

CAPAGIIC-Saúde:

Gestão, Informação, Inovação e Conhecimento

ORGANIZADORES

ELIANE LOURDES DA SILVA MORO

LIZANDRA BRASIL ESTABEL

ARIEL BEHR

FILIFE XERXENESKI DA SILVEIRA

GABRIELA FERNANDA CÉ LUFT

SÉRGIO WESNER VIANA



Porto Alegre

2022

Copyright © 2022

Permitida a reprodução sem fins lucrativos, parcial ou total,
por qualquer meio, desde que citada a fonte.

Tiragem:

2.000 exemplares

Revisão:

Gabriela Fernanda Cé Luft e
Vanessa Cristiane Vanzan de Oliveira

Produção Gráfica e Impressão:

Evangraf

Dados Internacionais de Catalogação na Publicação

C236

Capagiic-Saúde: gestão, informação, inovação e conhecimento /
Eliane Lourdes da Silva Moro ... [et al.], organizadores. -
Porto Alegre : Evangraf, 2022.

344 p. : il. color. ; 16x23 cm.

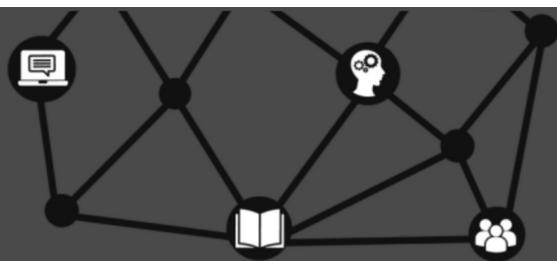
ISBN 978-65-5699-182-5

1. Educação a distância. 2. Informação em saúde. 3. Gestão.
4. Inovação. 5. Conhecimento. 6. Universidade Federal do Rio
Grande do Sul, Grupo de Pesquisa Leia.

CDU 37.018.43

Catalogação: Bibliotecário Filipe Xerxeneski da Silveira – CRB-10/1497

Impresso no Brasil



INSTRUMENTOS DE REPRESENTAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO: TAXONOMIAS, TESAUROS, ONTOLOGIAS E TERMINOLOGIAS

DIRCE MARIA SANTIN

1 INTRODUÇÃO

Informação e conhecimento são elementos centrais na civilização humana e têm papel essencial na atualidade, uma época caracterizada pela chamada Sociedade da Informação. (CASTELLS, 2009). A Organização e Representação do Conhecimento e da Informação assumem importância nesse contexto, como áreas estratégicas para lidar com esses elementos, tanto na Ciência da Informação e em outras áreas, dada a multidisciplinaridade desses objetos e sua importância nas diversas áreas do conhecimento.

Existe certa ambiguidade no uso dos termos Organização e Representação do Conhecimento e da Informação. Conhecimento e informação também são conceitos distintos: enquanto o conhecimento pode ser considerado o conteúdo da consciência humana, resultado da cognição, a informação é uma forma material do conhecimento, que existe fora da consciência humana e exerce sua função por meio de um suporte. (BRASCHER; CAFÉ, 2008).

A Organização e Representação da Informação e do Conhecimento têm distinções de caráter teórico e prático, mas estão em constante interação e possuem fortes intersecções em relação aos seus objetos e se referem aos conceitos do conhecimento, e não

aos registros de informação. Visam estruturar, de forma sistemática, as unidades do conhecimento (conceitos) com base em seus elementos (características) e a aplicação desses conceitos a assuntos ou objetos. (DAHLBERG, 2006). Hjørland (2008) divide a Organização do Conhecimento em dois grupos: a organização intelectual do conhecimento, que utiliza conceitos, sistemas conceituais e teorias; e a organização social do conhecimento, que compreende profissões, negócios e disciplinas.

2 A ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO

A Organização do Conhecimento refere-se à construção de sistemas conceituais, enquanto a Organização da Informação compreende a unidade desses sistemas como objetos de informação. (DAHLBERG, 2006). Brascher e Café (2008) distinguem dois processos de organização:

- a) Organização do Conhecimento, aplicada a unidades do conhecimento (conceitos) e à construção de modelos de mundo que se constituem em abstrações da realidade;
- b) Organização da Informação, aplicada às ocorrências individuais de objetos informacionais e voltada à organização desses objetos para arranjá-los sistematicamente.

A Organização e Representação da Informação abrangem as atividades de descrição, representação, indexação e classificação, além dos sistemas de informação e de organização do conhecimento. O objetivo desses processos é possibilitar o acesso à informação por meio da estruturação de seus elementos e envolve a descrição física e de assunto dos objetos informacionais, cujo produto é a representação da informação. Enquanto a organização compreende a análise, classificação e ordenação da informação, a representação está associada aos objetos de informação, à materialização e ao registro da simbologia que substituem os objetos ou ideias. Os dois processos visam à recuperação da informação.

