

# 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 tesouros;
  - 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).<sup>51</sup>

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.

<sup>52</sup> 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).<sup>53</sup>

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.

<sup>53</sup> 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 tesouros – 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 tesouro 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 tesouro: *Vocabulário Sistematizado*, *MeSH* e *Nuovo Soggettario*. Vamos conhecê-las?

## 5.6 VOCABULÁRIO SISTEMATIZADO

Figura 62 - Repensando a recuperação da informação



Fonte: Pixabay (2016).<sup>54</sup>

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*?

<sup>54</sup> 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 tesouros, 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 tesouros. 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.<sup>a</sup> *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 tesouros) 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 tesouros a partir do cabeçalho de assunto, podem ser identificadas nos seguintes pontos:

- abandono da distinção entre “termo principal” e “subdivisão”: todas as subdivisões são transformadas em termos;
- adesão às normas internacionais para o controle terminológico e indicação das relações semânticas (*ISO 25964-2011*, norma internacional para tesouros e interoperabilidade com outros vocabulários);
- adoção e indicação das relações próprias de um tesouro (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 (“descriptor”) 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 tesouro destinado a uma área específica do conhecimento, isso não aconteceria; as relações seriam somente entre os termos da área coberta.



## Explicativo

Assim como nos tesouros, 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 tesouros, 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 tesouro, 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) Tesouro
- (3) *Vocabulário Sistematizado*
- (4) Cabeçalho de assunto e *Vocabulário Sistematizado*
- (5) Tesouro e *Vocabulário Sistematizado*
- (6) Todas as linguagens
- (7) Linguagem natural
- (8) Cabeçalho de assunto, Tesouro 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 predefinida.
- (2) Por definição, o tesouro 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 discursivo.
- (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 tesouro 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 tesouro 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 tesouros e que dissemos que as regras para a construção de um tesouro 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 tesouros.

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 lembra, 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 tesouro (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).<sup>55</sup>

<sup>55</sup> 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--?).<sup>56</sup>

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).

<sup>56</sup> 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).<sup>57</sup>

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:

- sistema desenvolvido a partir de cabeçalhos de assunto;
- sistema atualmente conceito-orientado (cuja metodologia considera conceitos, e não palavras);
- unidade de representação: conceito;
- conceitos dispostos alfabética e hierarquicamente;
- sistema parcialmente pré-coordenado e pós-coordenado;
- existência de rede de remissivas;
- existência de regras para aplicação do vocabulário;

<sup>57</sup> 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](#)

<b>MeSH Heading</b>	Cardiovascular Physiological Phenomena
<b>Tree Number</b>	<a href="#">G09.330</a>
<b>Annotation</b>	general or unspecified; prefer specifics; DF: CARDIOVASC PHYSIOL
<b>Scope Note</b>	Processes and properties of the <a href="#">CARDIOVASCULAR SYSTEM</a> as a whole or of any of its parts.
<b>Entry Term</b>	Cardiovascular Physiological Concepts
<b>Entry Term</b>	Cardiovascular Physiological Phenomenon
<b>Entry Term</b>	Cardiovascular Physiology
<b>Entry Term</b>	Physiology, Cardiovascular
<b>Allowable Qualifiers</b>	<a href="#">DE</a> <a href="#">GE</a> <a href="#">IM</a> <a href="#">RE</a>
<b>Previous Indexing</b>	<a href="#">Cardiovascular System</a> /physiology (1966-1997)
<b>History Note</b>	2009 (1998)
<b>Date of Entry</b>	19990101
<b>Unique ID</b>	D002320

### MeSH Tree Structures

Fonte: *National Library of Medicine* (20--?).<sup>58</sup>

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).

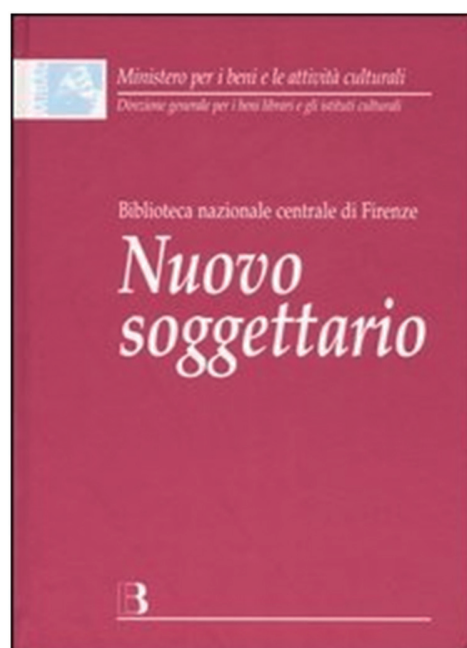
<sup>58</sup> Disponível em: <<http://www.nlm.nih.gov/MeSH/introduction.html>>.

### 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--?).<sup>59</sup>

O projeto *Nuovo Soggettario* é desenvolvido pela *Biblioteca Nazionale Centrale di Firenze (BNCF)*, dado seu papel institucional de arquivo nacional do livro e de agência bibliográfica nacional, responsável pela *Bibliografia Nazionale 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 Nazionale 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.

