

Interoperabilidade e mapeamentos entre sistemas de organização do conhecimento: Bioportal of National Center for Biomedical Ontology-NCBO

Doutora Juliatti de Andrade
Departamento de Ciência da Informação - GCI
Universidade Federal Fluminense - UFF
Email: juliattiandrade@id.uff.br

Doutora Marilda Lopes Ginez de Lara
Departamento de Informação e Cultura - ECA
Universidade de São Paulo - USP
Bolsista de Produtividade em Pesquisa – CNPq
Email: larama@usp.br

Resumo

Introdução e objetivos: São apresentados os resultados de pesquisa de doutorado realizada entre 2011-2015 referente à análise dos mapeamentos entre termos do BioPortal, portal na web desenvolvido pelo National Center for Biomedical Ontology-NCBO que dá acesso a aproximadamente 690 ontologias da área Biomédica. Discorre sobre o método utilizado para as análises relacionadas à consistência semântica dos resultados de busca e dos mapeamentos entre os termos das ontologias que compõem o BioPortal. **Método:** recorre à pesquisa bibliográfica, nos últimos 10 anos, em bases de dados das áreas de Ciência da Informação e Biblioteconomia, sobre os temas organização e recuperação da Informação, interoperabilidade e mapeamentos entre Sistemas de Organização do Conhecimento (SOC), e realiza avaliação das formas de busca e recuperação e dos mapeamentos entre os termos das ontologias do BioPortal com base nas recomendações da norma ISO 25964:2:2011, Parte 2: Interoperability with other vocabularies. **Resultados e conclusões:** considera relevante o uso de SOC interoperados e mapeados para diversas aplicações de organização e recuperação de informação, especialmente para seleção de termos para construção de estratégias de busca sensibilizadas, mas alerta para o fato de que a recuperação por coincidência de caracteres compromete a precisão porque recupera termos que não correspondem semanticamente ao que foi

procurado, impondo, conseqüentemente, a necessidade de avaliação e seleção de termos que sejam pertinentes para a busca.

Palavras-Chave: Mapeamento entre Sistemas de Organização do Conhecimento, Interoperabilidade, Estratégias de Busca Sensibilizadas.

Interoperability and mapping of knowledge organization systems: Bioportal of the National Center for Biomedical Ontology-NCBO

Abstract

Introduction and Objective: The study presents results of a PhD research conducted between 2011-2015 regarding the analysis of the mappings between terms of BioPortal, a web portal that gives access to approximately 690 ontologies of the Biomedical area, developed by the National Center for Biomedical Ontology-NCBO. It discusses the method used to analyze the semantic consistency of the search results and the mappings between the terms of the ontologies that compose the BioPortal library. **Method:** bibliographical research in the last 10 years in databases of Information Science and Library Studies on Information Organization and Retrieval, interoperability and mapping between Knowledge Organization Systems and evaluation of search and retrieval methods and mapping between the terms of the BioPortal ontologies based on the recommendations of ISO 25964: 2: 2011 Part 2: Interoperability with other vocabularies. **Results and conclusion:** the use of interoperated and mapped SOC for various information organization and retrieval applications, especially for the selection of terms for the construction of sensitive search strategies for information retrieval in the Health area is considered relevant. The recovery by coincidence of characters compromises the accuracy by means of the recovery of terms that do not correspond semantically to what was sought which consequently

imposes the necessity of evaluation and selection of terms.

Keywords: *Mapping between Knowledge Organization Systems, Interoperability, Sensitive Search Strategies.*

INTRODUÇÃO

A Saúde Baseada em Evidências (SBE) é uma área da saúde cujo objetivo é oferecer a melhor informação disponível para a tomada de decisão na assistência e na gestão hospitalar por meio da Epidemiologia Clínica, da Estatística, da Metodologia Científica e da Informática¹.

A produção e o uso do conhecimento técnico-científico demandam a busca qualificada de informações, o que requer a sistematização e a explicitação do caminho percorrido para sua realização. A sistematização é realizada por meio de estruturação de problemas de pesquisa com o uso do modelo PICO (problema, intervenção, controle e desfecho), o mapeamento e a seleção de termos e conceitos a partir desses problemas, a construção de estratégias de busca de acordo com o funcionamento dos Sistemas de Recuperação de Informações (SRI) e a organização dos resultados de busca de modo a que possam integrar as metodologias dos trabalhos, como as revisões sistemáticas. As revisões sistemáticas constituem os meios que permitem conferir maior credibilidade à SBE por explicitarem a metodologia adotada e tornarem o estudo reproduzível². A sistematização e a explicitação das ações de busca e recuperação de informações integram as revisões sistemáticas corroborando a reprodutibilidade do método.

Os Sistemas de Organização do Conhecimento (SOC) ou Knowledge Organization Systems (KOS), na terminologia da International Knowledge Organization Society (ISKO), incluem as Linguagens Documentárias que, por sua vez, incluem os tesouros, os sistemas de classificação, as listas de cabeçalhos de assuntos, entre outros. Os SOC têm papéis fundamentais para a busca qualificada de informações ao permitir o mapeamento de termos e conceitos para construção de estratégias de busca e, por conseguinte, para a produção do conhecimento técnico-científico, pois é a partir de conhecimentos já existentes que novos conhecimentos são produzidos, avaliados, atualizados ou descartados.

Os problemas que se apresentam nesse escopo são: como lidar com os fenômenos da linguagem, como a sinonímia, a semântica, a sintaxe e os aspectos morfológicos na construção e aplicação de estratégias de busca em bases de dados? Como acessar e utilizar diversos SOC de determinada área simultaneamente? Como obter uma amostra significativa de termos equivalentes que possibilite a recuperação de informações com revocação e precisão levando-se em consideração as características do problema do estudado, dos SOC e as formas de recuperação dos SRI?