A Organização do Conhecimento e da Informação produzem, segundo Brascher e Café (2008), dois tipos de representação:

- a) Representação do conhecimento: estrutura conceitual que representa modelos de mundo, os quais permitem descrever e fornecer explicações sobre os fenômenos da natureza;
- b) Representação da informação: conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo.

A Representação do Conhecimento busca materializar o que ocorre na mente humana e na dinâmica do conhecimento. Para tanto, utiliza modelos de representação dos processos que envolvem a aprendizagem, o raciocínio e as relações entre os conhecimentos. O conhecimento pode ser representado por taxonomias, ontologias, mapas mentais e mapas conceituais, inteligência artificial, modelos matemáticos e visuais, entre outros. (LIMA; ALVARES, 2012).

A Representação da Informação, por sua vez, envolve os processos de análise e síntese de objetos informacionais para gerar formas de representação que permitam identificá-los e recuperá-los a partir de seus atributos e características. Trata-se de um processo que explicita os dados de identificação dos produtores, do contexto de produção, das características e do conteúdo dos objetos, com vistas à recuperação da informação.

3 LINGUAGENS DOCUMENTÁRIAS

As linguagens documentárias são instrumentos para a Organização e a Representação da Informação. São construções artificiais ou convenções criadas para facilitar o conhecimento de um domínio, regidas por regras de utilização, a exemplo daquelas existentes para a língua natural, da qual representam uma imagem parcial ou reduzida. As linguagens documentárias se utilizam tanto da linguagem natural como da linguagem controlada, com terminologia, padrões e significados próprios para a representação e recuperação da infor-

mação. As linguagens são consideradas instrumentos de representação da informação, a exemplo das classificações, listas de cabeçalhos de assunto, tesouros e ontologias. Os produtos da representação documental são os catálogos, índices e resumos, os quais resultam do processo de representação. (DODEBEI, 2002).

A linguagem natural refere-se ao discurso com o qual nos comunicamos no dia a dia, oral e escrita. As linguagens documentárias se caracterizam pela eliminação da ambiguidade e pelo controle e padronização dos vocabulários. São linguagens controladas, normalizadas, que visam representar, de forma padronizada, o conhecimento de um domínio (área do conhecimento) e favorecer a ligação entre os usuários e o conteúdo dos objetos informacionais.

As linguagens documentárias levam em conta a abrangência do domínio, os relacionamentos de equivalência e as variações linguísticas. São baseadas na estrutura conceitual do domínio, com cada significado representado por uma única forma verbal. Cada conceito é um item de conhecimento do mundo real dos objetos e dos fenômenos, com seus atributos e relacionamentos. (DODEBEI, 2002).

Os conceitos são a essência da realidade, o que faz com que ela não possa ser diferente do que é, ou seja, são unívocos. Nas linguagens documentárias, a unidade informacional é o conceito. Os conceitos estão representados por termos, que são interligados por relações conceituais de equivalência, hierárquicas, partitivas, de oposição e funcional, conforme proposto por Dahlberg (1978) e descrito por Dodebei (2002):

- a) relações lógicas ou de equivalência: ocorrem quando um conceito pode ser representado por vários símbolos distintos ou quando se quer reduzir os níveis de implicação conceitual, como no caso de sinônimos, grafias diferentes, abreviaturas, etc.;
- b) relações hierárquicas: ocorrem quando dois conceitos diferentes possuem características idênticas e um deles possui uma característica a mais, de modo que se estabelece a relação hierárquica ou a relação de gênero-espécie;
- c) relações partitivas: ocorrem entre o todo e suas partes. Em relação

- aos conceitos e suas características, isto significa que o conceito de um todo inclui como características os conceitos de suas partes;
- d) relações de oposição: ocorrem na oposição ou contradição entre conceitos: numérico/não numérico, presença/ausência, entre outros;
 - e) relações funcionais: quando os conceitos expressam processos.

As linguagens documentárias incluem vários tipos de vocabulários controlados, que se diferenciam entre si de acordo com a complexidade. Vocabulários controlados podem ser definidos como um rol de termos explícitos e definidos de forma não ambígua e não redundante. As linguagens documentárias diferem dos sistemas de classificação, que representam as categorias de assunto, classes, divisões e arranjos do conhecimento. São exemplos de sistemas de classificação a Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU), a Classificação de Dois Pontos, de Ranganathan, e a Classificação Bibliográfica, de Bliss, 1935.

