrados nas categorias. Observe os termos:

- ACORDEÃO – Instrumento de fole e teclado que possui um único fole e utiliza um sistema de palhetas metálicas internas para produzir sons, um teclado como o do piano na mão direita e outro de botões para o acompanhamento dos baixos na esquerda.
- ACÚSTICA (MÚSICA) – Conjunto das qualidades sônicas de recintos e edificações, e da transmissão do som pela voz, por instrumentos musicais ou por meios elétricos.
- ANÁLISE MUSICAL – Parte do estudo da arte musical que tem como referência a própria música, e não parâmetros externos.
- APRECIAÇÃO MUSICAL – Disciplina teórica e prática que visa fornecer elementos para a audição musical através da história da música, com audições comentadas de obras musicais.
- ARCO (MÚSICA) – Acessório de instrumento musical que se constitui de uma vareta de madeira curvada.
- AVALIAÇÃO MUSICAL – Processo educativo, contínuo, sistemático e complexo que leva em conta duas dimensões no aprendizado de música: o que os alunos estão fazendo e o que eles estão aprendendo.
- BANDA DE MÚSICA – Conjunto de instrumentos cuja estrutura básica se apoia nos sopros. A banda moderna é formada por madeiras (flautas, clarinetas, saxes), metais (trompas, trompete, trombones e tubas) e percussão (tambores, *glockenspiel*, marimba, caixas-claras etc.).
- BANDA DE *ROCK* – Qualquer conjunto com qualquer formação que toque *rock*.

- CANÇÃO – Composição musical para a voz humana, geralmente acompanhada por instrumentos musicais e letras. É tipicamente interpretada por um único vocalista, mas também pode ser cantada por um dueto, trio ou mais vozes.
- CANTIGA – Pequena canção, trova, toada, feita para cantar o amor, sentimentos religiosos, acontecimentos históricos e outros assuntos ligados à vivência do povo.
- CANTOR – Indivíduo que tem o ofício de cantar.
- COMPOSITOR – Profissional que escreve música.
- CONGADA – Gênero musical brasileiro dançado, de espírito épico, que, no norte do Brasil, é mais conhecido como congo.
- DESENVOLVIMENTO MUSICAL – Processo pelo qual uma ideia musical é comunicada no curso de uma composição, referindo-se à transformação e redeclaração do tema inicial.
- DISCO (GÊNERO MUSICAL) – Gênero musical dos anos 1970 que mesclava ingredientes do *soul* e do *rock'n'roll*.
- DJ – Indivíduo que pilota toca-discos em casas noturnas.
- EDUCAÇÃO MUSICAL – Processo educativo que dá ao indivíduo oportunidade de acesso à música enquanto arte, linguagem e conhecimento.
- ESTILO MUSICAL – Maneira particular de compor e tocar, associada a um compositor, intérprete ou lugar (estilo alemão, napolitano, mozartiano, rococó, por exemplo).
- ETNOMUSICOLOGIA – Disciplina da Musicologia que aborda a música em seu contexto histórico, social e cultural, combinando abordagens estéticas e antropológicas.
- FALSETE – Técnica utilizada para alongar a extensão vocal que consiste no emprego de um modo de vibração reduzido das cordas vocais e simula, principalmente, no canto adulto masculino, vozes tipicamente femininas.
- FREVO – Gênero tocado por fanfarras, formadas principalmente por clarinetas e requintas, saxofones, trompetes, trombones, tubas e percussão.
- *FUNK* – Gênero musical que, nos Estados Unidos, é caracterizado por progressões harmônicas extremamente simples, forte acompanhamento percussivo e alto volume de amplificação, geralmente com uma seção de metais e saxofones.
- GRAVAÇÃO SONORA – Processo que envolve a conversão de sons em sinais elétricos para registro material temporário ou permanente.
- HISTÓRIA DA MÚSICA – Disciplina que estuda as origens e evolução da música ao longo do tempo.
- INSTRUMENTO MUSICAL – Artefato, dispositivo, aparelho ou qualquer objeto, construído, adaptado pelo homem ou encontrado na natureza, que é utilizado para produzir sons passíveis de serem identificados como música.
- LETRA DE MÚSICA – Texto poético da música.
- LINGUAGEM MUSICAL – Sistema de descrição, em palavras, de elementos musicais e a relação entre a simbologia da música e seu desempenho prático.
- MÚSICO – Pessoa que exerce atividades ligadas à música.

- POLIFONIA VOCAL – Estilo musical no qual várias vozes ou partes instrumentais são combinadas de maneira contrapontística (mantendo a individualidade da linha) em oposição à homofonia (um único som ou melodia).
- *RAP* – Gênero de música popular urbana, que consiste numa declamação rápida de rimas ritmadas e improvisadas de um texto.
- REGENTE – diretor de orquestra, banda, orfeão etc.; maestro.
- TEORIA MUSICAL – Disciplina que elabora os elementos de estudo musical.
- VIOLA – Instrumento de arco e cordas friccionáveis semelhante ao violino, porém de maiores dimensões e som mais grave.

COISA	GÊNERO MUSICAL	PROCESSO	AGENTE	ÁREA DO CONHECIMENTO
Banda de música		Avaliação musical	Músico	Apreciação musical
Letra de música	*Rap*	Desenvolvimento musical		História da música
	Congada			
		Educação musical		
Linguagem musical			Regente	

Agora, é só você ler com atenção as definições dos termos (que muitas já dão a dica do nome que a categoria deve ter), escolher as categorias adequadas e encaixar os termos em cada uma delas. Inspire-se no exemplo do quadro mostrado e crie o seu. Mas não se esqueça de consultar as categorias do *CRG* (veja Quadro 8 – Categorias do *Classification Research Group*).

Resposta comentada

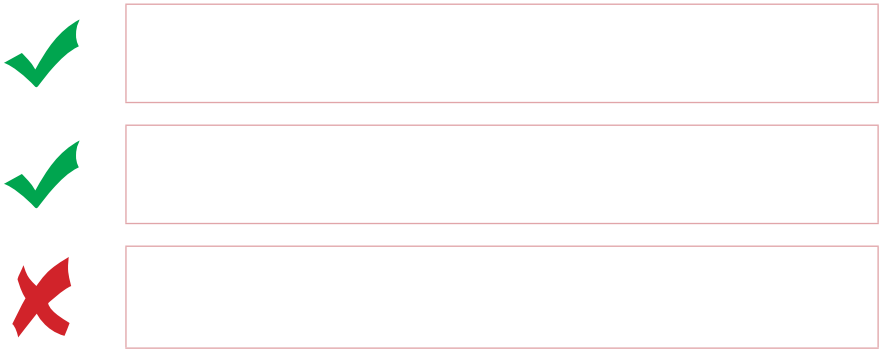
Mesmo sem ser um especialista na área musical, imagino que você não tenha sentido dificuldade em completar o quadro de categorias. Veja como ficou (Quadro 11):

Quadro 11 - Categorias

COISA	GÊNERO MUSICAL	PROCESSO	AGENTE	ÁREA DO CONHECIMENTO
Banda de música	FREVO	Avaliação musical	Músico	Apreciação musical
Letra de música	*Rap*	Desenvolvimento musical	CANTOR	História da música
BANDA DE *ROCK*	Congada	GRAVAÇÃO SONORA	COMPOSITOR	ETNOMUSICOLOGIA
INSTRUMENTO MUSICAL	DISCO (GÊNERO MUSICAL)	Educação musical	DJ	TEORIA MUSICAL
Linguagem musical	*FUNK*		Regente	

3.8 AVALIANDO UM TESAURO

Figura 44 - Avaliar mais para errar menos



Fonte: *Pixabay* (2016).[39]

Você já deve ter concluído que a existência de um instrumento de controle vocabular em um SRI é muito importante para contornar o problema da polissemia característica da linguagem natural, facilitando o processo de comunicação entre o sistema e o usuário. Se esse instrumento for um tesauro, o SRI estará com sorte, pois, como já dito, o tesauro é uma ferramenta que se apoia em princípios teórico-metodológicos estabelecidos internacionalmente, constituindo-se, assim, em uma ferramenta confiável.

Mesmo sabendo de sua utilidade, a construção de um tesauro não é tarefa fácil, envolvendo altos custos, mão de obra especializada e períodos longos para sua execução. Na maioria das vezes, as instituições querem resultados imediatos, porém a elaboração de um tesauro está longe disso.

Por isso tudo, são raras as iniciativas de construção de tesauros feitas por uma só instituição. Projetos desse porte devem ser desenvolvidos por um grupo de instituições que tenham interesses em comum. Como o tesauro se destina a cobrir uma área de assunto e, às vezes, outras correlatas, não é difícil serviços de informação especializados unirem-se para trabalhar em rede.

Todas as dificuldades que envolvem a construção dessa ferramenta fazem com que os SRI lancem mão de tesauros já prontos, embora essa situação não seja a ideal. É preciso muito cuidado com o reaproveitamento de tesauros existentes. Quanto a essa questão, uma primeira pergunta deve ser feita: "existe compatibilidade entre a metodologia da construção do tesauro já existente e a do tesauro que será construído?" Imagine se as unidades de representação de assuntos adotadas nos dois forem diferentes, se um adotou a palavra e o outro, o conceito!

Ainda que os dois tesauros tenham usado a mesma unidade de representação de assuntos e que essa unidade tenha sido o conceito, a questão é delicada. Conceitos variam de acordo com o contexto informacional: um conceito usado para os objetivos de uma instituição pode não servir, exatamente, aos de outra. Como exemplo, podemos lembrar dois

[39] Disponível em: <https://pixabay.com/pt/lista-de-verificação-teste-seleção-1402461>.

conceitos de AIDS. Em uma Secretaria de Saúde, talvez a melhor definição do conceito fosse: "doença sexualmente transmissível"(conceito aplicável, por exemplo, numa campanha para incentivar o uso do preservativo durante o Carnaval); já na *Fundação Oswaldo Cruz* (*FIOCRUZ*), talvez fosse melhor definir como: "doença infectocontagiosa"(já que uma das áreas da *FIOCRUZ* dedica-se à investigação de vírus e de doenças virais).

Apesar dessas considerações, o reúso de tesauros existentes é um expediente utilizado pelos SRI. Nesse caso, é necessário fazer uma análise cautelosa e, se necessário, tornar compatíveis os conceitos dos tesauros que servirem de base para o desenvolvimento de outros.

A literatura sobre avaliação de tesauros não é extensa. Ainda assim, pode-se sugerir alguns critérios para realizá-la. Eles são recomendados para fundamentar decisões de um serviço de informação quanto à adoção ou não de um tesauro existente (CAMPOS, 2003), conforme síntese a seguir:

a) Qual o domínio de conhecimento coberto pelo tesauro?
b) O tesauro apresenta introdução? Qual o conteúdo dela? Está redigida de forma clara a possibilitar o uso do instrumento?
c) Quanto à forma de apresentação do tesauro:
   c.1) Apresenta parte alfabética?
   c.2) Apresenta parte sistemática?
   c.3) Apresenta outra forma qualquer de exibição do esquema?
d) Qual o idioma do texto original? É monolíngue ou multilíngue?
e) Qual a unidade linguística utilizada: palavra, assunto ou conceito?
f) Quais os tipos de relação encontrados?
g) Quanto aos aspectos ligados à consistência:
   g.1) As relações entre os termos são consistentes?
   g.2) O uso do plural e do singular dos termos é consistente?
   g.3) O nível de especificidade é consistente?
h) Inclui informações sobre o termo, como notas de aplicação/escopo para esclarecer sobre seu uso? Apresenta a definição do termo para contextualizá-lo? Apresenta notas com instruções sobre como indexá-lo?

Esses critérios podem ser identificados a partir do exame do tesauro que se quer avaliar. Em outra direção, vários testes estatísticos para avaliação de tesauros têm sido propostos e aplicados, levando em conta medidas relativas a: termos descritores e não descritores, definições dos termos, nível de pré-coordenação, proporção do número de palavras que formam cada termo e outras medidas.

Nesta Unidade, nos dedicamos a tesauros, portanto, é oportuno esclarecer o que se entende por **descritores** e **não descritores**, já que são termos usados na bibliografia sobre essas ferramentas.

Descritor

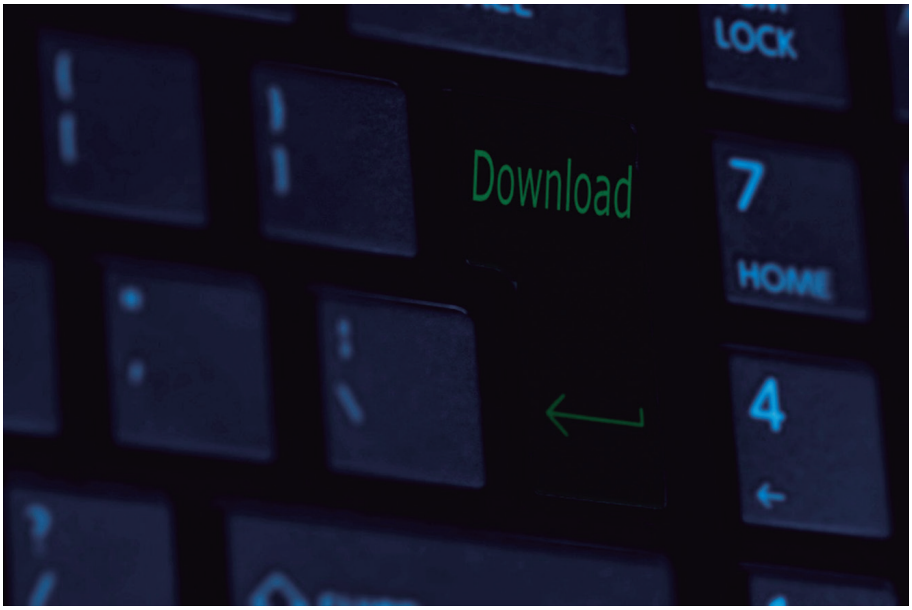
Termo preferido para representar um conceito que faz parte do vocabulário de entrada de um tesauro, criado por *Calvin Mooers* em 1948.

Não descritor

Termo que faz parte do vocabulário de entrada de um tesauro, mas não foi escolhido para uso.

3.9 AVALIANDO *SOFTWARES* PARA ELABORAÇÃO DE TESAUROS

Figura 45 - Como escolher o *software* adequado ao meu tesauro?



Fonte: *Pixabay* (2015).[40]

Já dissemos que a tarefa de construir um tesauro é difícil, cara e demorada. Esse trabalho é facilitado se dispusermos de um *software* que auxilie na construção e uso desses instrumentos. Um *software* para a construção de tesauros possui características fundamentais que valorizam sua aplicação: facilita o armazenamento, a manipulação de dados e a apresentação dos termos e de suas relações. Tendo isso em vista, a escolha desse dispositivo deixa de ser uma facilidade e passa a ser uma necessidade.

Daí a importância de selecionar um *software* que contemple não só aspectos metodológicos da construção de um tesauro, como também aspectos operacionais, independentemente da metodologia adotada, como mecanismos para a criação e controle de atualização de termos, possibilidade de reutilização de outros tesauros e elaboração de relatórios, por exemplo.

[40] Disponível em: <https://pixabay.com/pt/teclado-baixar-software-seta-www-967225>.

Multimídia

Avaliando um *software* para tesauros

Figura 46 - Não use gato por lebre!

Fonte: *Pixabay* (2016).[41]

Evite surpresas desagradáveis! Fique bem alerta aos critérios para a escolha de um *software* para seu tesauro. Você poderá saber muito mais detalhes deles consultando o "Estudo comparativo de *software* de construção de tesauros", no *link*: <http://www.scielo.br/pdf/pci/v11n1/v11n1a06>.

A seguir, veja critérios para a avaliação de *software* de tesauros, sistematizados a partir de um estudo realizado por uma equipe de profissionais da informação coordenada por *Campos* (2006). Os requisitos considerados são divididos em oito grupos, a saber: características gerais, tratamento de relações, tratamento de dados, interface/manipulação dos dados, relatórios, características de implementação, apoio metodológico, apoio ao usuário. Você verá também a relação desses requisitos, com suas respectivas especificações.

a) Características gerais:

a.1) Tem limite para número de termos a serem incluídos?

a.2) Permite o uso de termos compostos?

a.3) Permite registrar a fonte ou origem do termo?

a.4) Permite criar classes de assunto?

a.5) Permite registrar os termos em mais de um idioma?

b) Tratamento de relações:

b.1) Permite relações definidas pelo usuário?

[41] Disponível em: <https://pixabay.com/pt/gato-preto-e-branco-coelho-roux-1359634>.

b.2) Permite relações-padrão?

b.3) Permite as relações adotadas no padrão *ISO 5964* (1985)?

c) Tratamento de dados:

c.1) Permite guardar o tesauro em banco de dados?

c.2) Permite a importação/exportação de termos de/para outras fontes?

c.3) Permite crítica em relação à consistência de dados?

c.4) Permite controle de restrição de acesso para atualização do tesauro?

c.5) Permite gerenciamento da situação (*status*) de termos, como termos candidatos e obsoletos, por exemplo?

d) Interface/manipulação dos dados:

d.1) Permite busca flexível por termos?

d.2) Permite reorganizar a hierarquia de termos?

d.3) Possui interface *web*?

d.4) Permite criar o tesauro intuitivamente?

e) Relatórios: permite emitir relatórios? Quais?

e.1) ( ) alfabético;

e.2) ( ) sistemático (inclusive com relações hierárquicas e partitivas?);

e.3) ( ) KWIC;

e.4) ( ) hierárquico;

e.5) ( ) definido pelo usuário.

f) Características de implementação:

f.1) Depende de *software* não livre?

f.2) Roda em Unix?

f.3) Roda em *Windows*?

f.4) É gratuito?

g) Apoio metodológico (inclui apoio ao uso de alguma metodologia para construção de tesauros?):

g.1) Tem apoio para representação gráfica do escopo?

g.2) Tem apoio para o estabelecimento de relações entre os termos?

g.3) Para categorização com base no conceito?

g.4) Para o uso de características de divisão – as que permitem agrupar termos de uma mesma hierarquia de acordo com algum atributo comum?

h) Apoio ao usuário:

h.1) Possui suporte?

h.2) Possui grupo de usuários ativo?

h.3) Possui ajuda *on-line*?

h.4) Possui documentação explicativa do *software*?

Não existe dúvida quanto aos benefícios de um *software* para a construção de tesauros, mas eles possuem funcionalidades, custos e apoios metodológicos diferentes. Por isso, sua escolha deve ser criteriosa, orientada por requisitos mínimos de qualidade e funcionalidade, adequados aos objetivos e ao tipo de tesauro que se quer construir.

3.9.1 Atividade

Leia as afirmações e marque com V as que forem verdadeiras e com F as falsas.

1. ( ) Se minha instituição não tem recursos financeiros disponíveis, posso usar qualquer tesauro que existir sobre o assunto que me interessa.
2. ( ) A documentação do SRI onde trabalho é especializada em Economia. Isso significa que posso me basear num índice de periódicos só de Economia e, sobre ele, aplicar os princípios de construção de tesauros conceituais para melhorar meu vocabulário de indexação.
3. ( ) A apresentação alfabética de um tesauro é um dos critérios mais importantes em sua avaliação, e sufuciente para sua utilização.
4. ( ) Os tesauros que contêm definições dos termos têm maior probabilidade de apresentar relações entre eles estabelecidas em bases mais sólidas.
5. ( ) Se tenho que desenvolver um tesauro, a aquisição de um *software* para essa tarefa é o caminho mais fácil.
6. ( ) É possível adotar um *software* que não gere relatórios sistemáticos (hierárquicos), porque isso não vai fazer falta.
7. ( ) Os tesauros desenvolvidos por várias instituições e por uma equipe multidisciplinar de técnicos são mais bem-sucedidos.
8. ( ) Nos tesauros em que não há definição para todos os termos, é necessário que existam, pelo menos, notas de explicação sobre termos que sejam ambíguos.
9. ( ) Sou gerente de um SRI cujo vocabulário usado para indexação é uma relação de cabeçalhos de assunto. Tento convencer meu chefe a transformá-lo em um sistema baseado em conceitos, pois isso será feito sem custo e rapidamente.
10. ( ) O *software* que eu sugerir para aquisição deve possibilitar a categorização dos termos.