Para lidar com essas questões, considera-se relevante o uso de SOC mapeados e interoperados na busca e recuperação de informações, pois eles permitem a busca simultânea em vários instrumentos, assim como

a seleção de termos equivalentes para serem usados na construção e aplicação de estratégias de busca. Com essa premissa, objetiva-se, neste trabalho, apresentar buscas realizadas no BioPortal³ acompanhadas do método adotado, de modo a explicitar os passos para a realização das análises da recuperação da informação do ponto de vista semântico, no que se refere aos resultados de busca por classes* e mapeamentos entre classes de ontologias. Espera-se que, ao mostrar o caminho percorrido, possa-se generalizar o método.

MÉTODOS

Como apoio para a fundamentação deste trabalho, recorreu-se à realização de pesquisa bibliográfica sobre Organização e Recuperação da Informação, Interoperabilidade e Mapeamentos entre Sistemas de Organização do Conhecimento e Saúde Baseada em Evidências nas seguintes bases de dados: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação – BRAPCI; LISA: Library and Information, Information Science & Technology Abstracts (ISTA); Library, Information Science & Technology Abstracts with Full Text; PubMed, Scielo; Biblioteca Digital de Teses e Dissertações da USP; Site do Centro Cochrane do Brasil, periódico El Profesional de la Informacion. As buscas foram realizadas nos idiomas português, inglês e espanhol no período de 2000-2015, a partir dos seguintes descritores: Organização da Informação, Organização do

Conhecimento, Interoperabilidade, Mapeamento entre SOC, Sistema de Recuperação de Informações, Cruzamento de linguagem na recuperação de informação, Estratégias de busca, Saúde Baseada em Evidências, Revisão Sistemática, Biblioteconomia Baseada em Evidências, Uso da Informação Científica na Tomada de Decisões em Saúde.

Para o caso específico da pesquisa em saúde, elaboramos e aplicamos um método de avaliação dos resultados de busca que contou com a identificação dos objetivos do BioPortal, das formas de busca e recuperação de termos nas ontologias, e a avaliação da consistência semântica da recuperação por meio de comparação entre termos utilizados nas buscas, os recuperados, e os obtidos nos mapeamentos entre classes das ontologias do referido portal.

As análises foram feitas sob três perspectivas: a) análise dos tipos de busca e recuperação no que se refere à relação entre os termos utilizados nas buscas e os termos recuperados, b) análise da organização dos resultados de busca, c) análise dos mapeamentos entre classes.

Realizou-se a identificação dos tipos de busca e recuperação de informação, bem como registro, sistematização e reorganização dos resultados de busca para análise e classificação das equivalências. A classificação das equivalências nos níveis exata, inexata e parcial foi feita com base na norma ISO 25964:2:2011, assim como na observação de aspectos semânticos que dizem respeito à relação entre a forma

*O BioPortal utiliza a designação “classe” para referir-se aos termos que compõem as ontologias

linguística dos termos e suas definições; aspectos morfológicos que se referem à variação de número e ao uso de preposições, e aspectos sintáticos que se relacionam à variação da posição das palavras na constituição dos termos, assim como ao uso de pontuação.

As análises contaram com as seguintes etapas:

- Identificação dos objetivos do Portal e tipos de busca;
- Estabelecimento de critérios para busca e recuperação: critérios de inclusão e exclusão de tipos de busca e interfaces dos resultados, assim como definição dos termos utilizados para busca.
- Estabelecimento de modelo de registro e sistematização dos resultados de busca;
- Aplicação das estratégias de busca acompanhada de registro e sistematização dos resultados;
- Análise dos resultados de busca (equivalências e termos relacionados) com base no uso da norma ISO 25964-2:2011 e em aspectos semânticos, morfológicos e sintáticos para identificação e seleção de termos equivalentes resultantes do mapeamento entre SOC dos produtos analisados;
- Inclusão e exclusão de termos equivalentes;
- Considerações sobre os tipos de busca e recuperação e organização dos resultados de busca.

Elegemos o termo Knee Arthroplasty* e os seus respectivos termos equivalentes no tesouro

Emtree da base de dados Embase Biomedical Database Elsevier (2013), pelo fato de não fazer parte do BioPortal. Optamos por trabalhar com termos de um SOC que não fizesse parte do produto analisado para verificarmos o quanto o mapeamento entre SOC consegue integrar possíveis equivalências.

Interoperabilidade entre sistemas de organização do conhecimento

A possibilidade de acesso a termos e conceitos provenientes de vários SOC demanda a adoção de padrões. Estes dizem respeito à concepção dos SRI utilizados, aos princípios e conceitos de cada SOC, às formas de interoperabilidade, além dos padrões da Web Semântica e dos mapeamentos entre SOC.

A interoperabilidade é definida como a habilidade de dois ou mais sistemas usarem as informações trocadas sem esforço especial por parte de qualquer um dos sistemas^{4,5,6,7}. Para a norma ISO 25964-2:2011, o principal propósito da interoperabilidade entre SOC é tornar possível que uma expressão formulada usando um determinado SOC possa ser convertida em outra correspondente em outro SOC⁸.

Para estabelecer essa correspondência é necessária a realização de mapeamentos, os

*Termos e termos equivalentes do Emtree utilizados nas buscas: *knee arthroplasty* – termo preferido; equivalências: *arthroplasty*, *replacement*, *knee*; *arthroplasty, knee*; *knee arthroplasties*; *knee joint replacement*; *knee joint replacements*; *knee reconstruction*; *knee replacement*; *knee replacements*; *reconstruction, knee*

quais consistem no estabelecimento de relacionamentos semânticos entre termos de vários SOC de uma dada área do conhecimento.