4 TAXONOMIAS

As taxonomias são classificações estruturadas hierarquicamente, de forma a representar o conhecimento de determinado domínio ou aplicação, podendo ser compreendidas por homens e máquinas. Uma taxonomia compõe uma árvore hierárquica formada por subclasses, classes e superclasses. Nas classes situam-se os conceitos que compartilham conhecimentos comuns; nas subclasses, os conceitos com características extras que as especificam; e nas superclasses, características que os generalizam. A classe raiz possui a característica geral de um conceito e vai se especializando nas subclasses. Quanto mais subclasses, mais específica é a classificação.

As taxonomias se estabelecem numa relação todo-parte. Para uma classe, a superclasse tem um aspecto de todo, enquanto uma subclasse tem aspecto de parte. Representam relações semânticas hierarquizadas, com cada classe compondo uma entidade de informação que representa o mundo real. Essas classes são ligadas por al-

gum tipo de relação que as une, organiza e diferencia. Esse processo de conexão estabelece uma hierarquia semântica, com as entidades relacionadas num contexto que lhe atribui significados.

O uso mais comum das taxonomias é classificar informações em categorias, qualquer que seja o domínio ou aplicação. No contexto de sistemas de informação, são usadas como instrumentos de indexação e de classificação da informação com vistas à recuperação. No contexto da informação na web, são estruturas conceituais para a Web Semântica, podendo ser descritas como agentes ou motores de busca inteligentes para a recuperação da informação.

5 TESAUROS

Tesauros são vocabulários controlados que operam na transformação da linguagem natural para a artificial. Geralmente, são ligados ao conhecimento de um domínio, de modo que seu contexto é o domínio e não a linguagem natural. A origem etimológica da palavra é do latim *thesaurus*, que vem do grego *thesaurós*, com sentido equivalente à palavra em português. Os tesauros possuem uma característica singular no âmbito das linguagens documentárias, que é a organização do seu campo terminológico. São definidos por Gomes (1990) como linguagens documentárias dinâmicas que contêm termos relacionados semântica e logicamente e que representam, de forma compreensiva, um domínio do conhecimento.

Para Currás (2010), um tesouro é uma linguagem especializada, normalizada, pós-coordenada, usada para fins documentários, na qual os elementos linguísticos que a compõem (termos, simples ou compostos) estão relacionados entre si sintática e sistematicamente. Já para a ANSI/NISO Z39.19:2005, tesouro é um vocabulário controlado de um domínio e está organizado de forma que os relacionamentos entre os termos são dispostos e representados em relacionamentos padronizados e recíprocos.

Os tesauros se caracterizam pelo estabelecimento de conceitos, representados por termos, e das relações entre eles. Diferem das listas de termos, que não apresentam relações intrínsecas e recíprocas,

mas apenas uma ordenação alfabética, e das listas de cabeçalhos, nas quais os termos estão relacionados entre si em processo de pré-coordenação²⁰. Nos tesouros, os termos estão ordenados entre si de forma que permitem a combinação num processo de pós-coordenação²¹. A diferença básica entre os cabeçalhos de assunto e os tesouros refere-se às relações entre os termos e à forma como esses são considerados em cada instrumento. (CURRÁS, 1995).

A estrutura dos tesouros propicia o estabelecimento de relacionamentos equivalentes, hierárquicos e associativos entre os termos, que representam conceitos. Estabelecer relações entre conceitos significa conferir uma posição no sistema de conceitos, ou seja, estabelecer o gênero (termo genérico), a espécie (termo específico) e as possíveis associações entre eles. A apresentação do tesouro pode ser sistemática, com a ordenação de termos em classes ou facetas e a aplicação de um sistema de classificação, ou gráfica, representada em forma de árvore ou setas.

Os termos que compõem um tesouro são extraídos da análise do objeto do qual uma palavra ou conjunto delas se refere, compreendendo as relações entre eles e outros conceitos. Os termos representam conceitos, ou seja, são palavras ou expressões selecionadas para representar um dado significado. Enquanto os termos, que são palavras especializadas inseridas na linguagem especializada, têm apenas um sentido e se relacionam univocamente com o conceito especializado, as palavras podem ter diversos significados. Isso porque as palavras não estão restritas a um domínio, mas ao fenômeno da comunicação da língua natural. (GOMES, 1990).

Uma das principais características dos tesouros são os termos e as relações entre eles, representando conceitos. Os conceitos, segundo Dahlberg (1978), são unidades de conhecimento. É desse conjunto de relacionamentos entre os termos que emerge o caráter semântico dos tesouros. As relações entre os termos são esclarecidas com o

20 Pré-coordenação: controle e coordenação de termos realizados *a priori*, num processo que precede a utilização. Exemplos de linguagens documentárias pré-coordenadas: CDD e CDU.