Resposta comentada

1. (F) Todo tesauro deve ser avaliado antes de ser usado, para que seja possível saber se ele corresponde aos requisitos mínimos do sistema que se quer desenvolver.
2. (V) Os princípios técnico-metodológicos para a construção de tesauros aplicam-se à estrutura de qualquer vocabulário.

3. (F) A apresentação alfabética por si só empobrece o vocabulário.
4. (V) A definição do termo é importantíssima na geração de um vocabulário; as características das definições indicarão relacionamentos que deverão ser feitos entre os termos.
5. (V) Um *software* adequado é um fator essencial para o desenvolvimento de um tesauro, com garantia de qualidade e confiabilidade. Atualmente, o desenvolvimento manual desse tipo de ferramenta está fora de cogitação.
6. (F) A apresentação de relatórios sistemáticos é importantíssima, pois revela como a área do conhecimento representada se estrutura, servindo de guia para buscas mais refinadas.
7. (V) Um tesauro é uma ferramenta complexa e cara. Por isso, uma iniciativa conjunta entre instituições e técnicos possibilitará que o projeto seja mais bem-sucedido.
8. (V) Em um tesauro, são necessárias clareza e precisão dos significados dos termos.
9. (F) Você está certo em sugerir a transformação de seu vocabulário para um sistema de conceitos, mas deve alertar a chefia sobre os custos e necessidades de mão de obra especializada que isso envolverá.
10. (V) Se o que você deseja é ter um sistema mais preciso, ele deve ser baseado nos princípios adotados para tesauros-com-base-em-conceito, e um dos métodos usados em sua elaboração é o de faceta, que permite a categorização dos termos em classes abrangentes.

3.10 CONCLUSÃO

Apesar do aparecimento de algumas iniciativas modernas no campo da representação de assuntos para apoiar a indexação e a recuperação da informação, como as folksonomias, o tesauro documentário tem provado ser, até o momento, uma ferramenta eficaz para esse fim.

Há especialistas que argumentam que o tesauro é uma "lista de cabeçalhos de assunto disfarçada", porque ambos conteriam propostas idênticas. A nosso modo de ver, o tesauro representa um corte em relação às listas de cabeçalhos de assunto, principalmente por se caracterizar como um sistema baseado em conceito, transformando-se em um instrumento com características diferentes: a unidade de representação deixa de ser a palavra e passa a ser o conceito; os relacionamentos entre os termos são mostrados com clareza, indicando, inclusive, o tipo de relação estabelecida entre eles; o tesauro se conforma perfeitamente à tecnologia atual (computadores), além de adotar princípios calcados em três teorias (da Classificação Facetada, da Terminologia e do Conceito), tornando-se fonte mais segura na indexação e mais confiável nas buscas.

Enfim, o tesauro surge como um instrumento poderoso de comunicação entre o sistema e o usuário, orientando o analista sobre quais os melhores termos para representar o assunto do documento a ser indexado, ao mesmo tempo que também orienta o usuário na escolha de termos que melhor representem o assunto por ele procurado. É uma solução cara, que exige mão de obra especializada? Sim, mas nada que um consórcio não resolva. Surgirão novas iniciativas? Só o tempo dirá.

RESUMO

A palavra *tesauro* foi empregada há séculos para a organização de dicionários alfabéticos. Ela se popularizou com *Roget*, em 1852, devido a seu *Dicionário Analógico – Roget's Thesaurus of English Words and Phrases*, que é considerado o primeiro tesauro moderno. Porém, foi nos anos 1960 que surgiram vários tesauros com a proposta de servir como instrumentos de controle terminológico rigoroso.

Essa nova tecnologia promoveu uma ruptura em relação às listas de cabeçalhos de assunto, porque introduziu a possibilidade da busca pós-coordenada e o uso do conceito, e não da palavra, para representação dos assuntos de um documento. Com isso, conferiu às buscas maior precisão na recuperação da informação.

Atualmente, os tesauros são desenvolvidos com base em princípios estabelecidos em normas internacionais e se apoiam em três teorias (a da Classificação Facetada, a Geral da Terminologia e a do Conceito), sendo chamados de tesauro documentário. Eles têm, como unidade de representação, o conceito (tesauro-com-base-em-conceito).

Há vários tipos de tesauro, dependendo do aspecto pelo qual ele seja observado: monolíngue e multilíngue; multidisciplinar e especializado; macrotesauro e microtesauro.

Até hoje, o tesauro assim construído tem se mostrado um instrumento para representação temática eficaz, auxiliando na indexação/recuperação da informação. Seu desenvolvimento é oneroso e complexo, sendo aconselhável a formação de consórcios para sua execução.

Sugestão de Leitura

ÁLVARES, Lilian. Tesauro. **Lilian.alvarestech**, Brasília, [20--?]. Apresentações. Disponível em: <http://lillian.alvarestech.com/Fundamentos/Modulo2/Aula22Tesauros.pdf>. Acesso em: 12 jan. 2015.

DALHBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, [S.l.], v. 7, n. 2, 1978. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/1680/1286>. Acesso em: 13 jan. 2014.

GOMES, Hagar Espanha. **Elaboração de tesauro documentário:** aspectos teóricos e práticos. Rio de Janeiro: [S.n.], 1998. Manuscrito.

MARRONI, Gilza Núria Brandão. **Identificação e delimitação de relações associativas em tesauros:** um estudo de caso na área de Direito do Trabalho. 2006. 127f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2006.

REFERÊNCIAS

AITCHISON, Jean. The Thesaurofacet: multipurpose retrieval language tool. **Journal of Documentation**, London, v. 26, n. 3. p. 187-203, Sept. 1970.

AMERICAN NATIONAL STANDARDS INSTITUTE. **ANSI Z39.19:1980:** American National Standard Guidelines for Thesaurus Structure, Construction, and Use. Bethesda: Md. NISO, 1980. 20 p.

AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/ NISO Z 39.19:** guidelines for the construction, format and management of monolingual controlled vocabularies. Bethesda: Md. NISO, 2005. 176 p.

AZEVEDO, Francisco Ferreira dos Santos. **Dicionário analógico da língua portuguesa:** ideias afins/thesaurus. 2. ed. atual. rev. Rio de Janeiro: Lexikon, 2010.

CAMPOS, Maria Luiza de Almeida. **Linguagem documentária:** teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001. 133 p.

CAMPOS, Maria Luiza de Almeida et al. **Critérios para avaliação de tesauros**. Relatório acadêmico de monitoria. [S.l.: s.n.], 2003.

CAMPOS, Maria Luiza de Almeida et al. Estudo comparativo de *software* de construção de tesauros. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 68-81, 2006.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha; MOTTA, Dilza Fonseca da (Coord.). Elaboração de tesauro

documentário: tutorial. **Conexão Rio**, Rio de Janeiro, 2004. Disponível em: <http://www.conexaorio.com/biti/tesauro/index.htm> Acesso em: 21 jan. 2015.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, Brasília, v. 7, n. 2, 1978. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/1680/1286>. Acesso em: 13 jan. 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval. Suíça: ISO, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-2:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 2: Interoperability with other vocabularies. Suíça: ISO, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2788:** guidelines for the establishment and development of monolingual thesauri. Suíça: ISO, 1974.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 5964:** guidelines for the establishment and development of multilingual thesauri. Suíça: ISO, 1985.

LANCASTER, F. W. **Vocabulary control for information retrieval**. 2nd ed. Arlington: Information Resources Press, 1986.

MOTTA, Dilza Fonseca da. **Método relacional como nova abordagem para a construção de tesauros**. 1987. 90 f. Dissertação (Mestrado em Ciência da Informação) – UFRJ/IBICT, Rio de Janeiro, 1987.

MOTTA, Dilza Fonseca da. **Tesauro de cultura material dos índios no Brasil**. Rio de Janeiro : Museu do Índio, 2006. 250p.

RANGANATHAN, Shiali Ramamrita. **Prolegomena to Library Classification**. 3rd ed. Bombay: Asia Publising House, 1967.

UNITED NATIONS FOR EDUCATION, SCIENCE AND CULTURE ORGANIZATION. **Guidelines for the establishment and development of monolingual thesauri**. Paris: UNESCO, 1973.

WÜSTER, Eugen. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. In: RONDEAU, Guy; FELBER, Helmut. **Textes choisis de terminologie**. Québec: GIRSTERM – Université de Laval, 1981. p. 55-108.

UNIDADE 4

ONTOLOGIAS

4.1 OBJETIVO GERAL

Caracterizar as ontologias como instrumentos de representação do conhecimento e sua influência nas novas propostas de SRI.

4.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

a) compreender o que são ontologias, sua história, características e componentes;

b) identificar as funções e aplicações de uma ontologia;

c) identificar as principais metodologias, ferramentas e linguagens usadas para a construção de ontologias.

4.3 MAIS UMA LINGUAGEM A SERVIÇO DA RECUPERAÇÃO DA INFORMAÇÃO

Na Unidade 1, começamos nossos estudos objetivando compreender o que vem a ser "representação temática." Aprendemos a importância de representar os assuntos dos documentos adequadamente e que, para isso, deve-se utilizar um instrumento que padronize esses assuntos. Chamamos esses instrumentos de linguagens documentárias. Nas demais unidades, fomos aprendendo especificamente cada tipo de linguagem de representação temática existente, como usá-las e construí-las. Nesta Unidade, abordaremos um tipo de linguagem que, apesar de não ser voltada para o tratamento de documentos, guarda algumas semelhanças com as que já estudamos. Essa linguagem é o que a literatura chama de ontologias.

4.4 AFINAL, O QUE SÃO ONTOLOGIAS? DE ONDE VEM ESSA PALAVRA?

Figura 47 - Tirando dúvidas

Fonte: *Pixabay* (2017).[42]

[42] Disponível em: <https://pixabay.com/pt/ponto-de-interrogação-nota-duplicar-2110764>.

Se você fizer uma busca em dicionários, encontrará o termo "ontologia" definido como "teoria metafísica do ser"(FERREIRA, 1993). Em um estudo etimológico da palavra, você perceberá que *ontos* vem do grego e significa "ser". Já o sufixo *-logia* indica o estudo de algo. Sendo assim, temos "o estudo do ser". Mas o que isso tem a ver com nossos estudos? Os professores *Maurício Almeida* e *Marcelo Bax*, no que se refere ao termo "ontologia", explicam que:

> O termo original é a palavra aristotélica "categoria", que pode ser usada para classificar alguma coisa. Aristóteles apresenta categorias que servem de base para classificar qualquer entidade e introduz ainda o termo "differentia" para propriedades que distinguem diferentes espécies do mesmo gênero. A conhecida técnica de herança é o processo de mesclar differentia definindo categorias por gênero. (ALMEIDA; BAX, 2003, p.7).

Espécies do mesmo gênero? Herança? Se você pensou já ter lido algo parecido na seção 3, quando falamos de tesauros, você acertou! Falar de espécies de um mesmo gênero ou de herança de características é falar de classificação, e, como já aprendemos, um bom tesauro deve possuir uma estrutura classificatória.

Agora me responda: podemos dizer que tesauros e ontologias são a mesma coisa? A resposta é… Não! Apesar de guardarem semelhanças entre si, as ontologias possuem um nível de complexidade maior e isso se justifica por sua própria função.

Apesar de o nome "**ontologia**" ter sua origem no grego e se configurar como um empréstimo linguístico oriundo da filosofia por alguma semelhança, as ontologias que vamos estudar aqui têm outro significado.

Ontologia

No âmbito da Biblioteconomia, uma **ontologia** pode ser definida como um instrumento de representação que visa a padronizar os significados dos termos utilizados no tratamento de dados, com o objetivo de auxiliar os novos sistemas de recuperação na entrega de respostas inteligentes a seus usuários.

Figura 48 - Histórias são reveladoras...

Fonte: *Pixabay* (2015).[43]

4.4.1 Ontologias: um pouco de história

As ontologias surgem na área de Inteligência Artificial (IA) na década de 1990. Naquela época, a IA estava na fase chamada moderna, quando as pesquisas realizadas eram voltadas para a inserção de conhecimento nas máquinas, visando a capacitar o computador para o entendimento de ativi-

[43] Disponível em: <https://pixabay.com/pt/narrativa-história-sonho-dizer-794978>.

dades que somente um ser humano seria capaz de efetuar. Para isso, seria necessário que o conhecimento fosse representado em uma linguagem declarativa. Assim, surgiram as ontologias, com o intuito de descrever programas, por meio da definição de um conjunto de termos que representassem os domínios do saber e as tarefas a serem executadas por esses programas.

Multimídia

Para saber mais...

Segundo classificação do *Massachusetts Institute of Technology* (*MIT*), a IA é dividida em três fases, com os seguintes objetivos:

a) clássica: simular a inteligência humana através de métodos capazes de solucionar problemas gerais por meio da lógica;

b) romântica: simular a inteligência humana em situações predeterminadas, através de formalismos de representação de conhecimentos adaptados ao tipo de problema, visando à maior eficiência computacional;

c) moderna: simular o comportamento de um especialista humano ao resolver problemas em um domínio específico, através de um método de sistemas de regras, representação da incerteza, conexionismo.

Para saber mais, que tal assistir ao filme *A.I. Inteligência Artificial* (Figura 49), dirigido por *Steven Spielberg*?

Figura 49 - Cena do filme *A.I.: Inteligência Artificial*

Fonte: *Youtube* (20--?).

No entanto, foi a visão da ***web* semântica** que colocou as ontologias em voga. Trata-se de uma nova proposta de *web* que surgiu em 2001, a partir de um consórcio chamado *World Wide Web Consortium* (W3C), que reuniu os criadores da *web*, cientistas e empresas interessadas, com a finalidade de melhorar a *web* atual.

Web semântica

A ***web* semântica** tem como visão a ideia de ter dados na *web* definidos e ligados de tal maneira que possam ser usados por máquinas, não só com o objetivo de apresentação, mas por automação, integração e reúso de dados entre aplicativos (W3C, 2001).

A principal diferença entre a *web* atual e a *web* semântica está centrada no tipo de informação que elas armazenam e entregam ao usuário. O foco da primeira é o armazenamento de documentos, ao passo que o da segunda é o armazenamento de dados e seus relacionamentos, formando uma rede semântica de significados. Se considerarmos o fenômeno atual chamado *Big Data*, caracterizado pela enorme quantidade de dados gerados e disponibilizados na *web* via aplicativos, redes sociais, entre outros aparatos, podemos dizer que a visão da *web* semântica começa a se tornar uma realidade, já que hoje temos armazenados na *web* não apenas documentos, mas, principalmente, muitos dados já interligados.

Se, por um lado, o tipo de informação que a *web* armazena está mais diversificado, por outro, nossos mecanismos de busca continuam se comportando da mesma maneira que na década passada. Em termos práticos, hoje, quando fazemos uma pesquisa, consultamos os sistemas de buscas (ex.: *Google*, *Bing*) utilizando palavras-chave, e temos como resposta documentos e dados que guardam relação com essas palavras, mas nem sempre são exatamente aquilo que procuramos.

Para compreendermos as ontologias, precisamos entender o que realmente se espera com essa nova *web*. A fim de melhor ilustrar, vamos imaginar um buscador, onde, em vez de uma palavra-chave, possamos inserir questões e, em vez de documentos, possamos obter respostas.

A ideia que se encontra por trás da *web* semântica é a de que possamos entrar em um sistema de busca (*Google*, por exemplo) e, em vez de colocarmos termos e recebermos documentos, façamos uma pergunta específica, como: "Qual o carro mais indicado em viagem para seis pessoas, cada uma com duas malas, para rodar por Orlando?". E que, com base em uma ontologia e em uma série de regras de inferências, o sistema possa nos responder que um carro adequado, nesse caso, seria o *Chevrolet Suburban*, por exemplo.

Figura 50 - *Web* semântica: a resposta certa para cada pergunta

Fonte: *Pixaba*y (2015).[44]

[44] Primeira imagem: Bola de cristal. Disponível em: <https://pixabay.com/pt/médiuns-bola-de-cristalcartomante-1026092>; segunda imagem: *Google*. Disponível em: <https://pixabay.com/pt/google-seopesquisa-939112>.

Entendido o objetivo da *web* semântica, fica fácil compreendermos que, para um sistema nos oferecer repostas específicas, é necessário que a nossa linguagem e a linguagem utilizada pela máquina sejam compatíveis. A nossa linguagem é diferente da linguagem de máquina (por exemplo, o html), mas as **ontologias** podem ser um mecanismo de tradução dessas linguagens, por meio da padronização de seu vocabulário, conceitos, relações, regras de inferência, e da utilização de uma linguagem legível por máquina e inteligível por humanos.

Uma definição bastante usada para explicar o que são ontologias é a dada por *Borst* (1997, p.12): "uma ontologia é uma especificação formal e explícita de uma conceituação compartilhada." Nessa definição, "especificação explícita" diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas explicitamente definidos; "especificação formal" significa legível por computadores; "conceituação" diz respeito a um modelo abstrato de algum fenômeno do mundo real; e "compartilhado" significa que o conhecimento representado deve ser consensual.

Ontologia

No contexto da *web* semântica, uma **ontologia** é um documento ou um arquivo que define formalmente as relações entre os termos, sendo formada por uma taxonomia e regras de inferências (BERNERS-LEE; HENDLER; LASSILA, 2001). Além disso, ontologias são responsáveis pela padronização de significado, provendo a compatibilização de conceitos e diminuindo problemas relacionados à comunicação, não apenas entre usuário e sistema, mas também entre dois sistemas (interoperabilidade).

Explicativo

Figura 51 - Conceito e ontologias

Fonte: *Pixabay* (2017).[45]

De acordo com o *Dicionário Eletrônico Houaiss*, conceituação "é a ação ou efeito de dar ou criar um conceito; definição, formulação" (HOUAISS, 2007). Na área de Ciência da Informação, como já dito anteriormente, na Unidade 3, uma teoria importante e que é a base fundamental para a construção de tesauros é a Teoria do Conceito (DALHBERG, 1978). Ela nos ensina a pensar de forma abstrata e a formular conceitos sobre objetos individuais no mundo,

[45] Disponível em: <https://pixabay.com/pt/rede-sobre-mão-deixar-círculo-1987214>.

o que faz dela uma teoria importante a ser apreendida por quem pretende algum dia trabalhar construindo ontologias.

Você já leu um pouco sobre essa teoria na Unidade 3, mas pode saber mais, consultando este texto:

DALHBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, [S.l.], v. 7, n. 2, 1978. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/1680/1286>. Acesso em: 13 jan. 2014.

Outra definição, formulada no contexto da Ciência da Informação e que esclarece bastante sobre o que vêm a ser ontologias, é a de *Ramalho* (2006). Para esse autor, ontologias são:

> [...] instrumentos de representação do conhecimento definidos em uma linguagem formal e processável por máquina que possibilitam a descrição dos aspectos semânticos dos conteúdos informacionais, explicitando seus relacionamentos de modo detalhado a partir de restrições lógicas que possam ser processadas de forma automatizada, possibilitando, inclusive, relacionamentos baseados na Lógica. (RAMALHO, 2006, p. 59).

Com base nessas definições, podemos caracterizar as ontologias de acordo com seus componentes. É o que faremos na próxima seção.

4.4.2 Quais as características das ontologias?

De acordo com as definições estudadas, podemos destacar as seguintes características importantes sobre as ontologias:

a) possuem conceituações que devem ser compartilhadas;

b) devem possuir um vocabulário estruturado com termos, definições e relacionamentos, que devem expressar um acordo comum entre seus usuários;

c) devem ser descritas através de axiomas lógicos;

d) suas definições devem possuir uma forma-padrão, baseada em **axiomas**, que possam determinar a verdade das sentenças;

e) devem ser processáveis por máquina;

f) além de um vocabulário, devem possuir uma linguagem formal que propicie a interpretação desse vocabulário pelo computador.