Martinez Tamayo et al.⁹ analisam os objetivos da interoperabilidade entre SOC: “(...) a interoperabilidade entre os Sistemas de Organização do Conhecimento (SOC) procura harmonizar as relações conceituais e terminológicas que podem ser estabelecidas entre eles. É fazer com que os SOC possam trocar informações, independentemente do contexto em que foram criados, mantendo a eficiência, juntamente com outros SOC, na recuperação de informação. Esta troca de dados entre SOC pode ser estabelecida incluindo relações com os outros SOC, representando dados em padrões e usando sistemas que implementam protocolos comuns” (tradução nossa)⁹.

A possibilidade de uso simultâneo de SOC nos processos de busca e recuperação de informações demanda a interoperabilidade e o mapeamento semântico entre eles, os quais devem estar formalizados no que se refere à estruturação, abertura e ligação de dados no contexto da Web Semântica para, além de tornar possível a disponibilização de dados, estabelecer como fazer ligações entre eles de modo que uma pessoa ou máquina possa explorá-los.

Na Web Semântica proposta por Berners-Lee, Linked Data ou Linked Open Data corresponde a um conjunto de boas práticas para ligar e publicar dados estruturados na Web¹⁰. A

iniciativa Linked Data “designa um conjunto de procedimentos que visam promover a abertura dos dados de modo a permitir sua preparação, disponibilização e reaproveitamento” (tradução nossa)¹¹.

Para que os SOC possam ser utilizados em aplicações na Web, eles devem estar abertos e ligados no contexto da Web Semântica. A iniciativa referente à construção, ao uso e ao reuso desses sistemas na Web é denominada Linked Open Vocabularies (LOV), a qual se constitui em recomendações do W3Consortium visando estimular e fundamentar a publicação e a abertura dos dados relativos a vocabulários controlados¹¹.

Dentre as recomendações do W3Consortium (desde 2009) para publicação e abertura de dados dos SOC destaca-se o emprego do Simple Knowledge Organization System (SKOS). Para Martinez Tamayo et al.⁹, Lara¹¹, Pastor-Sánchez; Martínez-Méndez; Rodríguez-Munõz¹², o objetivo deste modelo é oferecer um padrão para migrar e construir os SOC no ambiente da Web Semântica. Segundo Martinez Tamayo et al.⁹, outros padrões concorrem para a interoperabilidade: o uso do padrão Resource Description Framework (RDF) e a Web Ontology Language (OWL) que, junto com o SKOS, permitem a interoperabilidade entre os SOC dentro de um ambiente Web. Tais padrões a) utilizam Extensible Markup Language (XML) como linguagem de marcação para estruturar dados; b) permitem que cada conceito seja identificado unívoca e universalmente mediante um Uniform Resource Identifier

(URI); c) e que cada recurso seja tratado como uma tripla que relaciona sujeito, predicado e valor.

No que tange ao vocabulário, a função do SKOS é contribuir na definição de relações entre os seus termos. “SKOS fornece um número de relações semânticas para estabelecer vínculos de mapeamento entre conceitos em diferentes esquemas. Isso indica se um conceito de um esquema é considerado idêntico a outro ou se você tem um significado próximo, genérico, específico ou relacionado. A nova norma ISO 25964 de tesouros propõe uma função semelhante para definir relações entre diferentes linguagens documentárias para ser utilizada em conjunto em operações de recuperação de informação” (tradução nossa)¹².

RDF é um padrão na construção da Web Semântica que usa modelos de metadados descritivos e orientados ao conteúdo, assim como supõe uma estrutura que permite a construção de linguagens lógicas que podem trabalhar juntas na Web Semântica ou codificar vocabulários existentes, como o Dublin Core^{9,12}. Para Méndez Rodríguez¹³, o RDF é uma maneira de usar XML orientado aos dados e aos metadados, e não apenas aos documentos. Caro Castro¹⁴ mostra que a representação de dados em RDF é utilizada com o SPARQL* como linguagem de consulta de dados, assim como observa que “embora o HyperText Markup Language (HTML) proporcione um meio para estruturar e ligar documentos na Web, o RDF proporciona um modelo de dados genérico que permite estruturar e ligar dados

que descrevem coisas no mundo em forma de triplas “sujeito-predicado-objeto” (tradução nossa)¹⁴.

Martinez Tamayo et al.⁹ descrevem a OWL como uma linguagem de marcação com base em RDF para representar ontologias, que são modelos que descrevem e representam significados relativos a uma porção do universo em um ambiente Web, por meio de codificação do conhecimento de um domínio, da descrição de classes, propriedades, relações e indivíduos.

O funcionamento da interoperabilidade via SKOS é viabilizado porque os elementos de SOC que correspondem a conceitos e a relações hierárquicas e associativas são formalmente representados¹². Considera-se que as iniciativas Linked Open Data e Linked Open Vocabularies associadas aos softwares referidos anteriormente, contribuem para construção, o uso e reuso de SOC por várias instituições, em diversas aplicações referentes à organização e à recuperação da informação na Web.

Mapeamento segundo a Norma ISO 25964-2:2011 INFORMATION AND DOCUMENTATION: thesauri and interoperability with other vocabularies

A “busca pela interoperabilidade é o que motiva a revisão das normas documentárias para elaboração de vocabulários controlados e tesouros”¹¹. A norma ISO 25964:2011 - Information and documentation - Thesauri and interoperability with other vocabularies é

* *Protocol and RDF Query Language*

organizada em duas partes, a saber: Parte 1: Thesaurus for information retrieval; Parte 2: Interoperability with other vocabularies. Elas constituem o resultado da revisão das normas Z39.19-2005, da BS 8723-4/5:2007/2008, das Guidelines for multilingual thesauri da International Federation of Library Association and Institutions 2009 e das propostas do SKOS pela W3C feitas em 2004 e 2005¹¹.

A segunda parte da norma estabelece critérios e orienta a elaboração e implementação de projetos de interoperabilidade e mapeamentos entre SOC. Analisamos, neste trabalho, as definições, os objetivos e os níveis de mapeamentos para operacionalizar concretamente suas recomendações.