21 Pós-coordenação: controle e normalização realizados no estudo dos termos e unidades de indexação ou classificação, atendendo princípios semânticos ou genéricos. Os termos se coordenam em processo posterior à sua fixação, no estabelecimento de seu uso. Exemplos: tesouros e taxonomias.

uso de indicadores de relacionamento, que identificam as relações hierárquicas, associativas ou de equivalência.

A seleção dos termos é um dos fatores determinantes na construção de tesouros e se orienta por dois princípios complementares de validação:

- a) princípio da garantia literária: baseado em classes e termos da literatura, de modo que a literatura determina o universo de termos que serão utilizados para representar o domínio;
- b) princípio do endosso do usuário: baseado nos termos de busca utilizados pelos usuários, de modo a confirmar que os termos utilizados estão de acordo com aqueles que usam o sistema.

A união dos princípios da garantia literária e do endosso do usuário é um forte indicador de que a linguagem documentária será eficiente na tradução da linguagem natural. Contempla tanto o universo de documentos representativos do domínio, como o universo de possíveis perguntas que serão formuladas acerca da existência de documentos pelos usuários. (DODEBEI, 2002).

A determinação do universo conceitual também pode se basear no conhecimento prévio do domínio e na observação do campo conceitual. Antes de criar um tesouro, é importante verificar se já existe algum tesouro para o domínio ou campo de aplicação e, caso exista, verificar como e até onde pode ser utilizado e adaptado às necessidades de uso. Os tesouros existentes servem como guias e fontes de consulta para estabelecer as relações entre os termos e sua ordenação em novos instrumentos.

Algumas condições devem ser cumpridas por todos os tesouros, como ser uma linguagem especializada, estar normalizado e permitir a inclusão ou supressão de termos para se manter constantemente atualizado. A revisão e atualização contínuas são essenciais aos tesouros, que devem acompanhar o desenvolvimento dos domínios e as possibilidades de aplicação.

Os tesouros têm duas funções principais no âmbito dos sistemas de informação: indexação e recuperação da informação. São utiliza-

dos por profissionais da informação para descrever objetos informacionais com a especificidade desejada e para a recuperação da informação pelos usuários.

De modo geral, também cumprem a função de instrumentos de controle terminológico, usados para transferir os descritores da linguagem natural para o sistema linguístico de um domínio. Nessa perspectiva, podem ser amplamente utilizados em diversas áreas do conhecimento.

Taxonomias e tesouros compartilham utilidades comuns, por servirem como instrumentos de classificação da linguagem e sistematização do conhecimento. Ambos servem à indexação e recuperação da informação e contribuem para o desenvolvimento da Ciência e da Tecnologia. (CURRÁS, 2010).

Há, também, claras diferenças. A principal função da taxonomia é a estruturação da informação, estabelecendo a ordenação ou organização de informação, e não a padronização terminológica, característica dos tesouros. (CAMPOS; GOMES, 2008). Além disso, os tesouros não estão restritos aos relacionamentos hierárquicos e partitivos, mas compreendem outros relacionamentos que possibilitam interações mais complexas entre termos.

No contexto dos sistemas de informação, uma das principais diferenças é a utilização praticamente única de tecnologias informáticas nas taxonomias. Os tesouros podem ser estruturados manual e mecanicamente, apesar de existirem sistemas para a construção automática. Nesse sentido, as taxonomias são mais utilizadas por profissionais da computação, enquanto os tesouros são instrumentos dos profissionais da informação. (CURRÁS, 2010).

Os tesouros são amplamente utilizados nas áreas biomédicas e da saúde. Entre os instrumentos existentes, destacam-se o Medical Subject Headings (MeSH), o Descritores em Ciências da Saúde (DeCS), o Embase Subject Headings (Emtree) e o The Unified Medical Language System (UMLS).

O MeSH é o vocabulário controlado da National Library of Medicine (NLM), dos Estados Unidos, utilizado para a indexação e a recuperação da literatura biomédica, em especial na base de dados ME-

DLINE. Está disponível em diversos idiomas e acessível livremente na internet, tanto em página própria (<https://meshb-prev.nlm.nih.gov>), como integrado à base de dados PubMed. (<https://www.ncbi.nlm.nih.gov/mesh>). É dividido em 16 classes principais (Anatomia, Organismos, Doenças, Produtos Químicos e Medicamentos, entre outras), com os cabeçalhos denotando significados padronizados dispostos em hierarquias múltiplas. A ordem hierárquica baseia-se no princípio de que todos os documentos indexados por determinado cabeçalho são também relevantes para qualquer descritor matriz, compreendendo aspectos que vão do mais geral para o mais específico. (FREITAS; SCHULZ; MORAES, 2009).