Axioma

Proposição tão evidente que não precisa ser demonstrada. Na lógica tradicional, um axioma ou postulado é uma sentença ou proposição que não é provada ou demonstrada, e é considerada como óbvia ou como um consenso inicial necessário para a construção ou aceitação de uma teoria. Por essa razão, é aceito como verdade e serve como ponto inicial para a dedução e inferências de outras verdades (dependentes de teoria) (FERREIRA, 1993).

4.4.3 Componentes essenciais

De acordo com as características citadas anteriormente e com base em *Gruber* (1993), podemos afirmar que ontologias são compostas por:

a) conceitos: são ideias básicas sobre o que se pretende formalizar. Podem ser classes de objetos, métodos, planos, estratégias, processos etc., e devem estar organizados em uma taxonomia;

b) relações: representam a interação entre os conceitos de um domínio, por exemplo: **parte_de** (Fígado **parte de** Sistema Digestivo), é_um (Cavalo é_um Mamífero), **causa** (Fumo **causa** Câncer);

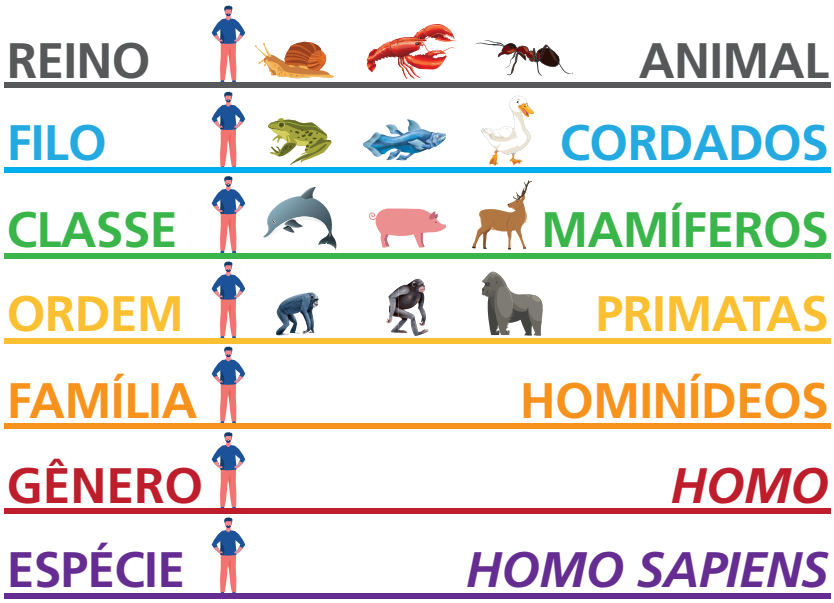
c) funções: são casos especiais de relações em que se identifica um elemento por meio do cálculo de uma função que considera vários elementos das ontologias. Ex.: **Categorizar-classe**; **atribuir-data**;

d) instâncias: são usadas para representar objetos específicos de determinados conceitos. Ex.: "*Biblioteca de Alexandria*" é uma instância do conceito "Biblioteca";

e) axiomas: são teoremas, isto é, sentenças verdadeiras que se declaram sobre relações que devem cumprir os elementos da ontologia. Ex.: se A e B são da classe C, então A não é subclasse de B.

Explicativo

Figura 52 - Vamos organizar o conhecimento?



Fonte: produção do próprio autor (2017).

Outro componente essencial das ontologias é a taxonomia. Os conceitos e as relações que compõem as ontologias devem ser organizados em uma estrutura classificatória. *Guarino*, um autor seminal da área de ontologias, diz que a taxonomia é a espinha dorsal de uma ontologia. Esse fato se explica porque a classificação é a base para a construção de argumentos lógicos. Assim, os conceitos, quando estruturados de maneira taxonômica, auxiliam os programas a processar logicamente o conhecimento explicitado nas ontologias.

Se liga!

A disciplina "Instrumentos de Representação Temática da Informação I" trata de instrumentos de classificação e, portanto, pode te ajudar a aprofundar seus conhecimentos sobre taxonomia. Fique de olho nela!

Multimídia

A solução para um problema

Figura 53 - Problema? Olha a solução!

Fonte: *Pixaba*y (2014).[46]

Vamos lá! Imaginemos que você tenha sido incumbido de organizar determinada área do conhecimento! O problema: por onde começar a classificação dos assuntos? Por que não pensar em fazer uma taxonomia? Ela é fundamental!

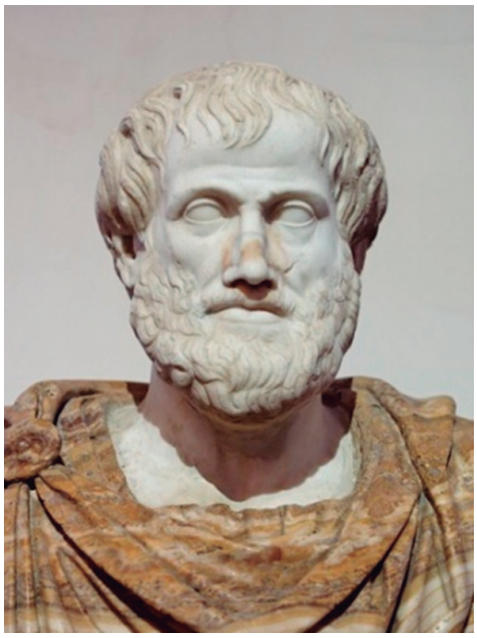
Pois então, acesse o *link*: <http://www.conexaorio.com/biti/taxonomianaweb.pdf>. Lá, você encontrará o trabalho *Taxonomia e a web*, *construção e uso*, de *Hagar Espanha Gomes* (2014), que abrange o assunto de forma clara e elucidativa, desde o significado da palavra taxonomia e a árvore de *Porfírio* – exemplo clássico de tal sistema (século III d.C.) –, até outros aspectos teóricos e práticos dessa ferramenta. Confira!

[46] Disponível em: <https://pixabay.com/pt/mãos-oferta-resposta-consultoria-460872>.

Explicativo

Um pouco sobre lógica...

Figura 54 - *Aristóteles*: o primeiro autor sobre lógica



Fonte: *Wikipédia* (2006).[47]

A lógica é uma disciplina que estuda as formas de argumentação e é discutida principalmente nas áreas de Filosofia (metafísica, ontologia, epistemologia e ética), Matemática (estudam-se as formas válidas de inferência de uma linguagem formal) e Ciência da Computação. Existem diversos tipos de lógica, cada uma é usada com uma finalidade. No escopo das ontologias, ainda há muita discussão sobre qual a lógica mais adequada para criar as regras de inferência necessárias para tornar a *web* semântica possível. Aqui temos uma breve explicação sobre as lógicas mais citadas na literatura sobre ontologias. No entanto, se você quiser se aprofundar mais ou conhecer outros tipos de lógica, deve buscar outros textos.

A lógica formal – também chamada de lógica simbólica – preocupa-se, basicamente, com a estrutura do raciocínio. Ela trata da relação entre as premissas e a conclusão, ou seja, lida com a relação entre conceitos e fornece um meio para compor provas sobre as declarações. Os conceitos de que se utiliza são definidos rigorosamente e as orações são transformadas em notações simbólicas precisas, compactas e sem ambiguidades. Para saber mais, leia o artigo do Prof. *Mundim* (2002), disponível em: <http://periodicos.pucminas.br/index.php/economiaegestao/article/view/113/104>.

Pode-se dizer que a lógica de primeira ordem – também chamada de cálculo de predicados de primeira ordem – consiste em um conjunto de axiomas (geralmente finito ou recursivamente enumerável) e de sentenças dedutíveis a partir deles. Os valores das variá-

[47] Disponível em: <https://en.wikipedia.org/wiki/Aristotle#/media/File:Aristotle_Altemps_Inv8575.jpg>.

veis são tirados de um mesmo universo de discurso, onde é possível tratar diferentes classes de objetos e criar generalizações a partir de fatos que valem para um indivíduo escolhido dentro desse universo predeterminado.

A lógica descritiva – também chamada de lógica de descrição ou terminológica – surgiu como uma evolução das redes semânticas e *frames*, com o intuito de resolver problemas e limitações do uso das outras lógicas no que tange a sua capacidade de ser computável. É o conjunto de formalismos que representam o conhecimento de um domínio a partir da descrição conceitual das propriedades dos objetos e indivíduos que fazem parte dele. Usada na área de inteligência artificial e também para fins de modelagem, essa lógica vem manifestando fundamental importância no provimento de um formalismo lógico para ontologias e *web* semântica, já que é a base da *Ontology Web Language* (*OWL*), uma linguagem própria para descrição de ontologias.

4.4.4 Atividade

A partir do que foi aprendido até agora, descreva uma ou duas situações hipotéticas em que um sistema com base em ontologia poderia ser útil para responder a uma questão específica.

Resposta comentada

Uma resposta possível seria um sistema de informações de um restaurante que, com base em um conjunto de conceitos sobre os pratos servidos, particularidades sobre os vinhos disponíveis e, ainda, uma série de inferências, pudesse sugerir ao usuário/cliente as opções de vinhos que harmonizariam com seu prato.

Outra opção seria uma ontologia de "material de aprendizagem" (item do conteúdo programático de uma unidade básica de determinada disciplina) para uma plataforma de ensino a distância: *Cursos On-Line* (COL). Com base em um conjunto de informações sobre esse material, além de inferências, a ontologia deveria prover respostas sobre os componentes desses materiais (disciplinas ensinadas; módulos existentes; profissionais responsáveis pela criação dos módulos; materiais de aprendizagem que comporiam a plataforma; pré-requisitos de determinado material; existência de materiais similares entre si; tipos, formatos e características dos objetos que comporiam os materiais de aprendizagem), permitindo a

alunos e professores a recuperação de informações de componentes de materiais de aprendizagem e possibilitando o acesso a esses materiais de forma rápida e inteligente.

Agora que você já entendeu o que são ontologias e como elas são formadas, podemos começar a enumerar suas aplicações e tipos. Vamos lá!

4.5 É PARA COMER OU PASSAR NO CABELO? PARA QUE SERVEM AS ONTOLOGIAS?

Figura 55 - Funcionalidade das ontologias

Fonte: *Freepik* (2010).[48]

Começamos perguntando se as ontologias são de comer ou de passar no cabelo com a intenção de ilustrar a confusão que existe acerca desse conceito. É muito comum as pessoas acharem que vão resolver o mundo construindo ontologias.

O estudo das ontologias é muito instigante, e, como já vimos, a ideia não é totalmente nova; o que faz das ontologias um conceito novo são seu uso e suas possibilidades de aplicação. Para sistemas de recuperação de informação de acervos documentais, os tesauros continuam cumprindo seu papel muito bem.

[48] Disponível em: <http://br.freepik.com/fotos-gratis/duvida_34773.htm>.

Apesar de os tesauros guardarem suas relações de semelhança com as ontologias, esses dois instrumentos possuem diferenças, principalmente quanto a sua aplicação. Como visto na Unidade 3, um tesauro é um instrumento de controle terminológico e sua aplicação é para tratamento e recuperação de informação, ao passo que uma ontologia é um instrumento mais complexo, pois sua finalidade vai além de ser um instrumento de recuperação da informação.

A principal finalidade de uma ontologia é auxiliar no processamento de conhecimento por máquina. No entanto, como o termo entrou na moda, é possível ver ontologias sendo utilizadas com função de tesauros e tesauros sendo chamados de ontologias. A relação que ambos os instrumentos guardam entre si se deve ao fato de que um pode ser extensão/complemento do outro, isto é, os tesauros podem servir de base para a construção de ontologias, e estas podem conter tesauros.

Outra relação é que a base teórica utilizada para a construção de tesauros conceituais (Teoria da Classificação Facetada de *Ranganathan*, Teoria do Conceito de *Dalhberg* e Teoria Geral da Terminologia de *Wüster*) pode oferecer subsídios para a construção de ontologias também.

Como já vimos durante o estudo da história das ontologias, esses instrumentos de padronização terminológica possuem funções específicas na área de IA e na *web* semântica. No entanto, na literatura, é possível encontrar outras funções para as ontologias em áreas específicas.

Quadro 12 - Aplicação de ontologias em diversas áreas

(continua)

ÁREA	APLICAÇÃO
Comunicação	Reduzem a confusão terminológica e conceitual através da unificação da linguagem de uma organização, permitindo que várias pessoas tenham o mesmo entendimento sobre determinado assunto.
Reúso	Auxiliam na importação e exportação de classes de objetos digitais em diferentes domínios. Para isso, as ontologias devem ser construídas de forma padronizada.
Confiabilidade	Conferem consistência ao *software,* servindo como um manual de verificação do modelo utilizado em seu desenvolvimento. Nesse caso, são usadas ontologias informais, em que um vocabulário em linguagem natural é acoplado ao sistema para ser consultado pelos programadores. Além disso, as ontologias também podem propiciar sistemas mais confiáveis, à medida que podem facilitar a integração entre eles, já que tornam explícitas as várias decisões tomadas pelos diferentes desenvolvedores de um mesmo sistema.
Classificação para ação	Permitem que as tarefas executadas sejam classificadas de acordo com a ordem de prioridade, além de padronizar essas tarefas e sua linguagem de execução para compartilhamento entre *softwares* distintos.
Interoperabilidade	Servem para integrar e permitir troca de informação entre sistemas de maneira consistente. A interoperabilidade pode ser entre sistemas iguais que precisam trocar dados ou entre sistemas de naturezas diferentes.

Quadro 12 - Aplicação de ontologias em diversas áreas

(conclusão)

Especificação de sistemas	Servem como regras que podem variar de acordo com o nível de formalidade e de automação do sistema, facilitando o processo de identificação de seus requisitos e a compreensão das relações entre seus componentes.
Representação e recuperação de informação	Proveem uma base para a representação do significado do texto, melhorando a precisão das informações recuperadas.
Usabilidade	Proveem interfaces corporativas e auxiliar na navegação em bases de informação, tornando o sistema mais amigável, à medida que o significado dos termos utilizados são padronizados. Além disso, as ontologias, por conterem em sua estrutura uma taxonomia, podem auxiliar, através de mecanismos de classificação, a arquitetura da informação, facilitando a interação entre usuário e sistema.
Processamento de textos em linguagem natural	Proveem eficácia na extração de conhecimento dos textos em linguagem natural, pois padronizam os significados dos termos utilizados no processo.
Engenharia de sistemas	As ontologias são usadas para apoiar o desenvolvimento de projetos de sistemas de *softwares*.

Fonte: produção do próprio autor, baseado em *Sales* (2006).

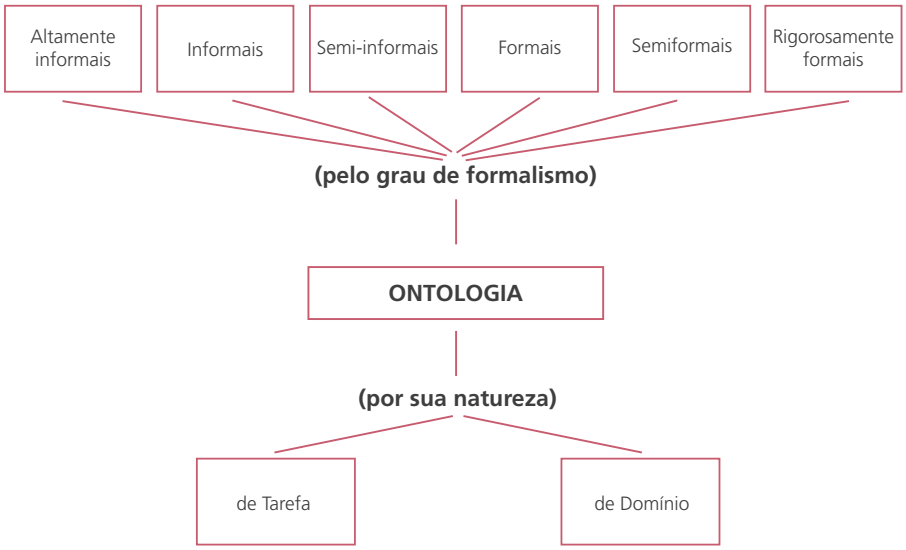
Apesar de todas essas aplicações da ontologia estarem mencionadas na literatura, se formos analisar cada caso na prática, em alguns deles verificaremos que também é possível a aplicação do tesauro, como nas áreas de comunicação, de representação e recuperação da informação e de usabilidade, por exemplo. Em alguns contextos, no entanto, a instituição prefere usar uma ontologia menos formal no lugar do tesauro ou, ainda, usar uma ontologia formal e compartilhar seus componentes para a solução de problemas distintos. Nesse caso, uma ontologia pode servir para melhorar a comunicação na instituição e a apresentação do *site*, ou também para processar conhecimento em sistemas com aplicações específicas.

4.6 ONTOLOGIAS: CONHECENDO SEUS TIPOS

Existem diversos tipos de ontologias. *Sales* (2006) propôs uma classificação para os tipos encontrados na literatura, considerando o grau de formalismo e a natureza delas.

De acordo com a natureza, as ontologias podem ser classificadas como Ontologias de Domínio e Ontologias de Tarefa. Por outro lado, de acordo com o grau de formalismo, elas podem ser altamente informais; informais; linguísticas ou terminológicas (semi-informais); formais; semi-formais e rigorosamente formais. Confira no Quadro13, a seguir:

Quadro 13 - Tipos de ontologias



Fonte: *Sales* (2006).

A seguir, estão listados os diversos tipos de ontologias, a saber:

a) altamente informal: expressa livremente, em linguagem natural, sem nenhum controle. Podemos citar como exemplo uma folksonomia;

Curiosidade

Folksonomia: que bicho é esse?

Figura 56 - Folksonomia: seus termos são chamados de *tags*

Fonte: produção do próprio autor (2017).

O termo "folksonomia" foi cunhado por *Thomas Van der Wal*, em analogia às taxonomias. Como o prefixo de origem inglesa ("folks", que significa pessoas) aponta, a folksonomia é construída por e para pessoas. Até aí, nada de muito diferente das taxonomias. O ponto central está na palavra "colaboração." Enquanto uma taxonomia é construída por uma ou mais pessoas para ser usada por muitos, a folksonomia é construída por todos e para todos.

Assim, trata-se de um vocabulário que usa o linguajar da comunidade que a constrói e utiliza. Os termos que constituem a folksonomia são chamados de *tags* (etiquetas) e são dados no momento da indexação da informação (em geral, objetos digitais). Como todo vocabulário construído, a folksonomia tem vantagens e desvantagens. Como desvantagem, podemos citar a dificuldade de controlar o sentido de cada *tag* usada; a vantagem está em usar como *tag* a palavra que realmente o usuário emprega para buscar a informação. Uma folksonomia pode ser um bom ponto de partida para iniciar a construção de um vocabulário controlado simples ou até de ontologias. Você aprenderá mais sobre esse assunto na Unidade 5 (5.7.3).

b) informal: também chamada de Ontologia Terminológica, especifica os termos que serão usados para representar o conhecimento em um domínio com certo controle. Podemos citar como exemplo os léxicos, as listas de vocábulos autorizados de uma biblioteca, entre outros;

c) semi-informal: expressa em linguagem natural de forma restrita e estruturada. Podemos citar como exemplo um tesauro, um esquema de classificação ou uma lista de cabeçalhos de assunto;

d) formal: utiliza um vocabulário controlado em linguagem humana e, a partir dele, constrói seu vocabulário em linguagem de máquina, com base em alguma lógica (descritiva, de primeira ordem, *fuzzy*, entre outras). Esse tipo de ontologia pode ser compreensível por humanos e por máquina;

e) semiformal: expressa diretamente em linguagem artificial definida formalmente, sem explicitar o vocabulário humano utilizado e sem se preocupar com uma lógica muito rígida;

f) rigorosamente formal: expressa por termos definidos com semântica formal, teoremas e provas. Construída para ser compreendida pela máquina sem ter a preocupação com o entendimento humano;

g) de Domínio: fornece vocabulários sobre conceitos, relacionamentos, atividades e regras que envolvem um domínio (ex.: medicina, automóveis e outros). Esse tipo de ontologia é próprio para representação e recuperação das informações em áreas genéricas e possui a particularidade de poder ser reusado em domínios relacionados e/ou para outros fins;

h) de Tarefa: fornece um vocabulário sistematizado de termos, especificando atividades genéricas, formado por ações que podem ou não estar no mesmo domínio (ex.: diagnóstico, vendas). Sua finalidade é possibilitar a execução de tarefas pela máquina.