Para estabelecer a interoperabilidade entre SOC é necessário realizar mapeamentos, que consistem no estabelecimento de relacionamentos semânticos entre termos de vários SOC de uma área do conhecimento. Os principais tipos de mapeamentos são os de equivalência, o hierárquico e o associativo, sendo “a equivalência o tipo mais comum e necessário”⁸.

As equivalências devem ser estabelecidas quando os conceitos correspondentes são encontrados em dois ou mais SOC diferentes, o que difere da equivalência entre dois termos num tesouro monolíngue quando um deles é designado como termo preferido e outro como termo não preferido. O mapeamento de equivalências entre SOC ocorre entre conceitos e não há diferença de status entre os conceitos,

termos preferidos ou notações que os representem⁸.

Os mapeamentos de equivalência são organizados em tipos e níveis. Os tipos compreendem os mapeamentos de equivalências simples e compostas. Para tratar a equivalência composta, as recomendações falam da intersecção de equivalências compostas, equivalência composta cumulativa e equivalência composta envolvendo vocabulários alvos (target vocabularies)⁸. Em relação aos níveis, a norma estabelece as equivalências exatas, inexatas e parciais. A equivalência exata acontece “quando os conceitos podem ser usados de forma intercambiável através de todas as aplicações que podem ser previstas para o mapeamento”; a inexata, quando conceitos correspondentes em dois ou mais vocabulários não são exatamente os mesmos, e a parcial relaciona-se às características genéricas ou específicas dos significados dos termos⁸.

A norma também aborda os mapeamentos hierárquicos e associativos. Os mapeamentos hierárquicos podem ser estabelecidos entre conceitos quando um é claramente mais genérico do que o outro. A relação parte-todo também pode justificar o mapeamento hierárquico. A norma não define mapeamento associativo, apenas observa que o mesmo pode ser estabelecido entre conceitos quando eles não qualificam equivalência ou mapeamento hierárquico, mas são semanticamente associados. Esclarece, ainda, que o limite entre mapeamentos associativos e equivalências

inexatas é indefinido e subjetivo, mas pode ser feito pragmaticamente de acordo com o contexto no qual o mapeamento será usado, citando como exemplo para o estabelecimento dessas distinções, o interesse dos usuários, o volume de recursos a ser pesquisados, a capacidade de recuperação e a visualização do sistema⁸.

Nesse trabalho, utilizamos os níveis de equivalência propostos pela norma para análise dos mapeamentos entre classes e termos recuperados no BioPortal como veremos a seguir.

Bioportal

Durante a realização de pesquisa de doutorado¹⁵, analisamos o BioPortal que fornece acesso a uma biblioteca que atualmente contém aproximadamente 690 ontologias biomédicas, incluindo tesouros, no campo da Biomedicina e a Biologia (na época da realização do experimento, o portal contava com 302 ontologias). Desenvolvido em OWL, Biological and Biomedical Ontologies (OBO), em RDF e Protégé Frames, o BioPortal permite navegar, buscar e descarregar ontologias. Os usuários registrados podem adicionar mapeamentos entre os termos e comentários em termos individuais dentro da ontologia¹⁶.

Tipos de busca e recuperação

No BioPortal é possível buscar classes dentro das ontologias (search for a class) nas modalidades simples e avançada, e por ontologia (find an ontology), por meio da

digitação de um assunto que deve fazer parte do título da ontologia, devido à recuperação por coincidência de caracteres. Também é possível acessar uma listagem com todas as Ontologias (Browse Ontologies) que pode ser ordenada sob os seguintes critérios: popular, tamanho, projetos, notas, data de upload.

Busca por uma classe: apresenta campo de busca livre por classe, cuja interface de resultado de busca apresenta o seguinte padrão:

- Número total de ontologias recuperadas;
- Lista de enunciados linkáveis formados pelos nomes das classes recuperadas seguidas do nome da ontologia à qual a classe pertence. Esse link dá acesso a uma ficha da classe com as categorias details, visualization, notes e class mapping, dentre as quais destacamos:

Detalhes (*details*): dessa categoria podem ser destacados os metadados: nome preferido (preferred name) e prefLabel, que corresponde ao que denominamos Termo Preferido no escopo de construção dos tesouros; link da classe no BioPortal; classificações (classifies), que se referem a termos sinônimos; CUI, identificador único do conceito extraído do Metatesouro do Unified Language System da National Library Medicine; Inverse of SIB e members of, que se referem a termos relacionados; SMQ_TERM_LEVEL, que se refere ao nível do termo na hierarquia e à subcategoria (subClassOf).

Visualização (*visualization*): representação gráfica do termo nas modalidades caminho para

raiz (path to root), termo vizinho (term neighborhood) e mapeamento para o termo vizinho (mappings neighborhood).

Mapeamento de classes (*class mappings*): corresponde às classes e suas respectivas ontologias em vários idiomas consideradas equivalentes ao termo selecionado.

Busca avançada em todas Ontologias: permite buscar uma classe do conjunto de ontologias do portal com a inclusão de informações sobre valores de propriedade, classes obsoletas, visualizar ontologias, assim como buscar classe por Correspondência Exata e Classes com Definições. Nessa modalidade, também é possível buscar por uma das 41 categorias correspondentes a diversos domínios da área da saúde e por ontologias os quais já vem default no sistema.

Estabelecimento de critérios de busca

Para identificar e analisar os mapeamentos entre classes de SOC distintos realizados pelo BioPortal, utilizamos os termos Knee Arthroplasty e seus respectivos termos equivalentes do tesouro Emtree do Embase, com os seguintes critérios de inclusão e exclusão:

Crítérios de Inclusão/Tipo de busca: busca de termos em todas as ontologias no modo avançado; Filtros utilizados: Classes com definições;

Crítérios de Exclusão/Filtros: Correspondência exata, pois dentro de nossa amostra de termos utilizados nas buscas, somente o termo Knee Arthroplasty retornou

resultados nessa modalidade, o que impossibilitou a avaliação de resultados da perspectiva de equivalências e termos relacionados. Também foram excluídos os filtros propriedades de valores, classes obsoletas e ver ontologia.