Os cabeçalhos MeSH possuem notas de escopo (definições) e atributos adicionais, como sinônimos e termos associados. Os qualificadores admissíveis também compõem o registro do cabeçalho, a exemplo de *diagnóstico*, *tratamento farmacológico* e outros no caso de *Doença de Alzheimer*, além de datas de inclusão e revisão do termo no tesauro.

O DeCS é um instrumento de controle terminológico capaz de traduzir a linguagem natural dos documentos em uma linguagem normalizada, composta por descritores, sinônimos e qualificadores relacionados. Os conceitos que compõem o vocabulário são organizados em uma estrutura hierárquica que possibilita a pesquisa em termos mais amplos ou específicos de uma mesma estrutura hierárquica, assim como ocorre com o MeSH e o Emtree. Os cabeçalhos do DeCS possuem definições e atributos adicionais, como, sinônimos, termos relacionados, notas de indexação e qualificadores.

A atualização do DeCS é feita em consonância com a atualização do MeSH e envolve, além da inserção de novos descritores, a remoção e o reposicionamento de outros na estrutura do tesauro. O esforço de atualização está associado à recuperação da informação na BVS e às prioridades da cooperação técnica da Organização Pan-Americana da Saúde (OPAS) e Organização Mundial da Saúde (OMS). Desta forma, o DeCS também vem sendo atualizado para incorporar descritores em temas transversais de gênero, equidade, direitos humanos e etnicidade. (BIREME, 2017). A importância do DeCS se revela

por seu histórico de tradição e atualização, além do uso amplo na indexação e recuperação da informação em saúde. A relevância não se estabelece apenas para a América Latina e Caribe, mas também para a Comunidade dos Países de Língua Portuguesa (CPLP), a Espanha e os países francófonos. (BIREME, 2017).

O Emtree é o vocabulário controlado da base de dados Embase, utilizado para a indexação e a recuperação da literatura biomédica na base e em outras fontes. Está disponível em inglês, com acesso restrito na própria base (www.embase.com). A Embase está disponível no Portal de Periódicos da Capes (www.periodicos.capes.gov.br) desde 2020 e pode ser acessada pela comunidade acadêmica participante do Portal na opção Buscar Base.

Os cabeçalhos do Emtree são organizados em 14 classes (Conceitos Anatômicos, Funções Biológicas, Doenças, entre outras) e apresentados na ordem hierárquica, com termos que vão dos mais gerais aos mais específicos. O Emtree foi modelado a partir do MeSH, de modo que as funções e a estrutura dos vocabulários são semelhantes. Apesar disso, os vocabulários não são equivalentes e seu uso deve considerar tanto o objetivo como a fonte de indexação e recuperação da informação (MEDLINE ou Embase, por exemplo). Em alguns casos, os instrumentos podem ser considerados complementares, como é o caso da identificação de sinônimos e termos livres.

A disposição dos cabeçalhos em hierarquias múltiplas faz com que possam estar associados a mais de uma classe do Emtree. Os cabeçalhos são acompanhados de sinônimos, definições provenientes de diversas fontes e histórico do termo. Os qualificadores usados para delimitar conceitos de medicamentos, doenças e dispositivos médicos e refinar seu significado nos processos de indexação e recuperação da informação.

O metatesauro UMLS é uma terminologia única e uma rede semântica em saúde criada em 1986 e desenvolvida pela NLM dos Estados Unidos para apoiar o acesso à informação biomédica em várias fontes terminológicas. Integra e distribui os principais padrões de terminologia para promover a criação de sistemas e serviços de informação biomédica mais eficazes e interoperáveis. (FREITAS; SCHULZ;

MORAES, 2009; NLM, 2019). O acesso é livre para indivíduos na página www.nlm.nih.gov/research/umls.

O UMLS constitui um conjunto de arquivos e *softwares* que reúne vocabulários e padrões de saúde para permitir a interoperabilidade entre sistemas de computador. Nessa perspectiva, pode ser considerado um conjunto de tesouros, ontologias e terminologias. Contém mais de um milhão de conceitos biomédicos de mais de 100 vocabulários, incluindo vários mencionados neste texto. O DeCS participa do UMLS com a atualização e envio dos termos MeSH em português e espanhol.