<sup>59</sup> Disponível em: <[https://upload.wikimedia.org/wikipedia/it/2/2e/Bibliografica\\_Nuovo\\_soggettario.jpg](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 tesouro geral, 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**spaço 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).<sup>59</sup>

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;

<sup>60</sup> Disponível em: <<https://pixabay.com/pt/sistema-rede-noticias-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 Soggettario* 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 Soggettario* 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*

The screenshot shows the search results for 'Tomografia assiale computerizzata' on the Nuovo Soggettario Thesaurus website. The page includes a navigation menu, search options, and a detailed entry for the term. A legend box highlights the following information:

- UF:** Usado por
- TT:** Categoria
- BT:** Termo abrangente
- RT:** Termo relacionado

The entry for 'Tomografia assiale computerizzata' includes the following details:

- Category:** Azioni:Attività
- UF:** TAC <Tomografia assiale computerizzata>
- TT:** Attività
- BT:** Tomografia
- RT:** Elaboratori, Raggi X, Tomografia computerizzata
- Definition:** Tecnica radiografica in cui l'organo in esame viene esplorato trasversalmente a varie profondità da un sottilissimo fascio di raggi X, da cui risulta una serie di immagini che permettono l'identificazione di piccole lesioni non apprezzabili con la radiologia comune (DeM)
- Reference:** BNI 1956-1985; VLJ; DeM; Zin; Dev; Dorland; ESG; MESH: Tomography, X-Ray Computed; LIUC; DDC21; Wikipedia(IT)
- Dewey Classification (Ed. 21):** 616.07572
- Partners:** BNI
- Record status:** Termine strutturato
- Id.:** 20395

Fonte: *Soggettario* (2007).<sup>61</sup>

<sup>61</sup> 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.

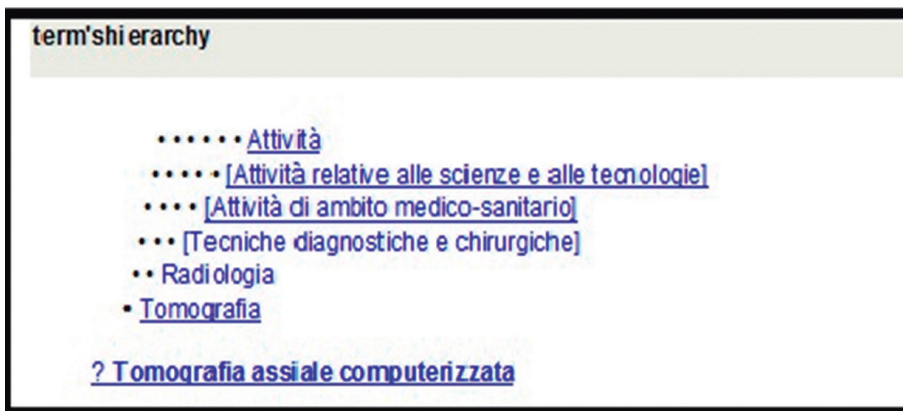
**Explicação:** Veneza é o conceito-chave e o objeto único de tratamento; a indicação de que se trata de um guia turístico desempenha um papel complementar de forma bibliográfica.

**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).<sup>62</sup>

### 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*

<sup>62</sup> 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”?

### 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.

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.

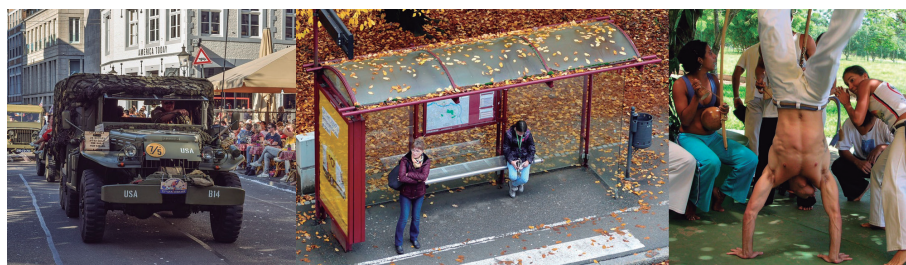
A folksonomia é também conhecida como: “classificação social”, “etiquetagem social”, “classificação popular”, “taxonomia popular”, “etno-classificação”, “classificação compartilhada”, “etiquetagem aberta”, “etiquetagem livre”, entre outras denominações.



## Explicativo

### Pegando carona no conceito de “parada”

Figura 71 - Qual o conceito de “parada”?



Fonte: Pixabay e Freeimages (2020).<sup>63</sup>

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.