Um tipo de ontologia que não entrou no Quadro 13, mas que também merece atenção, é a Ontologia de Topo. Trata-se de uma metaontologia, construída com termos de propósitos gerais para fundamentar a construção de ontologias de Domínio ou de Tarefa. Ela define conceitos mais básicos, que podem servir para a compreensão e o funcionamento de qualquer outro tipo de ontologia. A Ontologia de Topo é interessante porque possibilita a modelagem do sistema, permitindo que ontologias diferentes sejam construídas sob uma mesma base conceitual, reusadas ou integradas a outras.

4.6.1 Atividade

Você já teve a oportunidade de ver o filme *Operação Big Hero*, da Disney? É um filme voltado para o público infantil, mas existem alguns detalhes que só é capaz de entender quem compreende para que servem as ontologias.

Tente assistir. Você vai se divertir com o personagem *Baymax* – um robô com tamanho XXG, criado pelo personagem *Tadashi* para ser um atendente médico pessoal. Entre suas funcionalidades, está a capacidade de escanear uma pessoa, detectar e tratar seus níveis de estresse, dor e humor. Ele pode revolucionar a medicina moderna. Mas, para *Hiro* – irmão de *Tadashi* –, o robô acaba se transformando em muito mais do que isso: é um verdadeiro herói e seu melhor amigo. Depois de uma reprogramação, que inclui um pulso de foguete superforça e propulsores que lhe permitem voar, *Baymax* se torna um dos heróis da *Operação Big Hero*. O filme não diz, mas, para criá-lo e, depois, reprogramá-lo, *Tadashi* e *Hiro* precisaram de muita lógica e, também, de ontologias.

A proposta desta Atividade é que você assista ao filme e aponte as partes em que identifica o uso de ontologias para a criação de *Baymax*. Indique, também, a parte em que pode ter faltado uma ontologia em sua reprogramação.

_____

_____

_____

Resposta comentada

Se você prestou atenção ao filme, deve ter percebido que *Baymax* foi programado usando algumas Ontologias de Domínio (como Ontologias de Tipo de Doenças, de Medicamentos, de Sintomas, de Cuidados Paliativos), além de uma Ontologia de Tarefa – que lhe permitia saber exatamente o que fazer antes e depois de determinada descoberta, e, ainda, uma Ontologia de Topo, na qual devem ter sido definidos os conceitos que dariam o *start* do sistema (*liga* = "ui", "ai" ou qualquer outra onomatopeia de dor e *desliga* = expressão "eu estou bem").

Em outras situações, porém, você deve ter identificado ruídos na comunicação entre *Hiro* e *Baymax*, o que gerou cenas engraçadas ocasionadas pela eventual ausência de ontologias.

4.7 COMO ELABORAR ONTOLOGIAS? METODOLOGIAS, LINGUAGENS E FERRAMENTAS

Com o que já aprendemos até aqui, podemos perceber que as ontologias podem servir a diversas áreas, em cada uma com um objetivo diferente. *Gruber* (1993) afirma que "o importante é para o que a ontologia serve", ou seja, sua função e os objetivos para os quais ela vai se dirigir, pois são esses que irão guiar sua elaboração.

4.7.1 Metodologias para a construção de ontologias

Existe uma infinidade de metodologias já propostas na literatura para a elaboração de ontologias. Em 2003, *Almeida* e *Bax* apresentaram um levantamento sobre as metodologias existentes até então, e as classificaram em: metodologias para construção de ontologias e metodologias para construção de ontologias em grupo. Transcrevemos ambas nos Quadros 14 e 15, a seguir:

Quadro 14 - Metodologias para a construção de ontologias

(continua)

Metodologia	Breve descrição
Cyc	Codifica manualmente o conhecimento implícito e explícito das diferentes fontes e, quando já se tem conhecimento suficiente na ontologia, um novo consenso pode ser obtido por ferramentas que utilizam linguagem natural (LENAT; GUHA, 1990).
USCHOLD e KING	Identifica o propósito, os conceitos e relacionamentos entre os conceitos, além dos termos utilizados para codificar a ontologia e, em seguida, documentá-la (USCHOLD; KING, 1996).
GRUNINGER e FOX	Método formal que identifica cenários para uso da ontologia. Utiliza questões em linguagem natural para determinação do escopo da ontologia, executa a extração sobre os principais conceitos, propriedades, relações e axiomas definidos em Prolog (programação lógica) (GRUNINGER; FOX, 1995).
KACTUS	Método recursivo que consiste em uma proposta inicial para uma base de conhecimento; quando é necessária uma nova base em domínio similar, generaliza-se a primeira base em uma ontologia adaptada a ambas as aplicações; quanto mais aplicações, mais genérica a ontologia (BERNARAS; LARESGOITI; CORERA, 1996).

Quadro 14 - Metodologias para a construção de ontologias

(conclusão)

Metodologia	Breve descrição
Methontology	Constrói uma ontologia por reengenharia sobre outra, utilizando o conhecimento do domínio; as atividades principais são: especificação, conceitualização, formalização, implementação e manutenção (FERNÁNDEZ-LÓPEZ et al., 1999).
Sensus	Constrói ontologias a partir de outras, identificando os termos relevantes para o domínio e ligando-os à ontologia mais abrangente (Sensus, com 50 mil conceitos); um algoritmo monta a estrutura hierárquica do domínio (SWARTOUT et al., 1996).
On-to-knowledge	Auxilia a administração de conceitos em organizações, identificando metas para as ferramentas de gestão do conhecimento e utilizando cenários e contribuições dos provedores/clientes de informação da organização (STAAB et al., 2001).

Fonte: *Almeida* e *Bax* (2003).

Quadro 15 - Metodologias para a construção de ontologias em grupo

Metodologia	Breve descrição
CO4	Permite discussão sobre conhecimento introduzido em bases de conhecimento compartilhadas, que correspondem à ontologia, já que deve haver um consenso sobre o conhecimento ali representado. Quando uma mudança é proposta, os usuários são notificados e podem aceitá-la ou não (EUZENAT, 1996).
(KA)	Modela formas de aquisição do conhecimento usando ontologias desenvolvidas em conjunto por pessoas em diferentes locais, mas que utilizam o mesmo padrão; a comunicação e coordenação são feitas via agentes inteligentes (KIETZ; MAEDCHE; VOLZ, 2000).

Fonte: *Almeida* e *Bax* (2003).

Apesar de esse levantamento ser de 2003, o reutilizamos aqui para mostrar que, já naquele ano, existiam diversas metodologias para elaboração de ontologias. É certo que, desde então, outras metodologias foram propostas – o que aumentou ainda mais a variedade delas. No entanto, é importante sabermos que essa variedade e a falta de integração entre as metodologias para a construção de ontologias pode ser um problema e, por isso, diversos estudos vêm sendo realizados no sentido de unificar as metodologias existentes e compatibilizar as ontologias já construídas a partir delas. Você pode fazer uma busca sobre esse assunto no *Google Acadêmico* ou em outra base de dados e se informar mais.

Explicativo

Bibliotecas de ontologias

Figura 57 - Biblioteca

Fonte: *Pixabay* (2012).[49]

O reúso é um processo cada vez mais comum no escopo do desenvolvimento de ontologias. Uma vez que o conceito já foi trabalhado segundo os padrões da *web* semântica, não é mais necessário que ele seja tratado novamente dentro de outra ontologia. Devido a isso, sistemas de gerenciamento de ontologias vêm sendo disponibilizados, de modo a reunir e apoiar o compartilhamento de ontologias para seu posterior reúso. Esses sistemas são chamados de "bibliotecas de ontologias".

De acordo com *Breitman* (2005), as características desejáveis para as bibliotecas de ontologias são:

a) oferecer infraestrutura para armazenamento, busca e recuperação de ontologias;

b) ser de fácil acesso;

c) oferecer suporte à reutilização de ontologias existentes;

d) suportar vários padrões de linguagens;

e) oferecer mecanismo de tradução das ontologias para as linguagens recomendadas;

f) oferecer suporte para a edição de ontologias.

[49] Disponível em: <https://pixabay.com/pt/menina-biblioteca-educação-1721436>.

Multimídia

As bibliotecas de ontologias também são chamadas de repositórios de ontologias. Se você tiver facilidade em compreender a língua inglesa, listamos, a seguir, alguns exemplos, com seu endereço na *web*. Que tal navegar pelos *site*s e conhecer o conteúdo das bibliotecas?

a) *Ontolingua*: <www.ksl.stanford.edu/software/ontolingua>;

b) *webOnto*: <http://projects.kmi.open.ac.uk/webonto>;

c) *Biblioteca DAML*: <www.daml.org/ontologies>;

d) *Ontology Server*: <http://www.starlab.vub.ac.be/research/dogma/OntologyServer.htm>;

e) *Schemaweb*: <http://www.schemaweb.info>;

f) *Protégé Ontology Library*: <http://Protégéwiki.stanford.edu/wiki/Protégé_Ontology_Library>.

4.7.2 OWL: uma linguagem para a representação de ontologias

Figura 58 - Ontologia descrita com *OWL*

```
1  <?xml version = '1.0' encoding = 'ISO-8859-1' ?>
2  <rdf:RDF

3    <owl:Ontology rdf:about = 'Museum'/>

4    <owl:Class rdf:ID = 'Painter'>
5      <rdfs:label>Painter</rdfs:label>
6      <rdfs:subClassOf rdf:resource = '#Artist'/>
7      <rdfs:subClassOf>
8        <owl:Restriction>
9          <owl:onProperty rdf:resource = '#Painter.paints'/>
10         <owl:minCardinality  rdf:datatype = '&xsd#nonNegativeInteger'>0</owl:minCardinality >
11       </owl:Restriction>
12     </rdfs:subClassOf>
13   </owl:Class>

14   <Painter rdf:ID = 'PabloPicasso'>
15     <Artist.lastName rdf:datatype = '&xsd#string'>Picasso</Artist.lastName>
16     <Artist.firstName rdf:datatype = '&xsd#string'>Pablo</Artist.firstName>
17     <Painter.paints rdf:resource = '#Guernica'/>
18     <Painter.paints rdf:resource = '#Autoportrait'/>
19   </Painter>

20 </rdf:RDF>
```

Fonte: *Eclipse* (2017).[50]

[50] Disponível em: <www.eclipse.org>.

O consórcio *W3C* recomenda o uso de uma linguagem específica para definir e instanciar ontologias na *web*. O nome dessa linguagem é **OWL** e sua recomendação tem sido feita por se tratar de uma tecnologia importante para a futura implementação da *web* semântica. Sua recomendação pelo *W3C* faz dela um padrão que todos devem seguir para a construção de ontologias na *web*.

A OWL foi baseada nas linguagens OIL e DAML+OIL e XML, para o uso por aplicações que precisam processar o conteúdo da informação ao invés de apenas apresentá-la aos humanos. Essa linguagem foi desenvolvida para padronizar o processamento de conteúdo semântico da informação na *web*. Em relação às linguagens já existentes para descrição de conteúdos na *web*, como **XML**, **RDF** e **RDFS**, ela tem a vantagem de expressar a semântica dos recursos descritos, sendo considerada uma evolução dessas linguagens em termos de habilidade para tornar o conteúdo semântico da *web* compreensível por máquinas. Por ser baseada em XML, a informação pode ser facilmente compartilhada entre diferentes tipos de computadores, usando diferentes sistemas operacionais e linguagens de programação. Sua compreensão por humanos é possível apenas por meio de *softwares*, como o *Protégé*, que é um sistema próprio para construção de ontologias em OWL.

Uma característica interessante da linguagem OWL é que ela permite incluir descrições de classes/conceitos e suas respectivas propriedades e relacionamentos.

Atenção

É importante fazer aqui uma distinção. O que a área de Ciência da Computação chama de classes é o mesmo que nós, da Biblioteconomia e Ciência da Informação, entendemos como conceitos. Você aprendeu sobre o conceito na Unidade 3, em que estudou sobre tesauro. É importante compreendermos também que classe, para a Ciência da Informação, tem outro significado: trata-se de um conjunto de conceitos. Então, atenção para essas definições!

OWL

***O**ntology **W**eb **L**anguage* (Linguagem *web* para Ontologia). Essa linguagem, baseada em XML, é recomendada pelo *W3C* para representar ontologias por meio de descrições de classes, suas respectivas propriedades e relacionamentos. Foi desenvolvida para ser usada por sistemas *web*, propiciando a interpretação de seu conteúdo informacional não apenas por humanos, mas também pela máquina. Diferencia-se de outras linguagens, como o RDF, por fornecer vocabulário com uma semântica formal, sendo por isso considerada uma evolução dessas linguagens em termos de sua habilidade de representar conteúdo semântico na *web* e, portanto, ser mais efetiva que as demais para a implementação da *web* semântica.

XML

Sigla para *E**x**tensible **M**arkup **L**anguage* (Linguagem Extensível de Marcação Genérica). O XML é uma linguagem padrão para descrição de recursos na *web*. Sua principal característica é possibilitar a criação de uma infraestrutura única para diversas linguagens e seu maior objetivo é facilitar o compartilhamento de informações na *web*. Entre linguagens baseadas em XML, incluem-se: XHTML (formato para páginas *web*), SDMX, SMIL, MathML (formato para expressões matemáticas), NCL, XBRL, XSIL e SVG (formato gráfico vetorial).

RDF

Sigla para ***R**esource **D**escription **F**ramework* (Estrutura para Descrição de Recursos). Modelo para representação da informação na *web* recomendado pelo consórcio *W3C*, para descrever, por meio da linguagem XML, três elementos: sujeito-propriedade-objeto ou, também, como citado na literatura sobre o assunto, recurso-propriedade-valor. Ex.: a página <http://documentoampliado.blogspot.com.br/>, que tem como criadora *Luana Sales*. O endereço é o recurso ou sujeito; "criador" é uma propriedade desse recurso e "*Luana Sales*" é o valor dado a essa propriedade.

RDFS

Sigla para ***R**esource **D**escription **F**ramework **S**chema* (Esquema para Estrutura para Descrição de Recursos). Espécie de dicionário onde são definidos os termos que serão utilizados nas declarações RDF. É responsável por prover mecanismos para declaração de propriedades dos recursos RDF. Permite: definir os termos que foram usados nas declarações RDF, atribuindo-lhes significados específicos; modelar metadados com uma representação explícita de sua semântica; especificar restrições de tipos de dados para os sujeitos e objetos das triplas RDF.

4.7.3 Uma boa notícia: ferramentas para a construção de ontologias

Figura 59 - Extra! Extra! *Software* para construir ontologias!

Fonte: *Pixabay* (2013).[51]

Uma boa notícia é que atualmente existem inúmeros *softwares* que podem ajudar você a desenvolver ontologias, tornando o processo mais fácil. Muitas dessas ferramentas são gratuitas ou possuem uma versão demo, que você pode baixar para fazer alguns testes. Que tal experimentar? Minha preferida é o *Protégé*, mas você pode tentar qualquer uma das descritas no Quadro 16, ou ainda descobrir uma nova.

Quadro 16 - Ferramentas para a construção de ontologias

(continua)

Ferramentas	Breve descrição
CODE4 (*Conceptually Oriented Description Environment*)	Ferramenta de propósito geral que possui diferentes modos de herança e inferência, uma interface gráfica de fácil uso, um modo de hipertexto para navegação e utilitários para leitura de documentos e gerenciamento léxico (SKUCE, 1995).
VOID	Ambiente para navegação, edição e gerenciamento de ontologias. Por meio de simulações, possibilita o estudo de questões teóricas, como organização de bibliotecas de ontologias e tradução entre diferentes formalismos (SCHREIBER; TERPSTRA; SISYPHUS, 1995).
IKARUS (*Intelligent Knowledge Acquisition and Retrival Universal System*)	Explora as capacidades cooperativas do ambiente *web*. Utiliza uma representação hierárquica gráfica que permite herança múltipla. As declarações que contêm a informação são representadas como predicados com sintaxe e semântica definidos ou como fragmentos sem estrutura (SKUCE, 1996).

[51] Disponível em: <https://pixabay.com/pt/jornaleiro-menino-pessoa-notícias-146954>.

Quadro 16 - Ferramentas para a construção de ontologias

(continua)

Ferramentas	Breve descrição
Ontolingua	Conjunto de serviços que possibilitam a construção de ontologias compartilhadas entre grupos. Permite acesso a uma biblioteca de ontologias, tradutores para linguagens e um editor para criar e navegar pela ontologia. Editores remotos podem editar ontologias usando protocolos. (FARQUHAR; FIKES; RICE, 1996).
Ontosaurus	Consiste em um servidor de ontologias que usa uma linguagem para representação do conhecimento e um servidor de navegação por ontologias que cria páginas HTML dinamicamente e apresenta a hierarquia da ontologia (SWARTOUT et al., 1996).
GKB-Editor (*Generic Knowledge Base Editor*)	Ferramenta para navegação e edição de ontologias por meio de sistemas de representação baseados em *frames*. Oferece interface gráfica, em que os usuários podem editar diretamente a base de conhecimento e selecionar a parte que é de seu interesse (PALEY; KARP, 1997).
JOE (*Java Ontology Editor*)	Ferramenta para construção e visualização de ontologias. Proporciona gerenciamento do conhecimento em ambientes abertos, heterogêneos e com diversos usuários. As ontologias podem ser visualizadas como um diagrama entidade-relacionamento, como o gerenciador de arquivos do MS Windows ou como uma estrutura em árvore (MAHALINGAM; HUHNS, 1997).
APECKS (*Adaptive Presentation Environment for Collaborative Knowledge Structuring*)	É um serviço de ontologias que permite trabalho cooperativo mediante a criação de ontologias pessoais pelos usuários. Essas ontologias podem ser comparadas com outras, e é possível a discussão sobre as diferenças e similaridades entre elas (TENNISON; SHADBOLT, 1998).
OilEd	É um editor de ontologias de código aberto que permite construí-las utilizando linguagem OIL. Não é um ambiente completo para desenvolvimento de ontologias. A verificação da consistência e a classificação automática da ontologia podem ser executadas pela ferramenta FaCT (HORROCKS; SATTLER; TOBIES, 1999).
OntoEdit	É um ambiente gráfico para edição de ontologias que permite que elas sejam inspecionadas, navegadas, codificadas e alteradas. O modelo conceitual é armazenado usando-se um modelo de ontologia que pode ser mapeado em diferentes linguagens de representação. As ontologias são armazenadas em bancos relacionais e podem ser implementadas em XML, FLogic, RDF(S) e DAML+OIL (MAEDCHE et al., 2000).
COM (*Ontological Constraints Manager*)	Ferramenta para verificar a consistência de ontologias em relação a axiomas ontológicos. É composta por duas ferramentas de edição que possibilitam verificar a ocorrência de conflitos (KALFOGLOU et al., 2001).