RESULTADOS

Os termos recuperados foram analisados em relação aos termos utilizados nas buscas por meio da leitura da ficha do termo, onde observou-se: a designação do termo, hierarquias, definições e sinônimos, assim como o domínio do conhecimento coberto pela ontologia. Eles foram classificados nos níveis de equivalências exata, parcial e inexata com base na norma ISO 25964:2:2011 e em termos relacionados, os quais não foram classificados porque foram excluídos para os nossos objetivos.

A análise da organização dos resultados de busca permitiu verificar o emprego de interoperabilidade e mapeamentos entre as ontologias do BioPortal ao observamos o estabelecimento de equivalências entre o termo utilizado na busca e os termos recuperados, o emprego do conceito de relevância, assim como o de recuperação por coincidência de caracteres.

Na ordenação dos resultados de busca:

1. Recuperam-se todos os termos com as mesmas palavras na mesma sequência. Ex: ao buscar com o termo Knee Arthroplasty recupera-se o termo Knee Arthroplasty, o que compreendemos como uma equivalência exata;

2. Recuperam-se todos os termos com as mesmas palavras com sequência diferente da do termo utilizado na busca. Ex: ao buscar com o termo Knee Arthroplasty recupera-se o termo Arthroplasty Knee, o que corresponde ao estabelecimento de equivalência exata do ponto de vista semântico, relacionando variações sintáticas de um mesmo termo;

3. Recuperam-se todos os termos que contenham o termo utilizado na busca como equivalentes, com a mesma sequência de palavras. Ex: ao buscar com o termo Knee Arthroplasty recupera-se Total Knee Arthroplasty, que pode ser considerado como equivalência parcial em relação à Knee Arthroplasty;

4. Recuperam-se todos os termos que contenham ao menos uma palavra do termo utilizado na busca, em ordem alfabética. Ex: Arthroplasty;

5. Recuperam-se termos com mais de duas palavras que contenham os termos utilizados na busca com a mesma sequência de palavras, em ordem alfabética. Ex: Generic Planning for Knee Replacement.

Ao todo, foram feitas 10 buscas que resultaram um total de 54 termos, os quais totalizaram 195 pelo fato deles terem sido recuperados em mais de uma busca.

Quadro 1: Resultados de busca BioPortal com o termo *Knee Arthroplasty* e equivalências do tesauro Emtree do Embase.

Termos utilizados nas buscas	Total de classes (termos) recuperadas em cada busca
knee Arthroplasty	17
arthroplasty, replacement, knee	23
arthroplasty, knee	17
knee arthroplasties	15
knee joint replacement	24
knee joint replacements	17
knee reconstruction	24
knee replacement	23
knee replacements	16
reconstruction, knee	19
Total de termos recuperados	195

Fonte: Dados da pesquisa, 2018.

Do conjunto dos 54 termos, 6 foram classificados como equivalentes, 12 como termos relacionados e 36 não foram considerados equivalentes nem relacionados aos termos utilizados nas buscas. No conjunto das equivalências observa-se que apenas 2 termos se referem especificamente à Artroplastia do Joelho, pois os termos Arthroplasty e o Joint Reconstruction, ambos recuperados em 2 ontologias, foram classificados como uma equivalência parcial nessa perspectiva, por se tratar de um deslocamento genérico em relação à Artroplastia do Joelho.

Os termos relacionados foram assim tratados por não representarem a Artroplastia de Joelho de forma específica ou genérica, mas sim relacionados a ela por apresentarem relações semânticas, como por exemplo, o local da cirurgia (joelho), a prótese, lesões e patologias que geram a Artroplastia, entre outros.

Os 36 termos que não foram considerados equivalentes ou relacionados foram assim classificados porque não apresentaram características que os relacionassem semanticamente à Artroplastia de Joelho.

Observa-se que eles foram recuperados porque os termos utilizados na busca continham ao menos uma palavra em sua designação que coincidia com alguma palavra na designação do termo recuperado, assim como nos sinônimos. Essa situação foi constatada em todos os resultados de busca, o que nos permite concluir que esse fenômeno é resultante de recuperação por coincidência de caracteres.

Os quadros 2 e 3 são exemplos de termos classificados nos níveis de equivalência e em termos relacionados segundo a norma ISO 25964.

Quadro 2: Exemplo de Resultados de busca no BioPortal classificados por níveis de equivalência.

Termos utilizados nas buscas	Termos recuperados classificados por Níveis de Equivalência	Nome da Ontologia
arthroplasty, knee Knee Arthroplasty knee replacement knee joint replacement	Arthroplasty, Replacement, Knee Equivalência exata	Medical Subject Headings (MESH)
knee replacement arthroplasty, replacement, knee	Total Knee Replacement Equivalência Parcial	National Cancer Institute Thesaurus (NCIT)
knee replacement arthroplasty, replacement	joint reconstruction Equivalência Parcial	Radiology Lexicon (RADLEX)

Fonte: Dados da pesquisa, 2018.

Quadro 3: Exemplos de Resultados de busca no BioPortal classificados como Termos Relacionados.

Termos utilizados nas buscas	Termos recuperados classificados como relacionados	Ontologia
knee replacement knee joint replacement arthroplasty, replacement, knee knee arthroplasties Knee Arthroplasty arthroplasty, knee knee replacements knee joint replacements	Generic Planning for Knee Replacement	DICOM Controlled Terminology (DCM)
knee replacement Knee Arthroplasty arthroplasty, knee arthroplasty, replacement, knee knee joint replacement	Hip replacement	Human Physiology Simulation Ontology (HUPSON)
knee joint replacements knee joint replacement arthroplasty, knee	Knee Joint	National Cancer Institute Thesaurus (NCIT)
knee replacements knee joint replacements	Knee Injuries	Medical Subject Headings (MESH)
knee arthroplasties	Osteoarthritis, Knee	Medical Subject Headings (MESH)

Fonte: Dados da pesquisa, 2018.