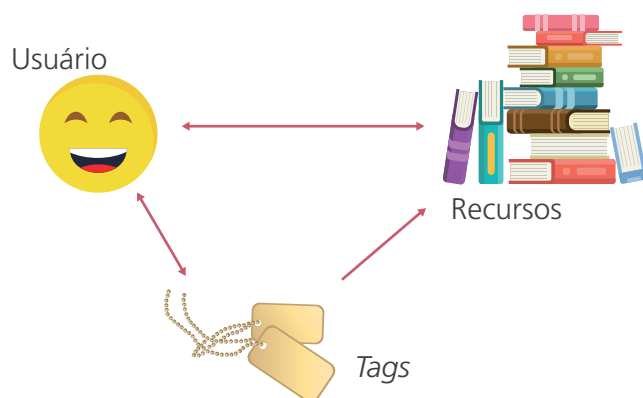
Na figura exibida, os três objetos têm a mesma denominação, mas se referem a objetos diferentes. Uma linguagem documentária com mecanismos de controle vocabular certamente daria conta de explicitar o significado do(s) objeto(s), orientando corretamente a pesquisa do usuário. Em uma folksonomia, em princípio, tal controle vocabular não é observado, uma vez que esses sistemas utilizam

<sup>63</sup> 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*)  
**Modelo básico para etiquetagem de sistemas**



Fonte: Columbia University (20--?).<sup>64</sup>

A folksonomia tem vários objetivos, dentre os quais:

- recuperação da informação a posteriori;
- gerenciamento de informação pessoal;
- social bookmarking* – compartilhamento público de *links*, resultando em uma rede de *tags* criada pelos usuários;
- reunião e compartilhamento de objetos digitais;
- melhoria na experiência de *e-commerce*.

### 5.7.3.2 Caracterização

Podemos citar como características da folksonomia:

- 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*;
- desenvolvida em um ambiente aberto que possibilita o compartilhamento;
- a atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros);
- a indexação é livre – baseada na linguagem natural;
- não possui controle de vocabulário;

<sup>64</sup> Disponível em: <[www.columbia.edu/cu/libraries/.../folksonomies.ppt](http://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 <i>web</i> 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).<sup>65</sup>

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.

<sup>65</sup> 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?

---

---

---

---

---

---

---

---

---

---

---

---



Figura 74 - F. W. Lancaster



Fonte: Pinterest (20--?).<sup>66</sup>

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.

<sup>66</sup> 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*

<b>Vocabulário</b>	<b><i>Vocabulário Sistematizado</i></b>	<b><i>Soggettario</i></b>	<b><i>MeSH</i></b>	<b>Folksonomia</b>
<b>Característica da indexação</b>	Pós-coordenada	Pré-coordenada e pós-coordenada	Pré-coordenada e pós-coordenada	Pós-coordenada
<b>Linguagem</b>	Controlada	Controlada	Controlada	Livre
<b>Unidade de representação</b>	Conceito	Conceito	Conceito	Palavra
<b>Componentes</b>	Vocabulário e regras para aplicação	Vocabulário, regras para aplicação e sintaxe	Vocabulário e regras para aplicação	Vocabulário
<b>Forma de apresentação</b>	Alfabética e sistemática	Alfabética e sistemática	Alfabética e sistemática	Alfabética
<b>Abrangência</b>	Multidisciplinar	Multidisciplinar	Especializado na área biomédica	Multidisciplinar ou especializada
<b>Compartilhamento de informação</b>	Institucional/ interinstitucional	Institucional/ interinstitucional	Institucional	Popular
<b>Navegação</b>	Hierárquica	Hierárquica	Hierárquica	Não hierárquica
<b>Acesso ao documento</b>	Indireto	Indireto	Direto/indireto	Direto
<b>Organização do conhecimento</b>	Apoio em princípios teórico-metodológicos	Apoio em princípios teórico-metodológicos	Apoio em princípios teórico-metodológicos	Apoio no conhecimento do usuário sobre o assunto
<b>Medida de recuperação da informação</b>	Precisão	Alta precisão	Precisão	Baixa precisão
<b>Mecanismo de refinamento de busca</b>	Presença de estrutura hierárquica	Presença de sintaxe	Navegação taxonômica	Organização categorial
<b>Produção do trabalho</b>	Colaboração institucional/ interinstitucional	Colaboração institucional/ interinstitucional	Colaboração institucional/ interinstitucional	Colaboração popular

Fonte: produção do próprio autor (2017).



Figura 75 - Não assumo que um tamanho serve para tudo!



Fonte: Pixabay (2016).<sup>67</sup>

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.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>>;

<sup>67</sup> Disponível em: <<https://pixabay.com/pt/fruta-apple-fundo-branco-branco-1160552>>.



## 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, conseqüentemente, maior custo.

Semestre

4

## 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 conseqüente 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](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 soggettoario thesaurus**: manual e aplicativo. Firenze: Biblioteca Nazionale Centrale Firenze, 2012. Disponível em: <[http://thes.bncf.firenze.sbn.it/Manuale\\_applicativo.pdf](http://thes.bncf.firenze.sbn.it/Manuale_applicativo.pdf)>. Acesso em: 10 jan. 2015.

BIBLIOTECA NAZIONALE CENTRALE FIRENZE. **Nuovo soggettoario 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 tesouros**. 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**



**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