Quadro 16 - Ferramentas para a construção de ontologias

(conclusão)

Ferramentas	Breve descrição
Protegé 2000	Ambiente interativo para projeto de ontologias, de código aberto, que oferece uma interface gráfica para edição de ontologias e uma arquitetura para a criação de ferramentas baseadas em conhecimento. A arquitetura é modulada e permite a inserção de novos recursos (NOY; FERGSON; MUSEN, 2000).
webODE	Ambiente para engenharia ontológica que dá suporte à maioria das atividades de desenvolvimento de ontologias. A integração com outros sistemas é possível, importando e exportando ontologias de linguagens de marcação (ARPÍREZ et al., 2001).
webOnto	Ferramenta que possibilita a navegação, criação e edição de ontologias, representadas na linguagem de modelagem OCML. Permite o gerenciamento de ontologias por interface gráfica, inspeção de elementos, verificação da consistência da herança e trabalho cooperativo. Possui uma biblioteca com mais de cem ontologias (DOMINGUE, 2001).

Fonte: *Almeida* e *Bax* (2003).

4.8 E AGORA, O QUE FAZER? DICAS PARA INICIAR A CONSTRUÇÃO DE UMA ONTOLOGIA

Agora que você sabe que existe uma infinidade de metodologias para construção de ontologias, e também ferramentas que podem auxiliá-lo em sua elaboração, deve estar se perguntando por onde começar. Como a finalidade principal das ontologias é oferecer respostas inteligentes a questões de busca, uma forma interessante de iniciar sua construção é levantar as perguntas que o sistema que utilizará a ontologia pretende responder. Essas perguntas são chamadas de "questões de competências". Além disso, nesse momento, deve-se definir também que tarefa o sistema deverá resolver. Essa primeira etapa o ajudará a identificar o quão geral ou específica a ontologia será.

A partir dessas questões, você pode selecionar as classes/conceitos que precisarão ser definidas e as relações entre elas. Em geral, as relações nas ontologias são diferentes das relações nas demais linguagens. En-

quanto as linguagens de indexação tradicionais fazem suas relações entre pares de conceitos (por exemplo: gênero-espécie, parte-todo, material--produto, entre outras), as relações nas ontologias aparecem na forma de verbo (exemplo: "*is_ a*", "*part_of*", "*product_of*", entre outras). Nesse momento, também é interessante já estabelecer a taxonomia, criando categorias e hierarquias de classes e subclasses.

Com esses passos e a escolha do *software* adequado, você já é capaz de criar uma ontologia.

4.8.1 Atividade

Um sistema está sendo construído para responder à seguinte questão de competência: "Qual o melhor tratamento de câncer para uma pessoa com 40 anos?"

Para responder-lhe, outras questões precisarão ser identificadas e respondidas corretamente pelo sistema. Algumas delas são:

Quais os tipos de câncer existentes?

Em que órgão está localizado o câncer?

Quais os tratamentos existentes?

Quais as possibilidades de cura de cada tipo de tratamento?

Quais as contraindicações de cada tipo de tratamento?

Quais as contraindicações para cada faixa etária?

A partir dessas questões, que termos precisarão ser definidos e organizados no escopo da ontologia?

_____

_____

_____

_____

Resposta comentada

Você deve ter identificado os seguintes conceitos para compor sua ontologia: câncer e todos os seus tipos (mieloma múltiplo, melanoma, carcinoma etc.) e tipos por locais de origem (câncer de mama, de tireoide, de estômago, entre outros); **tratamento de câncer** e seus tipos (quimioterapia, BCG, radioterapia etc.); **cura**, **contraindicação** e **faixa etária**.

Você pode ter especificado cada um desses conceitos e levantado outros, que propiciarão que o sistema ofereça respostas mais precisas, como tamanho do tumor, antecedentes familiares, outras doenças correlatas e medicamentos, por exemplo. Quanto mais conceitos você tiver definido na ontologia, mais precisa será a resposta de seu sistema.

Você pode, ainda, ter usado o sinônimo ou termo mais técnico, de acordo com o perfil do usuário que utilizará seu sistema. Se você está construindo um sistema para cidadãos comuns, o termo popular é o mais indicado. Se estava pensando em um sistema

para apoio a decisões médicas, você deve ter usado termos mais técnicos, por exemplo, carcinoma. Agora, você pode definir esses termos, categorizá-los (sistematizá-los hierarquicamente em classes e subclasses, construindo uma taxonomia) e, finalmente, implantá-los no *software* de elaboração de ontologias de sua preferência.

4.9 CONCLUSÃO

Falar de ontologia é falar de uma linguagem própria para indexação de assuntos no ambiente da *web*. Uma das características que distinguem as ontologias das demais linguagens de indexação é seu objetivo. Enquanto as demais linguagens são usadas para recuperar informação, as ontologias são usadas para processamento de conhecimento, isto é, para responder perguntas específicas de forma inteligente.

Não existe uma única forma de elaborar ontologias, mas existem várias alternativas viáveis e a possibilidade de compatibilizá-las com outras ontologias. A melhor solução quase sempre depende da aplicação e do uso que se pretende com uma ontologia.

A OWL – linguagem para descrição de ontologia – tem um papel fundamental na tradução de informações legíveis por humanos para informações legíveis por máquinas. Estudos continuam sendo realizados no intuito de permitir um número cada vez maior de aplicações, e vêm sendo foco de pesquisa para ferramentas, técnicas de inferências, fundamentos formais e extensões de linguagem.

É importante lembrar que, para uma ontologia atingir seus objetivos, faz-se necessário que as definições de seus conceitos sejam expressas por meio de lógica e que esses conceitos estejam organizados na forma de uma taxonomia.

Com o intuito de possibilitar o reúso de tesauros e vocabulários controlados na elaboração de ontologias, a *ISO*, em 2011, propôs uma nova norma para elaboração de tesauros e interoperabilidade com outros vocabulários: a norma internacional *ISO 25694* – pt. 1 e pt. 2. Enquanto a parte 1 atualiza as normas para construção de tesauros, a parte 2 estabelece regras que tornam os tesauros interoperáveis com outras linguagens de indexação, inclusive com as ontologias.

A construção de ontologias será mais eficiente se houver o trabalho cooperativo entre profissionais de diversas áreas, a saber: Biblioteconomia, Ciência da Informação, Linguística, Filosofia, Ciência da Computação, além de especialistas das áreas para as quais a ontologia estiver sendo construída. Nesse sentido, mais uma vez, se torna evidente a importância do trabalho colaborativo e interdisciplinar na elaboração de instrumentos de representação temática da informação.

RESUMO

À medida que a internet caminha em direção a uma *web* semântica, onde a informação codificada poderá ser interpretada por seres humanos e máquinas, aumenta a necessidade de métodos, técnicas e ferramentas que apoiem o desenvolvimento e o uso de vocabulários, como ontologias.

Nesta Unidade, foram apresentados conceitos que são fundamentais ao processo de construção de ontologias, passando por sua definição, história, características e componentes, bem como suas funções e aplicações.

Foram citadas, brevemente, as principais metodologias, ferramentas e linguagens usadas para a construção de ontologias. Mostramos que estas consistem em outro tipo de instrumento de representação temática da informação e que guardam semelhanças e diferenças com as outras linguagens igualmente destinadas à representação da informação, como o tesauro.

Sugestão de Leitura

TELLO, A. L. Ontologías en la web semântica. **Jornadas de ingenieria web**, [S.l.], n. 1, 2002. Disponível em: <http://www.informandote.com/jornadas/ngweb/articulos/jew02.pdf>. Acesso em: 30 out. 2004.

REFERÊNCIAS

ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, 2003.

BORST, W. N. **Construction of engineering ontologies**. 1997. Tese (Doutorado) – Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente, Enshede, 1997. Disponível em: <eprints.eemcs.utwente.nl/17377/01/t0000004.pdf>. Acesso em: 18 nov. 2014.

BREITMAN, Karin Koogan. ***Web* semântica:** a internet do futuro. Rio de Janeiro: LTC, 2005.

DALHBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, Rio de Janeiro, v. 7, n. 2, 1978. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/1680/1286>. Acesso em: 13 jan. 2014.

FERREIRA, A. B. de H. **Dicionário Aurélio eletrônico**. Rio de Janeiro: Nova Fronteira, 1993.

GRUBER, T. A Translation approach to portable ontology specifications. **Knowledge Acquisition**, [S.l.], v. 5, n. 2, p. 199-220, 1993.

HOUAISS, Antônio. **Dicionário Houaiss eletrônico:** versão 3.0. Rio de Janeiro: Objetiva, 2007.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval. Suíça: ISO, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-2:** Information and documentation: Thesauri and interoperability with other vocabularies: Part 2: Interoperability with other vocabularies. Suíça: ISO, 2013.

MUNDIM, Roberto Patrus. Alógica formal: princípios elementares. **Revista Economia & Gestão**, Belo Horizonte, v. 2, n. 3, 2002.

RAMALHO, R. ***Web* Semântica:** aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. Marília: Ed. UNESP, 2006.

SALES, Luana Farias. **Ontologias de domínio:** estudo das relações conceituais e sua aplicação. 2006. Dissertação (Mestrado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal Fluminense, Niterói, 2006.

WÜSTER, Eugen. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. In: RONDEAU, Guy; FELBER, Helmut. **Textes choisis de terminologie**, Québec: GIRSTERM – Université de Laval, 1981. p. 55-108.

UNIDADE 5

LINGUAGENS DE INDEXAÇÃO E RECUPERAÇÃO EM PRÁTICA

5.1 OBJETIVO GERAL

Apresentar outras linguagens de indexação/recuperação, já aplicadas em ambientes institucionais, para a escolha da linguagem mais adequada a determinado contexto.

5.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

a) conhecer outros tipos de instrumentos de representação temática da informação, desenvolvidos a partir dos princípios estabelecidos para a construção de tesauros;

b) identificar os princípios e características fundamentais que originaram os novos instrumentos;

c) dominar a metodologia de construção desses instrumentos de representação temática da informação, considerando sua aplicação para contextos específicos.

5.3 MIL CONEXÕES POR SEGUNDO

Figura 60 - O mundo conectado

Fonte: *Pixabay* (2016).[51]

Vivemos uma época em que nossas vidas são atingidas pela quantidade de informações, dados e conexões *on-line* que estão sendo feitas a cada segundo. Os serviços que prestam informação não ignoram esse fato e sabem que, neste exato momento, cada pessoa está buscando informação para resolver problemas de ordem pessoal ou institucional.

A partir daí, podemos afirmar que a tecnologia da informação e o armazenamento de dados não são modismos. Pelo contrário, a cada instante, administradores de SRI pensam em como desenvolver novas tecnologias, visando a seu objetivo máximo, que é atender os seus usuários. Nesse sentido, novas ideias e tecnologias surgem ao longo do tempo, visando a tornar a recuperação da informação mais eficaz.

5.4 APRESENTAÇÃO DE OUTRAS LINGUAGENS DE INDEXAÇÃO/ RECUPERAÇÃO DA INFORMAÇÃO JÁ APLICADAS EM AMBIENTES INSTITUCIONAIS

Nesta seção, você verá algumas iniciativas realizadas no Brasil e no exterior sobre linguagens de indexação/recuperação da informação, incluindo aspectos práticos, por exemplo, a adequação de cada tipo a determinada situação.

[52] Disponível em: <https://pixabay.com/pt/internet-global-terra-comunicação-1181587>.

Dentre esses instrumentos que conjugam os princípios do cabeçalho de assuntos e do tesauro, valendo-se dos fundamentos teóricos do tesauro, selecionamos três experiências de que temos notícia: o *Vocabulário Sistematizado* – linguagem que vem sendo aplicada a vários sistemas de recuperação da informação no Brasil –, o *Medical Subject Headings – MeSH* (E.U.) e o *Nuovo Soggettario* (Itália), como você poderá ver a seguir.

5.5 CABEÇALHO DE ASSUNTO E NOVAS LINGUAGENS

Figura 61 - Reciclando o cabeçalho de assunto

Fonte: *Pixabay* (2013).[53]

Antes de conhecermos novas linguagens, vamos fazer uma síntese de noções já abordadas neste Curso. O objetivo é rever o movimento da passagem do cabeçalho de assunto para novas linguagens de indexação/recuperação da informação, considerando que o cabeçalho de assuntos foi o ponto de partida para outras iniciativas que continuam acontecendo no mundo das linguagens documentárias.

Como visto na Unidade 2, o sistema de cabeçalhos de assunto foi amplamente usado nas bibliotecas dos EUA a partir da segunda metade do século XIX, assim como em outros países do mundo.

Também não é demais enfatizar que todo sistema de indexação/recuperação de assuntos, seja ele pré ou pós-coordenado, é determinado pela tecnologia existente na época. A tecnologia que determinou a sistematização de entradas dos cabeçalhos de assunto, feita por *Cutter*, como dito na Unidade 2, foram as fichas ou catálogos impressos.

[53] Disponível em: <https://pixabay.com/pt/reciclar-setas-verde-159767>.

O advento do computador representou um novo paradigma para o tratamento e a recuperação da informação. Para sua recuperação, particularmente, essa tecnologia passou a possibilitar o múltiplo acesso à informação, ou seja, ela passou a poder ser acessada por qualquer ponto pelo qual o documento tivesse sido indexado.

Se você tem boa memória, vai se lembrar do exemplo que demos sobre isso na Unidade 2, quando falamos da busca sobre "preços das pousadas em Natal". Está lembrado? Qualquer que fosse a ordem das palavras usadas na busca, você, certamente, obteria uma resposta. Assim, a importância da primeira palavra do cabeçalho de assunto para a entrada do cabeçalho ficou relativizada diante das possibilidades da nova tecnologia.

Constatou-se, então, a necessidade do desenvolvimento de instrumentos de controle terminológico mais adequados para o apoio à indexação e recuperação da informação. Isso se deu porque os sistemas informatizados propiciam acesso *on-line* à bases de dados na internet, organização de bibliotecas virtuais e outros instrumentos para a organização e disseminação da informação que exigem melhor controle da terminologia para filtragem de dados e recuperação mais eficiente.

Esse foi o momento em que se começou a pensar em sistemas que se valessem da vantagem de poder reunir unidades de representação temática (palavras/expressões que representassem os assuntos dos documentos) na hora da busca e pelo próprio pesquisador, ampliando suas possibilidades de satisfação na pesquisa. Esses eram os chamados sistemas pós-coordenados.

Iniciou-se aí o rompimento com o cabeçalho de assunto, que se tornou inadequado para as novas tecnologias, dando margem ao surgimento dos tesauros – sistemas pós-coordenados, que têm como unidade de representação o conceito, em vez da palavra.

Uma grande vantagem do conceito é que ele corresponde a um termo que pode ser constituído de várias palavras (na Unidade 3, citamos o Unitermo, sistema que restringia a representação dos documentos a uma única palavra). Sabemos que há assuntos que necessitam ser expressos por várias palavras. Você mesmo deve se lembrar de muitos exemplos. Na área de Biblioteconomia e Ciência da Informação, há um grande número deles. Só para lembrar alguns: "transferência da informação", "indexação de assuntos", "tecnologia da informação", "base de dados digital", "*Classificação Decimal Universal*" e tantos outros. Nas demais áreas do conhecimento não é diferente.

Você se lembra de que, na Unidade 3, falamos sobre conceito e Teoria do Conceito? Então, se tiver dúvidas sobre o assunto, volte lá e consulte o texto novamente.

Como visto naquela Unidade, o tesauro é um instrumento que se vale de bases teórico-metodológicas consistentes e estabelecidas por normas internacionais. Essas normas orientam a sistematização de conceitos em áreas do conhecimento, além de estabelecer princípios para a organização das relações entre conceitos e para a formação do termo. Os conceitos são elementos teóricos fundamentais para a elaboração de qualquer outro instrumento que venha atender as necessidades atuais de tratamento e recuperação de informação em base de dados.

Agora, você será apresentado às novas iniciativas de que falamos na introdução desta Unidade, que mesclam princípios do cabeçalho de assuntos e do tesauro: *Vocabulário Sistematizado*, *MeSH* e *Nuovo Soggettario*. Vamos conhecê-las?

5.6 *VOCABULÁRIO SISTEMATIZADO*

Figura 62 - Repensando a recuperação da informação

Fonte: *Pixabay* (2016).[54]

A questão é a seguinte: muitas bibliotecas e serviços de informação continuam usando maciçamente o cabeçalho de assunto para indexar/recuperar os documentos por seus temas. Quais seriam as razões para isso? Podemos dizer que elas têm fundamento na existência da *LCSH*.

Um dos motivos é que essa *Lista* formada por cabeçalhos de assunto vem sendo usada tradicionalmente pelas bibliotecas americanas desde o século XIX e esse padrão tem sido adotado por muitas bibliotecas do resto do mundo.

Tal fato nos leva a outro motivo, que também diz respeito à *LCSH*: a falta de opção. Não existem outras listas de cabeçalhos de assunto com cobertura tão grande de temas, com divulgação tão ampla e atualizada como a *LCSH*, podendo ser adquirida por qualquer indivíduo.

Paradoxalmente, esses mesmos motivos fazem com que a *LCSH* venha sendo utilizada pela maioria das bibliotecas que têm coleções gerais, sem que algum julgamento crítico seja feito sobre sua eficácia como instrumento de representação temática da informação e recuperação de documentos. Aliás, você já teve oportunidade de estudar isso antes. Se não estiver lembrado, consulte novamente a Unidade 2, seção 2.7 ("Avaliação de Cabeçalhos de Assunto").

Como proceder, então, para instituir novos instrumentos confiáveis destinados ao tratamento e recuperação da informação de documentos de coleções gerais que vêm sendo tratadas com cabeçalhos de assunto baseados na *LCSH*?

[54] Disponível em: <https://pixabay.com/pt/silhueta-cabeça-estante-sabe-1632912>.

No Brasil, para suprir essa lacuna, pode-se contar com uma proposta de instrumento terminológico compatível com a tecnologia atual. Trata-se do *Vocabulário Sistematizado*. Esse instrumento apropriou-se dos princípios científicos e normativos para a construção de tesauros, inovando, porém, quanto à metodologia: ela permite, entre outras coisas, a passagem do cabeçalho de assunto para um sistema baseado em conceitos, por meio da desconstrução do antigo sistema de representação de assuntos, para a reconstrução de um totalmente novo.

No exterior, estudos semelhantes aconteceram desde o final dos anos 1960, como no caso do *MeSH*, nos EUA, que já se valia de uma estrutura hierárquica pensando na melhoria da recuperação da informação (LANCASTER, 1986) e, mais tarde, do *Nuovo Soggettario*, na Itália. Adiante, você saberá mais detalhes dessas duas experiências. Por enquanto, vamos nos ocupar do *Vocabulário Sistematizado*.

Atenção

Afinal, o que é *Vocabulário Sistematizado*?

Trata-se de uma linguagem documentária destinada à indexação e recuperação de informações contidas em documentos de coleções gerais, que tem como unidade de representação o conceito.

Sua elaboração baseia-se nos mesmos princípios sugeridos para a construção de tesauros. Um dos principais aspectos metodológicos do *Vocabulário Sistematizado* é a transformação de cabeçalhos de assunto ou palavras-chave (baseados em palavras) em descritores (baseados em conceitos).

Motta (1987), preocupou-se com os resultados práticos da recuperação de documentos da área econômica, com a qual trabalhava. Estes resultados provaram-se insatisfatórios ao longo de algum tempo porque a indexação dos documentos era feita com base em um índice de periódicos de economia, elaborado a partir de cabeçalhos de assunto. *Motta* (1987) percebeu que as falhas na recuperação da informação deviam-se, principalmente, à pré-coordenação dos cabeçalhos de assunto. A solução encontrada foi desmembrar os cabeçalhos de assunto, considerar seus componentes como conceitos e não como palavras e passar a indexar os documentos com os termos de forma independente, sem estarem previamente atrelados um ao outro. Se você não está lembrado dos problemas que envolvem pré-coordenação e recuperação da informação, volte à Unidade 2 e reveja os prós e contras dos cabeçalhos de assunto.