Existem três fontes de conhecimento no UMLS: os conceitos do metatesauro; a rede semântica, com 133 categorias amplas e 54 relações do domínio; e o Especialista Lexicon & Lexical Tools, que fornece informações e programas lexicais para processamento de idiomas. O UMLS pode ser usado para processar textos e extrair conceitos e relacionamentos, desenvolver sistemas de recuperação da informação, extrair terminologias, facilitar o mapeamento entre terminologias, criar e terminologias locais, desenvolver serviços de terminologia, usar terminologias ou ontologias de pesquisa, vincular serviços médicos e farmacêuticos, coordenar o atendimento ao paciente entre setores e outras possibilidades. (NLM, 2019).

6 ONTOLOGIAS

Ontologias são instrumentos baseados na conceituação e artefatos compostos por vocabulários controlados e relações entre os conceitos, de modo a representar certa realidade. (GUARINO, 1995). O termo ontologia vem do grego *ontos*, que significa *ser*, e *logos*, que significa palavra ou estudo, recebendo o sentido de *estudo do ser* ou *estudo da existência do ser*.

As ontologias são expressas em formalismos baseados em lógica, que fornecem (meta) definições de classes (conceitos), relações, ocorrências e axiomas. Podem representar uma área de modo que os computadores possam processar as definições com suas semânticas, ao invés de empregar apenas termos de identificadores semânticos. (FREITAS; SCHULZ; MORAES, 2009).

Ontologias representam a realidade pela conceituação sistemática, compartilhada e formal num domínio de conhecimento. A conceituação pode ser explícita, formal ou compartilhada. Explícita por traduzir do implícito para o explícito, por descrever os elementos essenciais e particulares do objeto, visando eliminar as ambiguidades. Formal por ser sistematizada, estruturada em modelos que permitem, com a união do explícito e do formal, ser processada por máquinas e entendida por humanos. Compartilhada, por indicar a aceitação por uma comunidade. (SCHIESSL; SHINTAKU, 2012).

O fluxo de construção de uma ontologia indica uma sequência próxima a do desenvolvimento de *softwares*, iniciando com a especificação, que define a finalidade da ontologia, e finalizando com a manutenção, que não é um final, mas uma etapa que perdura enquanto a ontologia for válida.

A etapa de implementação se destina à aplicação da ontologia em sistemas de informação, ou seja, refere-se ao desenvolvimento da ontologia numa linguagem de representação do conhecimento que permita a sua utilização e a correção de possíveis problemas em sistemas informatizados.

A fim de apoiar os desenvolvedores nos processos de construção de ontologias, diferentes metodologias e métodos de construção foram elaborados, em especial nas décadas de 1990 e 2000. São exemplos de metodologias: Enterprise Ontology, Toronto Virtual Enterprise (TOVE), Methontology, método 101, método CYC, On-To-Knowledge Methodology (OTKM), NeOn, Up for Ontology, entre outras. (MENDONÇA; SOARES, 2017).

As principais funções das ontologias estão associadas à organização e à representação da informação em sistemas informatizados. As ontologias são elementos de significação e representação para a recuperação da informação na Web, em especial na Web Semântica, uma extensão da *World Wide Web*, que pressupõe questões de significação.

A Web Semântica interliga palavras com a finalidade de atribuir um significado aos conteúdos publicados na internet, de modo a torná-los compreensíveis por humanos e computadores. As ontologias

sustentam a ligação entre as palavras e atribuem significados aos conteúdos. Além disso, apoiam melhorias em aplicações de motores de busca, indexação automática, processos de gestão do conhecimento, entre outros.

As ontologias são semelhantes às linguagens documentárias, como os tesouros. Há um certo nexos entre ontologias e tesouros, uma semelhança entre ambos e uma certa dependência. Por outro lado, há diferenças significativas. A primeira delas está nos objetivos: os tesouros visam estabelecer a ligação entre a linguagem natural e artificial, enquanto as ontologias fornecem o mapa semântico de um domínio. Ontologias podem ir além da padronização do vocabulário, pois fornecem uma estrutura que viabiliza a identificação dos conceitos de forma automatizada em determinada realidade. Na comparação com as taxonomias, as ontologias são, em geral, mais complexas, pois vão além do relacionamento hierárquico e estabelecem relações semânticas.

Outra diferença está relacionada à finalidade dos termos. Enquanto nos tesouros os termos têm a função de restringir a linguagem para representar conceitos univocamente, nas ontologias a função é descrever e formalizar, no sentido de automatizar, os conceitos denotados pelos termos para um sistema. (SALES; CAFÉ, 2009). A questão de como as relações léxico-semânticas são expressas não é objeto das ontologias, pois estas se relacionam com entidades de forma independente da linguagem. (SCHULZ *et al.*, 2009).