Na época em que as análises foram realizadas, verificamos que a variação sintática referente à posição das palavras na constituição dos termos altera os resultados, como observamos em relação aos termos Knee Arthroplasty e Arthroplasty Knee. As buscas com ambos resultaram em 17 termos, porém não os mesmos. Com Knee Arthroplasty, foram recuperados os termos Knee Skin, do National Cancer Institute Thesaurus (NCIT) e o Knee region structure, da Systematized Nomenclature of Medicine Clinical Terms (SNOMEDCT), termos esses que não foram recuperados na busca com o termo Arthroplasty Knee. Este

último, por sua vez, recuperou os termos Knee Joint, do National Cancer Institute Thesaurus (NCIT) e o Knee effusion stroke test, também da Systematized Nomenclature of Medicine Clinical Terms (SNOMEDCT) que não foram recuperados na busca com Knee Arthroplasty. Entretanto, ao refazer o caminho em fevereiro de 2018, para atualizar a pesquisa em função deste artigo, observou-se o estabelecimento de equivalências entre os termos Knee Arthroplasty e Arthroplasty Knee, pois ambos recuperaram as mesmas 17 ontologias. Todavia, a mesma situação não ocorreu nas buscas com os termos Knee Reconstruction e Reconstruction Knee que na época resultaram em 24 e 19 termos, e em fevereiro de 2018, em 23 e 36 termos respectivamente.

Identificamos que não é estabelecida equivalência entre variações morfológicas correspondentes a número, como pudemos observar nas buscas com os termos Knee Arthroplasty e Knee Arthroplasties que resultaram em 17 e 15 termos respectivamente (os resultados foram os mesmos em Fevereiro de 2018).

** Definição de Knee Replacement no Medline Plus: Knee replacement is surgery for people with severe knee damage. Knee replacement can relieve pain and allow you to be more active. Your doctor may recommend it if you have knee pain and medicine and other treatments are not helping you anymore. When you have a total knee replacement, the surgeon removes damaged cartilage and bone from the surface of your knee joint and replaces them with a man-made surface of metal and plastic. In a partial knee replacement, the surgeon only replaces one part of your knee joint. The surgery can cause scarring, blood clots, and, rarely, infections. After a knee replacement, you will no longer be able to do certain*

activities, such as jogging and high-impact sports (BIORPORTAL, 2018).

A mesma situação foi verificada nas buscas com os termos knee joint replacement e knee joint replacements que recuperaram 24 e 27 termos respectivamente em 2014 e 29 e 25 termos em fevereiro de 2018.

Análise dos mapeamentos de classes por níveis de equivalência e aspectos semânticos, morfológicos, sintáticos e tipográficos

Após a identificação, análise e seleção de termos recuperados, elegemos Knee Replacement*, do MedlinePlus Health Topics (MEDLINEPLUS) cujo sinônimo é Knee Arthroplasty, para análise dos 25 mapeamentos entre classes propostos para o termo. Para fins de construção de estratégias de busca sensibilizadas adotamos os critérios de inclusão: termos em inglês, equivalentes e relacionados; critérios de exclusão: termos em outros idiomas e equivalências inexatas.

Foram feitos quadros (Quadros 4 (em anexo), 5 e 6) com os termos onde eles foram classificados, indicando-se também as inclusões e/ou exclusões e justificativas correspondentes, quando foi o caso.

Foi atribuída numeração sequencial aos termos para possibilitar a identificação de termos em duplicidades (15).

Quadro 5: Exemplo de grupo de termos incluídos e excluídos extraídos do mapeamento entre classes BioPortal para Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS): termo agrupado Arthroplasty of knee.

Numeração Sequencial	Termos Incluídos e Excluídos	Critérios de Inclusão
22	Arthroplasty of the knee - Read Codes, Clinical Terms Version 3 (CTV3)	Incluídos <ul style="list-style-type: none"> • Equivalência exata para <i>Knee Replacement</i> • Variação semântica: uso da palavra <i>arthroplasty</i> no lugar de replacement • Variação morfológica (número, verbo/substantivo com preposição).
24	Arthroplasty of knee - Medical Dictionary for Regulatory Activities	
20	Arthroplasty of knee, NOS - Systematized Nomenclature of Medicine, International Version	
25	Arthroplasty of knee - Systematized Nomenclature of Medicine - Clinical Terms	Excluídos <ul style="list-style-type: none"> • Duplicidades

Fonte: Dados da pesquisa, 2018.

Quadro 6: Exemplo de grupo de termo incluídos e excluídos do mapeamento entre classes BioPortal para Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS): termo agrupado Total replacement of knee.

Numeração Sequencial	Termos Incluídos e Excluídos	Critérios de Inclusão e Exclusão
14	Total replacement of knee - Systematized Nomenclature of Medicine, International Version	Incluídos <ul style="list-style-type: none"> • Equivalência Parcial para <i>Knee Replacement</i>. • Variação Semântica (específico em relação à <i>Knee Replacement</i>) • Variação Sintática: (posição das palavras dentro do termo, com e sem vírgula).
16	Total knee replacement - International Classification of Diseases, Version 9 - Clinical Modification	
19.	Knee total replacement - Medical Dictionary for Regulatory Activities	
17.	Total knee replacement - Medical Dictionary for Regulatory Activities	Excluídos <ul style="list-style-type: none"> • Duplicidades
18.	Total replacement of knee - Medical Dictionary for Regulatory Activities	

Fonte: Dados da pesquisa, 2018.