Mais tarde, uma nova metodologia desenvolvida pela Prof.ª *Maria Luiza Campos* – o *Vocabulário Sistematizado* – também partiu para a transformação de cabeçalhos de assunto em conceitos. Essa metodologia foi posta em prática na *Fundação Casa de Rui Barbosa* (BRASIL, 2002) e, mais tarde, em outras instituições (MOTTA et al., 2009).

O *Vocabulário Sistematizado* destina-se à conversão de um sistema de cabeçalhos de assunto em um novo, baseado em conceitos e não em palavras, modernizando-o e compatibilizando-o com novas tecnologias.

Uma grande vantagem do Vocabulário Sistematizado para serviços que disponibilizam informação sobre coleções gerais - como bibliotecas de universidades, que, em geral, utilizam listas de cabeçalhos de assunto para orientar a indexação de seus documentos - é que a transformação das palavras do cabeçalho de assunto em conceitos e o estabelecimento de suas inter-relações (operações que seguem as mesmas normas e padrões internacionais para o desenvolvimento de tesauros) possibilitam a visão e o acesso à coleção toda de forma sistêmica, revelando as relações entre termos de diversas áreas do conhecimento– exigência cada vez maior, dada a característica crescente da interseção dessas áreas. Isso fica visível nos exemplos do Quadro 17, a seguir, em que se pode observar que a inter-relação, ou seja, a relação entre os termos, acontece, mesmo eles pertencendo a áreas do conhecimento diferentes.

Quadro 17 - Relação entre termos de áreas do conhecimento diversas

Ex.:

ÁGUA DOCE ⟶ **(classe 550 *CDD*, Ciências da Terra)**

VT PEIXE DE ÁGUA DOCE ⟶ **(classe 590 *CDD*, Zoologia)**

ÁGUA DAS CHUVAS ⟶ **(classe 550 *CDD*, Ciências da Terra)**

VT ATMOSFERA ⟶ **(classe 520 *CDD*, Astronomia)**

LUZ ⟶ **(classe 530 *CDD*, Física)**

VT ECLIPSE ⟶ **(classe 520 *CDD*, Astronomia)**

VT VISÃO HUMANA ⟶ **(classe 590, Zoologia)**

Fonte: produção do próprio autor (2017).

5.6.1 Características do *Vocabulário Sistematizado*

Como revelaram as experiências de que se tem notícia, as principais características desse novo vocabulário, elaborado com base nas teorias que apoiam a construção de tesauros a partir do cabeçalho de assunto, podem ser identificadas nos seguintes pontos:

a) abandono da distinção entre "termo principal" e "subdivisão": todas as subdivisões são transformadas em termos;

b) adesão às normas internacionais para o controle terminológico e indicação das relações semânticas (*ISO 25964-2011*, norma internacional para tesauros e interoperabilidade com outros vocabulários);

c) adoção e indicação das relações próprias de um tesauro (equivalências hierárquicas e associativas);

d) adesão à estrutura classificatória (categorial) mono-hierárquica (os termos devem pertencer a uma e somente uma classe);

e) uso de linguagem pós-coordenada apropriada aos objetivos do SRI (a linguagem para uma biblioteca científica deverá ser diferente da usada para uma biblioteca escolar);

f) adoção do conceito ("descritor") como unidade de representação dos assuntos dos documentos;

g) adesão ao compromisso de representar, de forma padronizada, assuntos em coleções de acervos gerais.

5.6.2 Objetivos do *Vocabulário Sistematizado*

O *Vocabulário Sistematizado* tem como objetivos:

a) melhorar a recuperação da informação em coleções de acervos gerais (todas as áreas do conhecimento);

b) facilitar a estratégia de busca em coleções de acervos gerais;

c) melhorar a integração do usuário com o SRI de coleções de acervos gerais;

d) melhorar a integração com outros sistemas automatizados, propiciando interoperabilidade semântica.

5.6.3 Componentes do *Vocabulário Sistematizado*

Como as demais linguagens referidas neste Curso, o *Vocabulário Sistematizado* compõe-se de um vocabulário e de regras para sua aplicação:

a) vocabulário: formado por termos organizados alfabética e hierarquicamente e relações com outros termos. No exemplo a seguir, aparecem as seguintes relações: hierárquicas genérico-específicas (VTE) e associativas (VTA). Além das relações, estão exibidas no exemplo: categoria a que o termo pertence (aqui, categorização da *CDD*) e nota de definição (ND).

Exemplo:

ÁGUA (HIDROLOGIA)

Categoria: CIÊNCIAS DA TERRA (550)

ND: Parte líquida que cobre aproximadamente 70% da superfície terrestre, sob a forma de mares, lagos e rios.

VTE ÁGUA DAS CHUVAS
VTE ÁGUA DOCE
VTE ÁGUA TERMAL
VTA ABASTECIMENTO DE ÁGUA
VTA ANIMAL AQUÁTICO
VTA CICLO DA ÁGUA
VTA PLANTA AQUÁTICA
VTA TRATAMENTO DA ÁGUA

Será que você reparou que, embora o termo ÁGUA (HIDROLOGIA) pertença à categoria Ciências da Terra, ele se relaciona com termos de outras categorias, como Animal Aquático (Zoologia), Planta Aquática (Botânica) e Tratamento de Água (Meio Ambiente)? Em um tesauro destinado a uma área específica do conhecimento, isso não aconteceria; as relações seriam somente entre os termos da área coberta.

Explicativo

Assim como nos tesauros, as relações entre os termos, em um *Vocabulário Sistematizado*, podem ser de equivalência, hierárquicas ou associativas. Porém, a simbologia para representar tais relações pode variar de sistema para sistema, de acordo com o que foi convencionado previamente pela equipe responsável pelo desenvolvimento do *Vocabulário*. As relações de gênero-espécie, que, nos tesauros, são simbolizadas por TG/TE, nos *Vocabulários Sistematizados* de que se tem conhecimento apareceram como Ver Também (VT) ou como Ver também-termo genérico e Ver também-termo específico (VTG/VTE); as relações associativas também podem aparecer como termo relacionado (TR).

Além das relações, os *Vocabulários Sistematizados* incluíram outros elementos e seus respectivos símbolos, à semelhança de um tesauro, a saber: Notas de Definição (ND), muitas vezes chamadas de Notas de Escopo (NE), Notas definitórias ou, simplesmente, Definição (D) e Notas de Aplicação (NA), às vezes chamadas de Notas de Indexação (NI).

b) regras para aplicação do *Vocabulário* (NA): orientações sobre como usar os termos no momento da indexação do documento. Veja o exemplo a seguir, que inclui o termo "preservação ambiental" e a NA com a orientação sobre como usá-lo.

   Exemplo:

   PRESERVAÇÃO AMBIENTAL

   NA: Quando se tratar da preservação de determinado elemento do ambiente, indexe pelo descritor + elemento: PRESERVAÇÃO AMBIENTAL; ÁGUA (HIDROLOGIA).

5.6.4 Apresentação do *Vocabulário Sistematizado*

Como já mencionado, o *Vocabulário Sistematizado* dispõe de termos que podem ser apresentados em uma listagem alfabética e/ou sistemática. A seguir, você pode visualizar um exemplo de exibição sistemática dessa linguagem:

***NÚMERO**

**Conceitos numéricos

***Número natural

***Número inteiro

**Aritmética

***Operações

****Adição

****Subtração

****Multiplicação

****Divisão

5.6.5 Atividade

Relacione a primeira e a segunda colunas de acordo com as características das linguagens:

Coluna 1

(1) Cabeçalho de assunto
(2) Tesauro
(3) *Vocabulário Sistematizado*
(4) Cabeçalho de assunto e *Vocabulário Sistematizado*
(5) Tesauro e *Vocabulário Sistematizado*
(6) Todas as linguagens
(7) Linguagem natural
(8) Cabeçalho de assunto, Tesauro e *Vocabulário Sistematizado*

Coluna 2

( ) Sistema pós-coordenado
( ) Para tratamento de coleções de acervos gerais
( ) A unidade de representação é o conceito
( ) Sistema pré-coordenado
( ) Sistema para tratamento de acervos especializados
( ) Instrumento de recuperação da informação
( ) A unidade de representação é a palavra
( ) Facilita a indexação e a recuperação da informação
( ) Padroniza a linguagem usada nos SRI
( ) Possui relações explícitas de equivalência, hierárquicas e associativas
( ) Baseado em normas preestabelecidas para sua construção
( ) Possui regras para aplicação do termo
( ) Não se preocupa com padronização do vocabulário

Resposta comentada

(5) Ambos possibilitam a conjugação de conceitos no momento da busca (pós-coordenação).
(4) As duas linguagens se destinam a acervos gerais.
(5) As duas linguagens utilizam definições para contextualizar os termos, que nomeiam os conceitos.
(1) O cabeçalho de assunto conjuga as palavras de um assunto no momento da indexação, ou seja, a conjugação é predeterminada.
(2) Por definição, o tesauro se destina ao tratamento de acervos especializados.
(6) Todas as linguagens citadas prestam-se à recuperação da informação.
(7) A palavra usada na comunicação diária é a unidade que representa a linguagem natural e seu significado varia de acordo com o contexto discurso.
(8) São objetivos e funções das três linguagens citadas.

(5) São linguagens que utilizam teorias e normas e padrões internacionais com objetivo específico de padronizar a linguagem em SRIs.

(5) Apenas o tesauro e o *Vocabulário Sistematizado* explicitam as relações entre os termos. O cabeçalho de assunto só apresenta equivalências e relações de VT (Ver também).

(5) Apenas o tesauro e o *Vocabulário Sistematizado* são construídos a partir de regras internacionais para sua elaboração.

(8) As regras para aplicação de termos estão presentes nas três linguagens.

(7) A linguagem natural não faz controle vocabular.

5.6.6 Elaboração de *Vocabulário Sistematizado*

Você deve estar lembrado de que a Unidade 3 foi dedicada a tesauros e que dissemos que as regras para a construção de um tesauro baseado em conceitos podiam ser aplicadas à construção de qualquer outro vocabulário controlado. Pois bem, com o *Vocabulário Sistematizado* não é diferente. As regras básicas para sua construção e manutenção são as mesmas usadas para a elaboração de tesauros.

Porém, vale lembrar que a metodologia de construção do *Vocabulário Sistematizado* foi pensada principalmente para atualizar os vocabulários de instituições cujas coleções abrangem várias áreas do conhecimento e que ainda utilizam cabeçalhos de assunto para indexar seus documentos. Se você se recorda, a Unidade 2 revelou a fragilidade desse tipo de representação de assuntos para as necessidades de informação atuais.

Por isso a importância do *Vocabulário Sistematizado*, cujo objetivo principal é transformar os cabeçalhos de assunto (baseados em palavras pré-coordenadas) em um conjunto de conceitos (para serem pós-coordenados no momento da busca). Nesse caso, três ações iniciais devem ser feitas:

a) separar, numa listagem única, os nomes de pessoas, lugares, instituições, projetos e outros indivíduos que tenham identidade própria e não precisem ser definidos (a essa listagem vamos dar o nome de "identificadores");

b) separar todas as palavras que compõem os cabeçalhos de assuntos (essas palavras, depois de definidas, serão os conceitos do novo vocabulário);

c) procurar eleger conceitos já consagrados pela literatura das áreas; pode-se conseguir isso consultando-se, por exemplo, dicionários especializados, manuais, especialistas das áreas ou qualquer outra fonte de referência confiável.

Exemplo 1:

Cabeçalho de assunto: CAFÉ – EXPORTAÇÃO

Conceitos:

a) CAFÉ

b) EXPORTAÇÃO

Exemplo 2:

Cabeçalho de assunto: FÍSICA – ELETRICIDADE

Conceitos:

a) FÍSICA

b) ELETRICIDADE

Exemplo 3:

Cabeçalho de assunto: MÚSICA - HISTÓRIA - BRASIL

Conceitos:

a) HISTÓRIA DA MÚSICA

b) BRASIL

A partir daí, as demais etapas são as mesmas utilizadas para a construção de um tesauro (definição dos termos, categorização, estabelecimento de relações entre eles).

Como exemplos de experiências de *Vocabulário Sistematizado* no Brasil, podemos citar os desenvolvidos pela Fundação Casa de Rui Barbosa (Rio de Janeiro), pela Universidade Federal Fluminense (UFF), pela Financiadora de Estudos e Projetos (FINEP) e pelo Serviço Social do Comércio (SESC/DN) do Rio de Janeiro.

5.7 PARA ONDE CAMINHAM AS LINGUAGENS DOCUMENTÁRIAS?

Figura 63 - LD: inovar para melhorar

Fonte: *Pixabay* (2016).[55]

[55] Disponível em: <https://pixabay.com/pt/escolher-o-caminho-certo-1536336>.

Nesta seção, serão vistas algumas iniciativas de LD que foram desenvolvidas e que estão em uso corrente, quer em determinada área do conhecimento, como no caso do *Medical Subject Headings – MeSH* (EU), dedicado à área médica, quer em determinado país, como no caso do *Soggettario* (Itália), como veremos a seguir.

5.7.1 *MeSH*

Figura 64 - *MeSH*: uma experiência americana na área médica

Fonte: *Pexel* (20--?).[56]

No Quadro 18, você pode se certificar de algumas informações colhidas da página de introdução do *MeSH* (2014).

Quadro 18 - Como surgiu o *MeSH*?

- A primeira lista oficial de cabeçalhos de assuntos, publicada pela Biblioteca Nacional de Medicina dos Estados Unidos, surgiu em 1954, sob o título *Subject Heading Authority List* (Lista de Autoridade de Cabeçalho de Assunto). Ela foi baseada na lista de autoridade da instituição, que havia sido usada para publicação do *Current List of Medical Literature*. Este, por sua vez, tinha incorporado cabeçalhos do Catálogo-índice da Biblioteca e do *Quarterly Cumulative Index Medicus*;
- em 1960, uma nova edição do *Subject Heading Authority List*, totalmente atualizada, surgiu sob o nome de *Medical Subject Headings* (*MeSH*). Posteriormente, em 1963, a segunda edição do MeSH, projetada especificamente para ser empregada em um sistema pós-coordenado (MEDLARS, atual MEDLINE), continha 5.700 descritores, comparados com 4.400 termos de edição de 1960. Dos cabeçalhos usados em 1960, 113 foram retirados em favor de novos termos. A edição de 2015 contém 27.455 descritores;
- inicialmente, o sistema *MeSH* foi elaborado com base em cabeçalhos de assunto, tendo, como unidade de representação, a palavra. Era, portanto, um sistema pré-coordenado. Posteriormente, ele se torna um sistema-descritor, baseado em conceitos, com o objetivo de se tornar compatível com o *Unified Medical Language System* (UMLS). (NATIONAL LIBRARY OF MEDICINE, 2014).

Fonte: *MeSH* (2014).

[56] Disponível em: <https://www.pexels.com/photo/black-sphygmomanometer-33258>.

Desde o começo, o *MeSH* pretendeu ser uma linguagem dinâmica, com procedimentos para recomendação e exame de incorporação de novos cabeçalhos necessários. O conteúdo do vocabulário era relacionado ao uso de termos da própria literatura e evoluiu para a incorporação de novos conceitos no campo da Medicina. O uso do computador tornou possíveis revisões mais práticas e sistemáticas.

Figura 65 - O que é *MeSH*?

Fonte: *Pixabay* (2013).[57]

O *Medical Subject Headings* (*MeSH*®) *Thesaurus* é um vocabulário controlado produzido pela *National Library of Medicine* (NLM) dos EUA, utilizado para indexação e busca de informações em documentos da área biomédica e de saúde, cujos objetivos e características veremos a seguir.

5.7.1.1 *Objetivos*

O objetivo do *MeSH* é auxiliar a representação e recuperação da informação na área biomédica.

É usado para indexar artigos dos principais periódicos biomédicos no mundo e para alimentar a base de dados *MEDLINE®/PubMED®*. É também usado para tratamento de todo material documental adquirido pela *NLM*.

5.7.1.2 *Caracterização*

Dentre as características do *MeSH*, podem ser destacadas as seguintes:

a) sistema desenvolvido a partir de cabeçalhos de assunto;

b) sistema atualmente conceito-orientado (cuja metodologia considera conceitos, e não palavras);

c) unidade de representação: conceito;

d) conceitos dispostos alfabética e hierarquicamente;

e) sistema parcialmente pré-coordenado e pós-coordenado;

f) existência de rede de remissivas;

g) existência de regras para aplicação do vocabulário;

[57] Disponível em: <https://pixabay.com/pt/rastreamento-de-pulso-163708>.

h) existência de histórico dos termos: data de entrada do termo no sistema e alterações;

i) atualização periódica;

j) sistema para acervo especializado na área biomédica;

k) termos selecionados e tratados por um *staff* composto de especialistas da área biomédica e outras áreas, indexadores e outros profissionais (LANCASTER, 1972).

Figura 66 - Dados de um descritor do *MeSH*

National Library of Medicine - Medical Subject Headings

2015 MeSH

MeSH Descriptor Data

Return to Entry Page

Standard View. Go to Concept View; Go to Expanded Concept View

MeSH Heading	Cardiovascular Physiological Phenomena
Tree Number	G09.330
Annotation	general or unspecified; prefer specifics; DF: CARDIOVASC PHYSIOL
Scope Note	Processes and properties of the CARDIOVASCULAR SYSTEM as a whole or of any of its parts.
Entry Term	Cardiovascular Physiological Concepts
Entry Term	Cardiovascular Physiological Phenomenon
Entry Term	Cardiovascular Physiology
Entry Term	Physiology, Cardiovascular
Allowable Qualifiers	DE GE IM RE
Previous Indexing	Cardiovascular System/physiology (1966-1997)
History Note	2009 (1998)
Date of Entry	19990101
Unique ID	D002320

MeSH Tree Structures

Fonte: *National Library of Medicine* (20--?).[58]

Na Figura 66, muitos sinônimos, quase-sinônimos e conceitos intimamente relacionados estão incluídos como entrada ("*Entry term*") para ajudar os usuários a encontrarem o descritor *MeSH* mais relevante para o conceito que estão procurando. Em bancos de dados *on-line* da *NLM*, muitos termos digitados pelos pesquisadores são automaticamente mapeados para descritores *MeSH*, para facilitar na obtenção de informações relevantes.

O Sistema *MeSH* é parcialmente pré-coordenado e parcialmente pós-coordenado. Nele, os conceitos que aparecem frequentemente juntos podem ser combinados em um termo pré-coordenado e esse termo pode, então, ser coordenado com outros na hora da pesquisa. Um exemplo é o termo "neoplasmas do fígado", que deve ser coordenado na hora da busca com o termo "radioterapia", se estivermos buscando o assunto "radioterapia de neoplasmas do fígado". (LANCASTER, 1972).

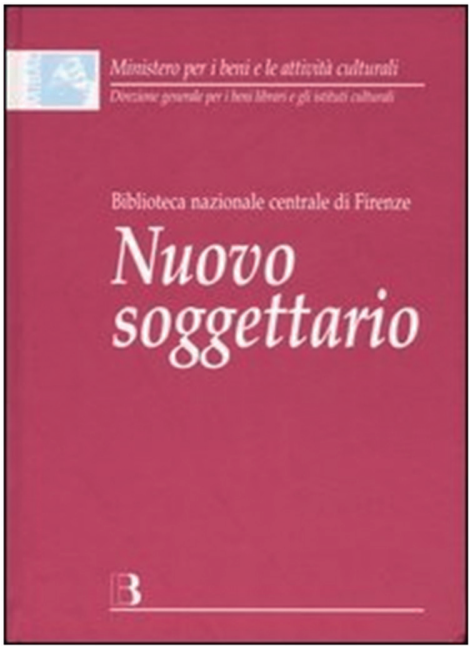
[58] Disponível em: <http://www.nlm.nih.gov/*MeSH*/introduction.html>.

5.7.1.3 *Componentes*

Como toda linguagem que se destina à indexação e recuperação da informação, o *MESH* é composto de um conjunto de termos (descritores) e relações (equivalências e hierárquicas), bem como regras para aplicação da linguagem.