A estrutura dos instrumentos é outra característica que os diferencia. Nos tesouros parte-se de uma ordenação dos termos, em hierarquias e relações semânticas, enquanto nas ontologias a ordenação é concebida de forma distinta, considerando as peculiaridades e propriedades dos termos. Por fim, a diferença também está no uso desses instrumentos. No caso dos tesouros, o alcance das definições dos termos se evidencia como um instrumento de registro terminológico, utilizado essencialmente por seres humanos. As ontologias, por sua vez, se destacam pela formalização em sistemas de informação, aspecto essencial para viabilizar inferências e processamento por máquinas. (SCHIESSL; SHINTAKU, 2012). As áreas biomédicas e

da saúde estão entre os campos mais comuns de aplicação das ontologias. Essas áreas vêm utilizando, de forma crescente, aplicações ontológicas para a representação de entidades biológicas e reestruturação das terminologias no aprimoramento dos registros de pesquisa e serviços de saúde. (CAMPOS *et al.*, 2009; SCHULZ *et al.*, 2009).

As principais motivações para o uso das ontologias nessas áreas são o armazenamento e a comunicação da informação e de dados de pacientes e a integração com aplicativos baseados em dados, os quais fornecem suporte à pesquisa básica e clínica e tornam mais eficientes os serviços de saúde. (GONÇALVES; ZAMBORLINI; GUIZZARDI, 2009).

Entre os esforços globais que fomentam as representações ontológicas biomédicas, destaca-se a *Open Biological and Biomedical Ontology* (OBO), um coletivo que reúne mais de 150 ontologias, como *The Foundational Model of Anatomy* (FMA), *Gene Ontology* (GO), entre outras (OBO Foundation, 2020).

7 TERMINOLOGIAS EM SAÚDE

As áreas da saúde dispõem de ampla gama de terminologias, artefatos linguísticos que unem os diversos sentidos de termos e conceitos. As terminologias geralmente são construídas com fins definidos, como a recuperação de documentos, o registro de estatísticas de doenças, mortalidade e morbidade, além de questões relacionadas às pesquisas e aos serviços de saúde. (FREITAS; SCHULZ; MORAES, 2009).

Terminologias são agrupamentos de termos, definições, anotações e relacionamentos que definem a linguagem especializada de um domínio. Diferem-se de tesouros e ontologias: enquanto as primeiras definem a linguagem utilizada, os tesouros e ontologias também representam o conhecimento, ou seja, os conceitos do domínio e suas relações.

O uso de terminologias universalmente aceitas garante o intercâmbio da informação, preserva seu significado e qualifica a comunicação. As terminologias se caracterizam pela representação extensa

do domínio (abrangência); pela granularidade (profundidade); pela composição do núcleo, baseado em termos, conceitos, relacionamentos e definições; e por outras características que garantem sua funcionalidade e utilidade. (LEÃO, 2012).

A OMS é responsável por diversas terminologias aceitas universalmente. A Família de Classificações Internacionais da OMS é formada por três classificações de referência e por um conjunto de classificações derivadas e relacionadas. As classificações de referência da OMS (WHO, 2018) são descritas como:

- a) **CID** – Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde: é a base para as estatísticas de saúde. Mapeia a condição humana desde o nascimento até a morte, codificando as lesões, doenças ou causas de morte, além dos fatores que influenciam a saúde ou causas externas de mortalidade e morbidade. Fornece uma visão holística de todos os aspectos da vida que podem afetar a saúde. A edição atual é a 11^a (CID-11), de 2015;
- b) **CIF** – Classificação Internacional de Funcionalidade, Incapacidade e Saúde: utilizada para medir a saúde e a deficiência dos indivíduos e da população. É a base para a compreensão dos determinantes, resultados e condições de saúde, classificando as funções e a incapacidade humana no meio ambiente e os fatores contextuais;
- c) **CIIS** – Classificação Internacional de Intervenções de Saúde: ferramenta para relatar e analisar intervenções de saúde. Abrange diagnóstico, medicina, cirurgia, saúde mental, atenção primária, apoio funcional, reabilitação, medicina tradicional e saúde pública.

REFERÊNCIAS

BRASCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo, 2008.