Após o reagrupamento dos termos*, foram incluídos 8 termos considerados equivalentes aos termos Knee Replacement, e 4 foram excluídos por serem duplicidades geradas pelo mapeamento de termos com mesma designação, mas pertencentes a ontologias distintas. Ou seja, dos 25 mapeamentos de classe propostos pelo BioPortal para o termo Knee Arthroplasty,

* Apresentamos 5 termos incluídos, devido ao limite do número de páginas. Os outros termos incluídos foram: *Prosthetic total arthroplasty of the knee - Read Codes, Clinical Terms Version 3 (CTV3)*; *Arthroplasty, Replacement, Knee - Medical Subject Headings* e *Knee arthroplasty - Medical Dictionary for Regulatory Activities*, conforme ANDRADE (2015).

apenas 7 termos equivalentes foram considerados válidos para compor uma estratégia de busca, dentro dos critérios estabelecidos.

E para finalizar, verificamos que apenas 3 dos 10 termos propostos para Knee Arthroplasty, do Emtree-Embase incluindo o termo preferido constam nos 25 mapeamentos entre classes do termo Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS) no BioPortal-NCBO.

Quadro 7: Verificação de termos do Emtree no mapeamento entre classes de Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS) realizado por BioPortal-NCBO.

Termo preferido e equivalentes Emtree	Consta no mapeamento entre classes de Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS) no BioPortal-NCBO ?
arthroplasty, replacement, knee knee arthroplasty knee replacement	Sim
knee replacements arthroplasty, knee knee arthroplasties knee joint replacement knee joint replacements knee reconstruction reconstruction, knee	Não

Fonte: Dados da pesquisa, 2018.

Como podemos observar, mesmo com os múltiplos mapeamentos realizados no sentido de agrupar as possíveis variações semânticas, sintáticas, morfológicas e tipográficas dos termos, ainda assim algumas possibilidades não foram mapeadas.

CONCLUSÃO

Destinados principalmente aos desenvolvedores de sistemas na criação de aplicações referentes

ao processamento de linguagem natural, à indexação automática da literatura biomédica em bases de dados especializadas, ao mapeamento de sinônimos da terminologia biomédica e à recuperação da informação, o BioPortal, do The National Center for Biomedical Ontology representa um modelo de interoperabilidade e mapeamento entre SOC na área da saúde, embora com algumas limitações geradas pela recuperação por coincidência de caracteres.

De acesso livre, permite o registro de usuários que podem adicionar mapeamentos entre os termos, fazer comentários individuais sobre a ontologia, bem como realizar revisões de ontologias. Segundo Whetzel et al.¹⁶, o conteúdo gerado pelo usuário fornece uma avaliação crítica e mecanismo de feedback para os desenvolvedores de ontologias, o que os autores consideram uma característica distintiva do sistema.

No que concerne aos resultados da interoperabilidade e mapeamento entre classes observa-se que a adoção dos padrões de interoperabilidade na web semântica nem sempre garante o estabelecimento de relações de equivalência correspondentes às variações morfológicas, como as variações de número, variações sintáticas relacionadas à posição das palavras na constituição dos termos, especialmente na língua inglesa. Esse fenômeno é ocasionado pela recuperação por coincidência de caracteres cujo objetivo é aumentar a revocação, mas com comprometimento da precisão.

Considera-se relevante o uso de produtos que integram SOC interoperados e mapeados para a identificação e seleção de termos equivalentes para compor estratégias de busca na área da Saúde, tanto do ponto de vista da recuperação de documentos para elaboração dos trabalhos, como para composição da metodologia de trabalhos na Saúde Baseada em Evidências.

Em resumo, considera-se que BioPortal cumpre uma das funções da interoperabilidade e de mapeamento entre SOC, ou seja, permitir a busca simultânea de termos em vários SOC e/ou ontologias. Entretanto, recomenda-se o conhecimento relativo às formas de busca e recuperação, incluindo as limitações geradas pela recuperação por coincidência de caracteres, assim como a avaliação e a seleção dos termos recuperados. Considera-se que a avaliação e a seleção de termos devem levar em consideração o funcionamento dos SRI onde os termos serão utilizados, pois a inclusão e exclusão de termos equivalentes dependerá da capacidade do sistema em estabelecer as equivalências dos pontos de vista semântico, morfológico, sintático e tipográfico.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Realização de doutorado sanduíche na Universidade de Salamanca, Espanha, 2014.

REFERÊNCIAS

1. Centro Cochrane do Brasil. Medicina baseada em evidências. [Acesso em 01 nov 2017.] Disponível em: <http://www.centrocochranedobrasil.org.br/mbe.html#>.
2. Atallah, ÁN. A incerteza, a evidência e a ciência. *Diagn. Tratamento* 2004; 9(1):27-28, 2004. [Acesso em 02 dez. 2017]. Disponível em: http://www.centrocochranedobrasil.org.br/apl/artigos/artigo_442.pdf.
3. National Center For Biomedical Ontology. BioPortal. [Acesso em: 20 fev. 2018]. Disponível em: <https://bioportal.bioontology.org/>.
- 4 - Association For Library Collections & Technical Services, A Division Of The American Library Association. Committee On Cataloging: Description And Access. Task force on metadata: final report. Chicago: ALCTS, 2000. [Acesso em 07 nov. 2017]. Disponível em: <https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>.
5. National Information Standards Organization. ANSI/NISO Z39.19 – 2005: guidelines for the construction, format, and management of monolingual thesauri. Bethesda: NISO Press, 2005.
6. BSI Group. Structured vocabularies for information retrieval guide: part 4 interoperability. London: BSI, 2007. 55 p.