5.7.2 *Nuovo Soggettario*: uma experiência na Itália

Figura 67 - Instrumento de controle de assunto



Fonte: *Wikipédia* (20--?).[59]

O projeto *Nuovo Soggettario* é desenvolvido pela *Biblioteca Nacional Central de Florença (BNCF)*, dado seu papel institucional de arquivo nacional do livro e de agência bibliográfica nacional, responsável pela *Bibliografia Nacional Italiana (BNI)*.

O projeto nasceu da convicção, apoiada pela prática de indexação por muitas bibliotecas italianas, de que o *Soggettario* (lista de cabeçalhos de assunto) de 1956 estava defasado e necessitava de revisão geral à luz dos desenvolvimentos contemporâneos da área de Biblioteconomia e Ciência da Informação. Trata-se de desenvolvimentos que foram produtos de teorias, normas internacionais e outros sistemas de indexação.

O *Nuovo Soggettario* é um instrumento construído pela *BNCF*, Itália, que implica a reformulação do cabeçalho de assunto tradicional, visando a um modelo flexível para a indexação de assuntos feita pelos serviços de informação participantes de um sistema nacional e pela *Bibliografia Nacional Italiana*.

Uma das características desse instrumento é que ele pode ser tanto um sistema pré-coordenado como pós-coordenado, pois a unidade de representação usada em seus cabeçalhos de assunto é o conceito e sua sintaxe é estabelecida com rigor. A seguir, você poderá verificar os objetivos desse trabalho, suas características e componentes.

[59] Disponível em: <https://upload.wikimedia.org/wikipedia/it/2/2e/Bibliografica_Nuovo_soggettario.jpg>.

5.7.2.1 *Objetivos*

O objetivo do *Nuovo Soggettario* é renovar o catálogo de assuntos da *BNCF* de 1956 – usado desde então pela maioria das bibliotecas na indexação de assuntos –, a fim de disponibilizá-lo para as instituições italianas que lidam com acervos documentais.

5.7.2.2 *Caracterização*

O *Soggettario* de 1956 era um sistema de cabeçalhos de assunto, tendo a palavra como unidade de representação.

O *Nuovo Soggettario* partiu de sua análise para a identificação dos termos que formariam o novo sistema, tendo o conceito como unidade de representação. Devido a essa característica, o vocabulário do *Nuovo Soggettario* é considerado um tesauro geral, por por possibilitar a visão-geral das áreas do conhecimento, integrando-as em um só instrumento. Além disso, outras características desse instrumento são:

a) é um sistema pré e pós-coordenado;

b) estabelece a sintaxe com rigor (ver exemplo na subseção 5.7.2.5: O *Nuovo Soggettario* em prática);

c) realiza o trabalho em rede;

d) segue princípios de especificidade, exaustividade e coextensão;

e) utiliza princípios de sistemas analítico-sintéticos;

f) observa princípios da Teoria da Classificação Facetada;

g) fundamenta a linguagem em padrões internacionais (Ex.: *ISO* e FRBR);

h) propicia interface amistosa sistema-usuário;

i) possibilita a interoperabilidade com outros instrumentos de representação temática;

j) atualiza periodicamente o sistema;

k) possui versatilidade para indexação de documentos de diversas naturezas;

l) possui um conjunto de termos controlados pela composição de assuntos;

m) fundamenta sua estrutura em um cabeçalho principal e subdivisões. A ordem de citação é baseada na tipologia das categorias fundamentais de *Ranganathan* (**p**ersonalidade, **m**atéria, **e**nergia, e**s**paço e **t**empo).

5.7.2.3 *Componentes*

O *Nuovo Soggettario* é constituído de: um conjunto de regras sintáticas e semânticas destinadas ao tratamento do termo que representa o assunto de um documento; um dicionário de sinônimos multidisciplinar em italiano, contendo a terminologia a ser utilizada na indexação, e, ainda, catálogos *on-line* e outros bancos de dados para pesquisa de assunto pelos usuários.

Os termos do *Nuovo Soggettario* são provenientes: do *Soggettario* de 1956 e atualizações; de repertórios e enciclopédias; de catálogos e índices; de termos propostos pela *BNI*, pela *BNCF* (especialmente fotografias) e por outras bibliotecas.

Explicativo

Uso do trabalho colaborativo para construção do *Nuovo Soggettario*

Figura 68 - Trabalhar em rede é um grande negócio!

Fonte: *Pixabay* (2015).[59]

O *Nuovo Soggettario* é um trabalho de grande abrangência terminológica, sendo composto por mais de 30 mil termos. Isso mostra que o universo abrangido por esse instrumento é rico em variedade linguística. Vale, então, ressaltar a importância do trabalho em rede. Basta observar que a base para elaboração do *Nuovo Soggettario* foram os termos utilizados pelas bibliotecas italianas e pela *BNI*, experiência de diversos profissionais.

Outro trabalho importante realizado colaborativamente foi a criação de normas para o controle do vocabulário e para a construção dos cabeçalhos de assunto e o protótipo do tesauro multidisciplinar em formato eletrônico e passível de consulta na rede. Você pode visualizar este trabalho, na língua italiana, consultando o *link:* <http://www.bncf.firenze.sbn.it/pagina.php?id=183>. Ele foi elaborado entre 2004 e 2006 por um grupo de trabalho composto por bibliotecários do setor de indexação e classificação de assunto da *BNI* e colaboradores externos. É interessante observar, ainda, que essa colaboração não se deu somente entre as bibliotecas citadas, já que a criação do *Nuovo Soggettario* aconteceu a partir da troca de experiências com outras bibliotecas nacionais, institutos, centros de documentação e outras organizações que trabalham com materiais não exclusivamente bibliográficos, mas, também, museológicos, arquivísticos e assemelhados, sem descuidar dos padrões internacionais.

5.7.2.4 *Estágios de elaboração e aplicação do Soggettario*

Os estágios através dos quais se desenvolveu o projeto para os novos cabeçalhos de assunto foram os seguintes:

a) **2000-2002:** estudo de viabilidade sobre a renovação do *Soggettario* por uma equipe de consultores externos. O estudo analisou e definiu princípios, características e requisitos para a nova linguagem de indexação, produzindo um conjunto de recomendações para desenvolver um planejamento a curto e médio prazos;

[60] Disponível em: <https://pixabay.com/pt/sistema-rede-notícias-pessoal-927147>.

b) **2003-2005:** realização de testes com a nova linguagem, com um experimento em "casos" bibliográficos específicos. Início da fase de elaboração de protótipo e escolhas tecnológicas; preparação de regras sintáticas e semânticas; implementação de *thesaurus on-line*;

c) **2006:** publicação do *Nuovo Soggettario: Guida al sistema italiano di indicizzazione per soggetto Prototipo del Thesaurus*, contendo regras (sintáticas e semânticas) e de pesquisa para o dicionário de sinônimos do usuário *on-line*, como um serviço de assinatura;

d) **2007:** início do uso do *Nuovo Soggettario* pela *BNI* e início da ampliação terminológica do tesauro;

e) **2008:** início do uso do *Nuovo Soggettario* por outras bibliotecas italianas;

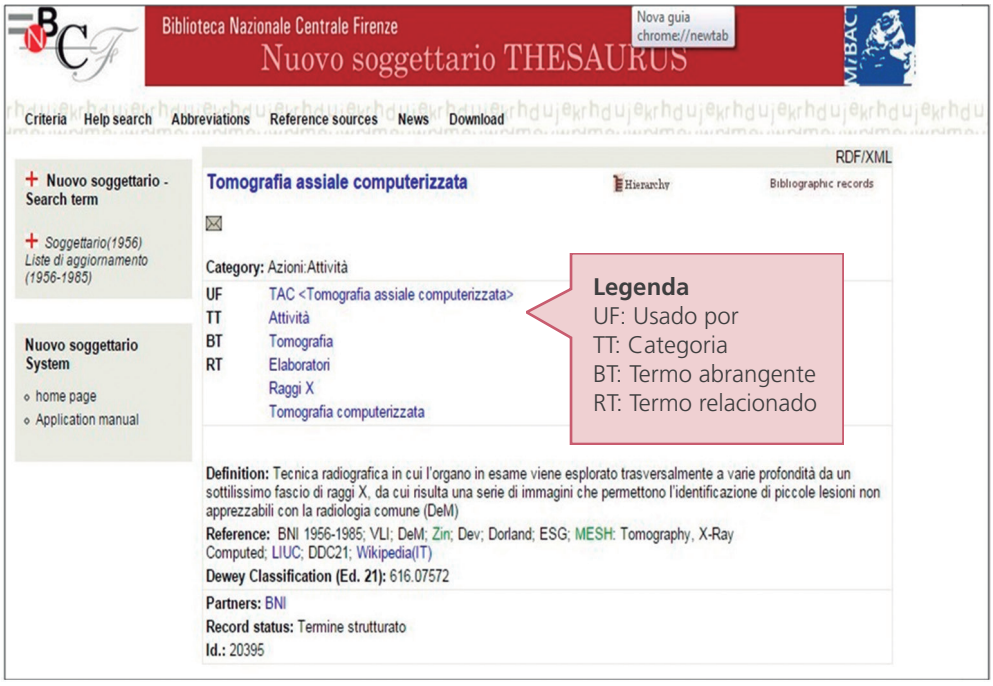
f) **2010:** disponibilização *on-line* do *Nuovo Sogettario* para acesso gratuito, no *site* da *BNCF*. Disponibilização, também, do *Manual de aplicação para auxílio à indexação*, periodicamente atualizado. Realização do protótipo de conversão do formato dos metadados do vocabulário do *Nuovo Soggetario* para o padrão *Simple Knowledge Organization System/Resource Description Framework* (SKOS/RDF – padrões indicados pelo consórcio *W3C*), considerando, ainda, a existência de dados ligados e a possibilidade de interoperabilidade com outras ferramentas *on-line*;

g) **2011:** início do experimento de indexação automática de recursos ativos digitais através do *thesaurus*;

h) **2013:** ativação da interoperabilidade recíproca com a *Wikipédia* italiana.

5.7.2.5 *O Nuovo Soggettario em prática*

Figura 69 - Página de busca do *Nuovo Soggettario Thesaurus*



Fonte: *Soggettario* (2007).[61]

[61] Disponível em: <http://thes.bncf.firenze.sbn.it>.

5.7.2.6 *Origem da sintaxe*

Veja exemplos de origem da sintaxe (ordem dos termos) no *Nuovo Soggettario,* a partir de declarações de assuntos:

a) **declaração 1:** ocupação alemã da Itália durante a última guerra.

Explicação: a Alemanha é agente da ação transitiva direta "ocupação bélica" para o objeto "Itália".

Sequência de termos: ocupação bélica [*da parte della*] Alemanha – 1943-1945;

b) **declaração 2:** Veneza durante o século XVIII.

Explicação: Veneza é o conceito-chave e o objeto único de tratamento; a indicação do "Século XVIII" desempenha o papel complementar de tempo.

Sequência de termos de indexação: Veneza - Séc. XVIII;

c) **declaração 3:** guia turístico de Veneza.

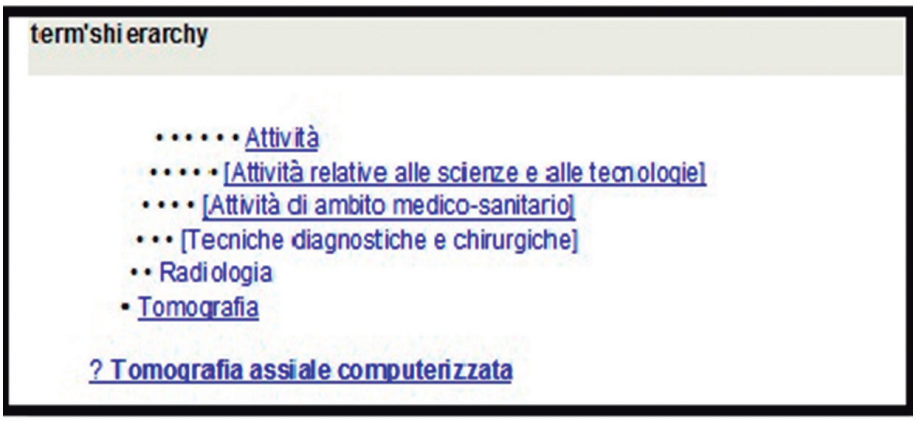
Explicação: Veneza é o conceito-chave e o objeto único de tratamento; a indicação de que se trata de um guia turístico desempenha um papel complementar de forma bibliográfica.

Sequência de termos: Veneza - Guia.

5.7.2.7 *Forma de apresentação do* Nuovo Soggettario

O *Soggettario*, assim como o *Vocabulário Sistematizado*, possui uma forma de apresentação em ordem alfabética e outra em ordem sistemática, conforme o exemplo a seguir:

Figura 70 - Forma sistemática de apresentação dos termos do *Soggettario*



Fonte: *BNCF* (2007).[62]

5.7.3 Folksonomia: um instrumento de indexação colaborativo

A palavra "folclore" você conhece bem, não é mesmo? E deve saber que tem a ver com "cultura popular." Como saber nunca é demais, vamos ser curiosos e ver de onde ela veio? O *Dicionário Eletrônico Houaiss*

[62] Disponível em: <thes.bncf.firenze.sbn.it>.

da Língua Portuguesa nos informa que a palavra tem origem na língua inglesa e foi criada a partir da junção de "*folk*" (povo, nação) + "*lore*"(ato de ensinar, instruir, educação, lição) (HOUAISS, 2007).

Agora, pergunto a você: o que "folclore" e "folksonomia" têm em comum? Reparou que ambas as palavras têm o mesmo radical "*folk*", indicando que as duas têm a ver com "povo"?

Pois bem, em 2004, *Thomas van der Wal* criou o termo **folksonomia**, a partir da junção de "*folk*" (povo, pessoas) com "*taxonomy*" (taxonomia: ciência ou técnica que lida com a descrição, identificação e classificação dos seres). Esse novo termo foi criado para se referir ao novo paradigma de organização dos conteúdos (indexação) dos recursos digitais na *web*, de cuja organização os próprios usuários participam.

A folksonomia é também conhecida como: "classificação social", "etiquetagem social","classificação popular", "taxonomia popular", "etnoclassificação", "classificação compartilhada", "etiquetagem aberta", "etiquetagem livre", entre outras denominações.

Folksonomia

De acordo com *Thomas van der Wal*, é o resultado da atribuição livre e pessoal de etiquetas (*tagging*) a informações e objetos (qualquer coisa com *URL*), visando a sua recuperação. Isso é feito em um ambiente social (comumente compartilhado e aberto a outros). A folksonomia é criada a partir do "*tagging*" pelo próprio consumidor da informação.

Explicativo

Pegando carona no conceito de "parada"

Figura 71 - Qual o conceito de "parada"?

Fonte: *Pixabay e Freeimages* (2020).[63]

Vamos pegar essa carona para recordar algo interessante. Reparou quantos significados um mesmo nome pode ter?

Só para refrescar sua memória: lembra-se de quando falamos da necessidade de se controlar um vocabulário, quando citamos as funções e objetivos de um vocabulário controlado (Unidade 1, seção 1.6)? Pois estamos diante de um exemplo típico de palavras equivalentes, que, se não forem submetidas a um mecanismo de controle, podem ocasionar muito ruído num SRI, afetando a qualidade do resultado de uma busca.

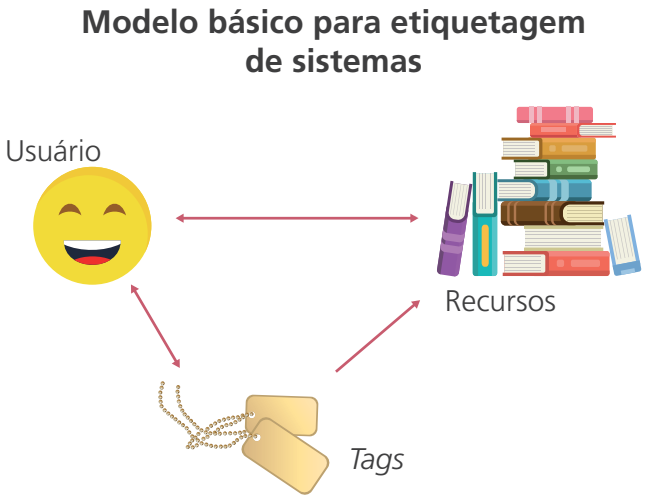
Na figura exibida, os três objetos têm a mesma denominação, mas se referem a objetos diferentes. Uma linguagem documentária com mecanismos de controle vocabular certamente daria conta de explicitar o significado do(s) objeto(s), orientando corretamente a pesquisa do usuário. Em uma folksonomia, em princípio, tal controle vocabular não é observado, uma vez que esses sistemas utilizam

[63] Primeira imagem: Parada militar. Disponível em: <https://cutt.ly/qku3VDO>; segunda imagem: Parada de ônibus. Disponível em: <https://cutt.ly/6ku3ZC4>; terceira imagem: Parada de mão. Disponível em: <https://cutt.ly/Eku3Mj5>.

palavras-chave atribuídas pelos usuários –- que empregam termos provenientes da linguagem natural, dispensando o uso de vocabulários controlados, contribuindo, dessa forma, para a escolha subjetiva das palavras que integrarão a folksonomia.

5.7.3.1 *Objetivos*

Figura 72 - Modelo de etiquetagem (*tagging*)



Fonte: *Columbia University* (20--?).[64]

A folksonomia tem vários objetivos, dentre os quais:

a) recuperação da informação a posteriori;

b) gerenciamento de informação pessoal;

c) *social bookmarking* – compartilhamento público de *links*, resultando em uma rede de *tags* criada pelos usuários;

d) reunião e compartilhamento de objetos digitais;

e) melhoria na experiência de *e-commerce*.

5.7.3.2 *Caracterização*

Podemos citar como características da folksonomia:

a) resultado de uma indexação livre feita pelo próprio usuário do recurso, ou seja, não é o autor nem o profissional de indexação que indexam o recurso da *web*;

b) desenvolvida em um ambiente aberto que possibilita o compartilhamento;

c) a atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros);

d) a indexação é livre – baseada na linguagem natural;

e) não possui controle de vocabulário;

[64] Disponível em: <www.columbia.edu/cu/libraries/.../folksonomies.ppt>.

f) não possui regras e/ou políticas de indexação;

g) reflete as escolhas linguísticas e terminológicas dos usuários;

h) é rapidamente adaptável às mudanças nas necessidades de informação dos usuários e do próprio vocabulário;

i) a adição de novos termos à folksonomia não implica adição de custo, nem para o usuário, nem para o sistema;

j) não é necessariamente oposta aos vocabulários controlados, pode ser complementar a eles ou derivá-los, estabelecendo uma linguagem mais próxima do usuário;

k) permite representar assuntos ou quaisquer outros elementos de metadados, tais como tipo ou formato;

l) instrumento de natureza dinâmica que se adapta ao funcionamento dos recursos informacionais;

m) as barreiras para cooperação são mínimas;

n) não possui relação claramente definida entre os termos do vocabulário, diferentemente do que acontece em outras linguagens de indexação, nas quais há várias espécies de relações (ex.: termo geral, termo específico e termos relacionados).

5.7.3.3 *Vantagens e desvantagens da folksonomia*

Observe, no Quadro 19, algumas vantagens e desvantagens da folksonomia:

Quadro 19 - Folksonomia: vantagens e desvantagens

Vantagens	Desvantagens
Forma de organizar a *web* com bom índice de custo-benefício.	Não possui o mesmo nível de clareza que um vocabulário controlado.
Benefícios sociais: trabalho colaborativo em que todos aproveitam o resultado.	Sistema polissêmico (mesma palavra com vários significados).
Adaptável a vários ambientes com sucesso.	Termos gerais imprecisos, impactando negativamente a recuperação da informação.
Sistema que permite a participação de qualquer indivíduo.	Não possui controle de sinônimos, de homógrafos e da forma da palavra (ex.: singular e plural), criando cabeçalhos redundantes.
Provê resposta individual e social de imediato.	Não possui regras de sintaxe.
	Não possui notas de escopo para contextualização do termo.
	Não possui rede de referências cruzadas para agrupar termos relacionados.