- CAMPOS, A. L. A. *et al.* Aspectos Metodológicos no Reuso de Ontologias: um estudo a partir das anotações genômicas no domínio dos tripanosomatídeos. **Reciis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 3, n. 1, p. 64-75, 2009.
- CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e Classificação: o princípio de categorização. **DataGramZero** – Revista de Ciência da Informação, v. 9, n. 4, p. 1-22, 2008.
- CASTELLS, M. **A Sociedade em Rede**. São Paulo: Paz e Terra, 2009.
- CURRÁS, E. **Ontologias, Taxonomia e Tesouros em Teoria de Sistemas e Sistemática**. Brasília: Thesaurus, 2010.
- CURRÁS, E. **Tesouros: linguagens terminológicas**. Brasília: IBICT, 1995.
- DAHLBERG, I. Knowledge organization: A new science? **Knowledge Organization**, Wurzburg, v. 33, n. 1, p. 11-19, 2006.
- DAHLBERG, I. Teoria do Conceito. **Ciência da Informação**, v. 7, n. 2, 1978.
- DODEBEI, V. L. D. **Tesouro: linguagem de representação da memória documentária**. Rio de Janeiro: Interciência, 2002.
- FREITAS, F.; SCHULZ, S.; MORAES, E. Pesquisa de Terminologias e Ontologias atuais em Biologia e Medicina. **Reciis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 3, n. 1, p. 8-20, 2009.
- GOMES, A. P. et al. O Papel dos Mapas Conceituais na Educação Médica. **Revista Brasileira de Educação Médica**, v. 35, n. 2, p. 275-282, 2011.
- GOMES, H. E. **Manual de Elaboração de Tesouros Monolíngues**. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990.
- GONÇALVES, B.; ZAMBORLINI, V.; GUIZZARDI, G. Uma Análise Ontológica do Eletrocardiograma. **Reciis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 3, n. 1, p. 49-63, 2009.
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human-Computer Studies**, London, v. 43, n. 5-6, p. 625-640, 1995.
- HJØRLAND, B. What is knowledge organization? **Knowledge Organization**, Wurzburg, v. 35, n. 2-3, p. 86-101, 2008.
- LEÃO, B. F. **Terminologias em Saúde: conceitos básicos e exemplos**. 2012. Disponível em: www.sgc.goias.gov.br/upload/links/arq_589_SISAGoiasABeatrizA2012AMarA07A-APadrosesATerminologia.pdf. Acesso em: 22 mar. 2020.
- LIMA, G. A. B. O. Mapa Conceitual como Ferramenta para Organização do Conhecimento em Sistemas de Hipertextos e Aspectos Cognitivos. **Perspectivas em Ciência da Informação**, v. 9, n. 2, p. 134-145, 2004.

- LIMA, J. L. O.; ALVARES, L. Organização da Informação e do Conhecimento. *In: ALVARES, L. et al. (Org.). **Organização da Informação e do Conhecimento**: conceitos, subsídios interdisciplinares e aplicações.* São Paulo: B4 Editores, 2012. Cap. 1, p. 21-48.
- NLM - U. S. NATIONAL LIBRARY OF MEDICINE. **Medical Subject Headings**. 2020. Disponível em: <https://meshb.nlm.nih.gov>. Acesso em: 24 mar. 2020.
- NLM - U. S. NATIONAL LIBRARY OF MEDICINE. **The Unified Medical Language System (UMLS)**. 2016. Disponível em: www.nlm.nih.gov/research/umls/new_users/online_learning/OVR_001.html. Acesso em: 18 mar. 2020.
- NLM - U. S. NATIONAL LIBRARY OF MEDICINE. **Unified Medical Language System (UMLS)**. 2019. Disponível em: <https://www.nlm.nih.gov/research/umls/index.html>. Acesso em: 18 mar. 2020.
- OBO Foundation. **The Open Biological and Biomedical Ontology**. 2020. Disponível em: www.obofoundry.org. Acesso em: 22 mar. 2020.
- SALES, R.; CAFÉ, L. Diferenças entre Tesouros e Ontologias. **Perspectivas em Ciência da Informação**, v. 14, n. 1, p. 99-116, 2009.
- SCHIESSL, M.; BRÄSCHER, M. Do Texto às Ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, v. 40 n. 2, p.301-311, 2011.
- SCHIESSL, M.; SHINTAKU, M. Sistemas de Organização do Conhecimento. *In: ALVARES, L. et al. (Org.). **Organização da Informação e do Conhecimento**: conceitos, subsídios interdisciplinares e aplicações.* São Paulo: B4 Editores, 2012. Cap. 2, p. 49-118.
- SCHULZ, S. Vantagens e Limitações das Ontologias Formais na Área Biomédica. **Reciis: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 3, n. 1, p. 33-48, 2009.
- TAVARES, R. Construindo Mapas Conceituais. **Ciências & Cognição**, v. 12, p. 72-85, 2007.