Fonte: produção do próprio autor (2017).

As folksonomias são formadas por uma coleção agregada de *tags* e possuem estrutura categorial indutiva (partindo de termos específicos para categorias).

Explicativo

Não confunda *tags* com palavras-chave!

Figura 73 - *Tags*

Fonte: *Wikimedia Commons* (2011).[65]

Às vezes, estes dois termos – *tag* e palavra-chave – são usados um no lugar do outro. É bom saber a diferença entre eles:

a) palavras-chave fazem parte de um sistema, podendo ser padronizadas por um vocabulário, e têm o papel de auxiliar na identificação e na recuperação da informação, não estando linkadas diretamente a um objeto digital específico. Por outro lado, *tags* são frequentemente agregadas a objetos digitais específicos, tendo a função de recuperá-los;

b) palavras-chave geralmente não são hiperlinkadas, pois podem servir para indexar objetos físicos que não estão presentes digitalmente no sistema;

c) palavras-chave implicam recuperação de informação, *tags* implicam linkagem e recuperação de objetos digitais;

d) *tags* são muito mais para conectar itens do que para categorizar assuntos;

e) *tags* não têm natureza hierárquica, mas podem contribuir para um sistema hierárquico;

f) *tags* são meios de criar *links* entre itens através da criação de conjuntos de objetos digitais;

g) *tags* contribuem para a conexão entre pessoas com interesses comuns.

[65] Disponível em: <https://commons.wikimedia.org/wiki/File:Nube-etiquetas.png>.

5.7.3.4 *Outras considerações*

A folksonomia pode ser usada, também, para organizar recursos para uma intranet. Outra possibilidade de uso é valer-se do conhecimento do usuário para organizar catálogos de bibliotecas: os clientes podem organizar e "taguear" itens de seu interesse a partir do catálogo ou de fontes externas, se disponíveis.

A folksonomia é, sem dúvida, alvo de problemas típicos de qualquer vocabulário sem controle terminológico, mas sua popularidade crescente sugere que pessoas estejam motivadas e interessadas em atribuir suas próprias *tags* a itens de seu interesse.

5.7.4 **Atividade**

Frederick Wilfrid Lancaster foi professor da *Graduate School of Library and Information Science da Universidade de Illinois* (EUA), onde lecionou cursos de armazenamento e recuperação da informação, entre outros. Em um de seus textos, em que fala sobre vocabulários controlados e vocabulários sem controle, ele faz a seguinte afirmação:

"Sistemas híbridos" – o termo 'híbrido', no contexto da Biblioteconomia e Ciência da Informação, refere-se a sistemas de recuperação da informação operando em uma combinação de termos controlados e linguagem natural. Incluem tanto sistemas nos quais termos são atribuídos por indexadores humanos, como sistemas nos quais a base de dados pode ser pesquisada em uma combinação de termos controlados, atribuídos por humanos, ou palavras da linguagem natural ocorridas em títulos, resumos ou na mente do pesquisador.

Desde os primeiros dias de sistemas baseados em computadores, nos anos 1950, existe uma tendência evidente para a simplificação da recuperação da informação. Cada vez mais, torna-se amplamente reconhecido que é possível operar sistemas efetivamente com o nível mínimo de vocabulário controlado ou sem nenhum controle."

Diante da afirmação do autor e com o que foi aprendido até agora, dê sua opinião sobre a linguagem livre, vocabulários controlados ou vocabulários híbridos para tratamento e recuperação da informação: o que você acha mais eficaz e por quê?

Resposta comentada

Figura 74 - *F. W. Lancaster*

Fonte: *Pinterest* (20--?).[66]

Concordamos com *Lancaster*, por considerarmos que os sistemas híbridos são uma boa solução para a recuperação da informação, e explicamos por quê. Considerados separadamente, tanto a linguagem natural quanto o vocabulário controlado têm vantagens e desvantagens. Para citar apenas uma, a linguagem natural possibilita a recuperação de maior número de itens, porque qualquer palavra que o usuário empregue dentro do contexto de seu interesse pode servir de ponto de acesso para documentos, independentemente de as palavras serem sinônimas ou homógrafas e, exatamente por isso, a quantidade de documentos inúteis a seu interesse será imensa e redundante. Já o vocabulário controlado trará seu resultado de busca com menos itens, contudo mais preciso, pois o controle de sinônimos e homógrafos terá sido feito. Se tais sistemas de recuperação têm vantagens e desvantagens, a combinação dos dois pode resultar em muitos benefícios. O sistema híbrido oferecerá duas opções de busca: ela poderá ser feita por meio de consulta ao vocabulário controlado e por meio da linguagem natural. Dessa forma, a recuperação será, ao mesmo tempo, precisa e abrangente.

5.8 ESCOLHENDO UMA LINGUAGEM ADEQUADA PARA CADA CONTEXTO DE UM SRI

Nesta Unidade, foram vistas linguagens documentárias diferentes das mostradas anteriormente neste Curso, que vêm sendo aplicadas em ambientes institucionais. Cada uma delas possui características, algumas peculiares, outras comuns a todas as demais.

[66] Disponível em: <https://br.pinterest.com/pin/445363850619996739/>.

No Quadro 20, a seguir, você poderá conferir uma síntese dessas características, que poderá auxiliá-lo a distinguir uma linguagem da outra, facilitando a resolução de um problema encontrado na prática de atividades de um SRI.

Quadro 20 - Síntese comparativa de características do *Vocabulário Sistematizado*, do *Soggettario*, do *MeSH* e das folksonomias

Vocabulário / Característica da indexação	***Vocabulário Sistematizado***	***Soggettario***	***MeSH***	**Folksonomia**
	Pós-coordenada	Pré-coordenada e pós-coordenada	Pré-coordenada e pós-coordenada	Pós-coordenada
Linguagem	Controlada	Controlada	Controlada	Livre
Unidade de representação	Conceito	Conceito	Conceito	Palavra
Componentes	Vocabulário e regras para aplicação	Vocabulário, regras para aplicação e sintaxe	Vocabulário e regras para aplicação	Vocabulário
Forma de apresentação	Alfabética e sistemática	Alfabética e sistemática	Alfabética e sistemática	Alfabética
Abrangência	Multidisciplinar	Multidisciplinar	Especializado na área biomédica	Multidisciplinar ou especializada
Compartilhamento de informação	Institucional/ interinstitucional	Institucional/ interinstitucional	Institucional	Popular
Navegação	Hierárquica	Hierárquica	Hierárquica	Não hierárquica
Acesso ao documento	Indireto	Indireto	Direto/indireto	Direto
Organização do conhecimento	Apoio em princípios teórico-metodológicos	Apoio em princípios teórico-metodológicos	Apoio em princípios teórico-metodológicos	Apoio no conhecimento do usuário sobre o assunto
Medida de recuperação da informação	Precisão	Alta precisão	Precisão	Baixa precisão
Mecanismo de refinamento de busca	Presença de estrutura hierárquica	Presença de sintaxe	Navegação taxonômica	Organização categorial
Produção do trabalho	Colaboração institucional/ interinstitucional	Colaboração institucional/ interinstitucional	Colaboração institucional/ interinstitucional	Colaboração popular

Fonte: produção do próprio autor (2017).

Figura 75 - Não assuma que um tamanho serve para tudo!

Fonte: *Pixabay* (2016).[67]

No Quadro 20, você pôde verificar a variedade de características das linguagens abordadas. Isso revela que nem todas servem aos mesmos objetivos. Cada SRI deve desenvolver seus sistemas de linguagem, tendo em vista os objetivos da instituição, sua clientela e os serviços que pretende oferecer. Embora os recursos para o desenvolvimento de linguagens nas instituições geralmente sejam reduzidos, não é aconselhável a "importação" de uma linguagem desenvolvida para outra instituição com realidades diferentes.

Sendo assim, é bom ficar atento a situações em que você talvez tenha que tomar decisões ou participar delas, e observar as características dos trabalhos construídos utilizando novas metodologias. Possivelmente uma delas poderá nortear a linguagem que você venha a desenvolver.

Multimídia

Caso você queira conhecer mais sobre as experiências abordadas nesta Unidade, como detalhes sobre suas metodologias, por exemplo, você pode navegar nos *links* a seguir:

a) para o *Vocabulário Sistematizado*:

<http://www.casaruibarbosa.gov.br/dados/DOC/artigos/a-j/FCRB_Vocabulario_sistematizado_a_experiencia_da_Fundacao_Casa_de_Rui_Barbosa.pdf>;

<http://www.lbd.dcc.ufmg.br/colecoes/ontobras/2010/0023.pdf>;

[67] Disponível em: <https://pixabay.com/pt/fruta-apple-fundo-branco-branco-1160552>.

<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/545/O%20vocabul%C3%A1rio%20-%20Almeida.pdf?sequence=5>;

b) para o *MeSH*:

<http://www.nlm.nih.gov/MeSH/> (no idioma inglês);

c) para o *Nuovo Soggettario*:

<http://thes.bncf.firenze.sbn.it/index.html> (no idioma italiano);

d) como exemplo de plataformas que se utilizam de folksonomia para facilitar o agrupamento e a recuperação do conteúdo compartilhado, observe:

<https://www.instagram.com/?hl=pt-br>;

<https://www.youtube.com/?gl=BR&hl=pt>.

5.8.1 **Atividade**

Imagine que você foi contratado para organizar a coleção de uma biblioteca pública. Ao chegar lá, você se dá conta de que só existe uma lista de palavras usadas como indexadores dos documentos (palavras soltas, cabeçalhos de assunto, expressões, entre outras). A recuperação da informação está problemática. Dificilmente o usuário encontra o que quer. Você sente que precisa tomar uma decisão quanto àquela lista.

Considere que sua biblioteca tem uma coleção geral, para atender a um público amplo, que merece ter respostas mais precisas. Nessa situação, qual das experiências abordadas nesta Unidade você acha que poderia servir de exemplo para a nova linguagem a ser desenvolvida? Consulte o texto, faça sua escolha e justifique-a no espaço a seguir.

Resposta comentada

Eu não pensaria nem no *MeSH*, nem em uma folksonomia. O primeiro foi desenvolvido para atender acervos especializados na área biomédica e poderia ficar muito longe da sua realidade. Já o segundo tem como característica principal o uso da linguagem natural, que recupera muitos itens, mas sem precisão. Linguagem natural você já tem na lista que encontrou.

Sobraram as experiências do *Vocabulário Sistematizado* e do *Soggettario*. Ambos servem para acervos gerais e são baseados em conceitos, o que garantiria precisão na recuperação. Se você quer dar mais flexibilidade a seus usuários, para que manipulem a busca segundo seus interesses, as duas experiências seriam válidas, já que ambas baseiam-se em recuperação pós-coordenada. A única diferença é que o *Soggettario*, além do vocabulário e regras para aplicação, inclui uma sintaxe, para fixação da ordem dos elementos usados na linguagem. Isso acarreta mais gente envolvida na construção do vocabulário e, consequentemente, maior custo.

5.9 CONSIDERAÇÕES FINAIS

Chegando ao final desta Unidade, é bom observar algumas condições ideais para o desenvolvimento de SRIs:

a) os vocabulários usados para a indexação deveriam também ser disponibilizados para os usuários;

b) o sistema deveria permitir o acesso direto ao documento ou, pelo menos, ligar o assunto a seu registro na base;

c) seria útil usar folksonomias cuidadosa e estrategicamente para incrementar a recuperação nos SRI e aproveitar seus benefícios, explorando formas de combinar as duas linguagens: controlada e livre. Ou seja, o sistema deveria acatar e retrabalhar as sugestões de seus usuários quanto ao vocabulário. Para isso, seria desejável haver uma coordenação central de uma ou mais pessoas destinadas à mediação de resoluções a tomar quanto à escolha do vocabulário adequado;

d) vocabulários devem ser constantemente atualizados, acompanhando a evolução da linguagem e do conhecimento;

e) é importante considerar os objetivos da indexação. Nos sistemas desenvolvidos por uma instituição, os usuários e os objetivos da própria instituição são os primeiros aspectos a serem levados em conta. Porém, no caso das folksonomias, as pessoas indexam coisas, primeiramente, porque querem encontrá-las, não por querer que outros as encontrem. Esse fato impactará a qualidade dos termos e sua frequência de ocorrência;

f) vocabulários controlados poderiam ser mais bem utilizados do que comumente são. Por exemplo, as estruturas de assunto

são subutilizadas na Biblioteconomia porque os vocabulários controlados existentes não estão sendo exportados para a *web*;

g) deve-se considerar, ainda, que os termos bem conectados facilitam as descobertas, por isso, as linguagens de recuperação mais modernas disponibilizam sistemas de referências cruzadas, por meio das quais vários termos possam, eventualmente, despertar interesse numa busca, possibilitando sua expansão.

5.10 CONCLUSÃO

A despeito de os instrumentos de representação temática da informação vistos nesta Unidade adotarem, quer controle vocabular rígido (como no caso do *Vocabulário Sistematizado*, do *MeSH* e do *Nuovo Soggettario*), quer linguagem livre (como no caso das folksonomias), ou ambos (sistemas híbridos), a necessidade de indexação dos documentos para que sejam recuperados é evidente.

Apesar das diversas iniciativas existentes – algumas citadas nesta Unidade –, é sabido que a indexação e recuperação precárias permanecem na maior parte dos SRI. Os motivos podem ser vários, desde a consciência da necessidade de tais instrumentos até questões de custo-benefício de sua implementação, passando, ainda, por questões técnicas.

Mesmo entendendo que os benefícios podem ser muitos para os sistemas, algumas instituições consideram somente o aspecto do custo elevado. Para minimizar o alto custo de tais iniciativas, seria oportuno enfatizar que uma boa medida para a construção de linguagens de recuperação é o trabalho colaborativo em rede, que deve acontecer entre instituições ou pessoas com interesses em comum.

Nessa direção, dentro do que foi visto nesta Unidade, relembramos os exemplos aqui citados das experiências do *Nuovo Soggettario* e da folksonomia. É sabido que outras experiências iniciadas (casos de *Vocabulários Sistematizados*, por exemplo) foram interrompidas por falta de recursos humanos e/ou financeiros, ou ainda pela dissolução das redes cooperativas. Além disso, para que as linguagens documentárias tenham sua utilidade garantida no decorrer do tempo, faz-se necessário que elas sejam atualizadas permanentemente, acompanhando a dinâmica do conhecimento e seu consequente impacto na transformação da linguagem.

RESUMO

O advento do computador possibilitou o múltiplo acesso à informação, ampliando a probabilidade de recuperar um documento num SRI. Isso levou ao aparecimento de sistemas pós-coordenados, permitindo a

conjugação dos termos na hora da busca e conferindo maior flexibilidade aos usuários para encontrar a informação desejada.

Assim, o cabeçalho de assunto, como tradicionalmente usado, teve sua importância limitada pela nova tecnologia. Porém, muitas bibliotecas continuam a usá-lo, embora ele restrinja bastante as possibilidades de recuperação. Esse panorama fez com que surgissem novas linguagens documentárias, dentre as quais aquelas que se aproveitaram do cabeçalho de assunto já instituído.

Dentre as experiências conhecidas, encontram-se o *Vocabulário Sistematizado* (experiências nacionais desenvolvidas para acervos gerais), o *Medical Subject Headings (MeSH)* (experiência americana para a área médica) e o *Nuovo Soggettario* (experiência italiana para acervos gerais).

Outra experiência moderna de linguagem, que foge aos princípios do vocabulário controlado, sendo usada, sobretudo, para recuperação da informação, é a folksonomia. Trata-se de uma experiência colaborativa, que usa linguagem livre, em que o próprio usuário sugere os termos de indexação – *tags*.

Sugestão de Leitura

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na web. **Datagramazero**, Rio de Janeiro, v. 8, n. 3, 2007.

DIAS, Elizabeth Abib Vasconcelos; BECKER, Ilva Pereira Lima. **Vocabulário controlado:** uma experiência interdisciplinar no controle terminológico do Sistema de Bibliotecas da UFF. Niterói: [s.n], 2010.

LUCARELLI, Anna. **Authority control semantico e nuovo Soggettario**. [S.l.: s.n., 20--?]. Disponível em: <http://eprints.rclis.org/bitstream/10760/4167/1/lucarelli_ita.pdf>. Acesso em: 15 jan. 2015.

LUCARELLI, Anna; BULTRINI, Leda; CHETI, Alberto. **Il Nuovo Soggettario italiano:** dallo studio al progetto. [S.l.: s.n.], 2004. Disponível em: <http://www.bncf.firenze.sbn.it/documenti/nsog.pdf>. Acesso em: 15 jan. 2015.

SAVAGE, Allan. Changes in MeSH data structure. **NLM Tech Bull**, Maryland, v. 313, p. 2, 2000.

REFERÊNCIAS

ALMEIDA, Tatiana de; SOUZA, Rosali Fernandes. O vocabulário controlado como instrumento de organização e representação da informação na FINEP. **Organização e Representação do Conhecimento**, [S.l.], p. 424, 2011. Disponível em: <http://hdl.handle.net/123456789/545>. Acesso em: 18 jan. 2015.

BIBLIOTECA NAZIONALE CENTRALE FIRENZE. **Nuovo soggettario thesaurus:** manual e applicativo. Firenze: Biblioteca Nazionale Centrale Firenze, 2012. Disponível em: <http://thes.bncf.firenze.sbn.it/Manuale_applicativo.pdf>. Acesso em: 10 jan. 2015.

BIBLIOTECA NAZIONALE CENTRALE FIRENZE. **Nuovo soggettario thesaurus:** Versatilità per l"indicizzazione di documenti di varia natura. Firenze: Biblioteca Nazionale Centrale Firenze, 2007. Disponível em: <http://thes.bncf.firenze.sbn.it/ricerca.php>. Acesso em: 10 jan. 2015.

BRASIL, Maria Irene et al. Vocabulário Sistematizado*:* a experiência da Fundação Casa Rui Barbosa. In: INTEGRAR: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS, 1., 2002, São Paulo. **Anais...** São Paulo: Fundação Casa Rui Barbosa, 2002. p. 81-94.

HOUAISS, Antônio. **Dicionário Houaiss eletrônico:** versão 3.0. Rio de Janeiro: Objetiva, 2007.

LANCASTER, F. W. **Vocabulary control for information retrieval**.Tradução de M. A. R. Piedade. Washington: Information Resources Press, 1972.

LANCASTER, F. W. **Vocabulary control for information retrieval.** 2nd ed. Arlington: Information Resources Press, 1986.

MOTTA, Dilza Fonseca da. **Método relacional como nova abordagem para a construção de tesauros**. 1987. 90 f. Dissertação (Mestrado em Ciência da Informação) – UFRJ/IBICT, Rio de Janeiro, 1987.

MOTTA, Dilza Fonseca da et al. Vocabulário Sistematizado de linha conceitual para as bibliotecas do SESC: uma ferramenta cliente-orientada. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 22., 2009, Bonito. **Anais...** São Paulo: FEBAB, 2009.

WAL,T.V. **Folksonomy coinage and definition**. Maryland: [s.n.], 2007. Disponível em:<http://vanderwal.net/folksonomy.html>. Acesso em: 1 abr. 2017.

Universidade Federal
do Rio de Janeiro

CAPES MINISTÉRIO DA EDUCAÇÃO

BRASIL GOVERNO FEDERAL

UAB UNIVERSIDADE ABERTA DO BRASIL

Agência Brasileira do ISBN
ISBN 978-85-85229-24-5
9 788585 229245

Agência Brasileira do ISBN
ISBN 978-85-85229-25-2
9 788585 229252