7. International Standard Organization. ISO 25964: thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneve: International Standard Organization, 2011a.
8. International Standard Organization. ISO 25964: thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. Geneve: International Standard Organization, 2011b.
9. Martinez T, Ana M. et al. Interoperabilidad de sistemas de organización del conocimiento: el estado del arte. *Inf. cult. soc.*, Ciudad Autónoma de Buenos Aires, n. 24, p. 15-37, 2011. [Acesso em 12 abr 2014]. Disponível em: <http://eprints.rclis.org/17198/>.
10. Berners-lee T. Linked data, 2006. [Acesso em 10 set 2014]. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>.
11. Lara MLG. Documentary languages and knowledge organization systems in the context of the semantic web. *Transinformação*, Campinas; 2013; 25(2):145-150, 2013. [Acesso em 10 abr 2017]. Disponível em: <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/articloe/view/1954>.
12. Pastor-Sánchez JA, Martínez-Méndez FJ, Rodríguez-Muñoz JV. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. *El profesional de la información*, Barcelona 2012; 21(3):245-253. [Acesso em 13 maio 2017]. Disponível em: <http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/04.html>.
13. Rodríguez EMM. La Web Semántica: una Web más bibliotecaria. *Boletín de la SEDIC Madrid*; 2004; 41. [Acesso em 21 nov 2014]. Disponível em: http://www.sedic.es/p_boletinclip41_confirm.htm.
14. Castro CC. Vocabularios estructurados, Web Semántica y Linked Data: oportunidades y retos para los profesionales de la documentación. In: Murgia EI; Rodrigues EF (Org.). *Arquivologia, Biblioteconomia e Ciência de Informação: identidades, contrastes e perspectivas de interlocução*. Niterói: UFF, 2012; p. 139-155. [Acesso em 22 mai 2014]. Disponível em: http://gredos.usal.es/jspui/bitstream/10366/121953/3/DBD_UFF_ccaro.pdf.
15. Andrade J. Interoperabilidade e mapeamentos entre sistemas de organização do conhecimento na busca e recuperação de informações em Saúde: estudo de caso em Ortopedia e Traumatologia. 2015. 327 p. (Doutorado). Escola de Comunicações e Artes. Universidade de São Paulo, 2015.
16. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic*

Acids Res. 2011 Jul;39(Web Server issue):W541-5. [Acesso em 15 jan 2018].
Disponível em:
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3125807/pdf/gkr469.pdf>.

Quadro 4: Mapeamento de classes para o termo Knee Replacement do MedlinePlus Health Topics (MEDLINEPLUS) no BioPortal-NCBO.

Numeração Sequencial	Termos e Ontologias	Crítérios de Inclusão e Exclusão: Idioma/Nível de Equivalência
1	artroplastia de la rodilla - SNOMED Terminos Clinicos	Excluir: Espanhol
2	Remplacement total du genou - MDRFRE	Excluir: Francês
3	Arthroplastie du genou - MDRFRE	Excluir: Francês
4	Remplacement total du genou - MDRFRE	Excluir: Francês
5	Arthroplastie totale du genou - MDRFRE	Excluir: Francês
6	Arthroplastie du genou - MDRFRE	Excluir: Francês
7	Arthroplastie prothétique de genou - Thesaurus Biomedical Francais/Anglais [French translation of MeSH]	Excluir: Francês
8	Arthroplasty, knee, tibial plateau; with debridement and partial synovectomy - Current Procedural Terminology	Excluir: equivalência inexata
9	Arthroplasty, knee, condyle and plateau; medial AND lateral compartments with or without patella resurfacing (total knee arthroplasty) - Current Procedural Terminology	Excluir: equivalência inexata
10	Arthroplasty, knee, tibial plateau - Current Procedural Terminology	Excluir: equivalência inexata
11	Arthroplasty, femoral condyles or tibial plateau(s), knee; with debridement and partial synovectomy - Current Procedural Terminology	Excluir: equivalência inexata
12	Arthroplasty, femoral condyles or tibial plateau(s), knee - Current Procedural Terminology	Excluir: equivalência inexata
13	Arthroplasty, knee, condyle and plateau; medial OR lateral compartment - Current Procedural Terminology	Excluir: equivalência inexata
14	Total replacement of knee - Systematized Nomenclature of Medicine, International Version	Incluir: Equivalência Parcial
15	Arthroplasty, Replacement, Knee - Medical Subject Headings	Incluir: Equivalência Parcial
16	Total knee replacement - International Classification of Diseases, Version 9 - Clinical Modification	Incluir: Equivalência Parcial
17	Total knee replacement - Medical Dictionary for Regulatory Activities	Incluir: Equivalência Parcial
18	Total replacement of knee - Medical Dictionary for Regulatory Activities	Incluir: Equivalência Parcial
19	Knee total replacement - Medical Dictionary for Regulatory Activities	Incluir: Equivalência Parcial
20	Arthroplasty of knee, NOS - Systematized Nomenclature of Medicine, International Version	Incluir: Equivalência Parcial
21	Prosthetic total arthroplasty of the knee - Read Codes, Clinical Terms Version 3 (CTV3)	Incluir: Equivalência Parcial
22	Arthroplasty of the knee - Read Codes, Clinical Terms Version 3 (CTV3)	Incluir: Equivalência exata
23	Knee arthroplasty - Medical Dictionary for Regulatory Activities	Incluir: Equivalência exata
24	Arthroplasty of knee - Medical Dictionary for Regulatory Activities	Incluir: Equivalência exata
25	Arthroplasty of knee - Systematized Nomenclature of Medicine - Clinical Terms	Incluir: Equivalência exata

Fonte: Dados da pesquisa, 2018.

Como citar este artigo

Andrade Jde, Lara MLGde. Interoperabilidade e mapeamentos entre sistemas de organização do conhecimento: bioportal do national center for biomedical ontology-ncbo. Revista de Saúde Digital e Tecnologias Educacionais. [online], volume 03, número especial. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, mês e ano, p. 43-61. Disponível em: <http://periodicos.ufc.br/resdite/index>. Acesso em “dia/mês/ano”.

Data de recebimento do artigo: 24/09/2018

Data de aprovação do artigo: 06/11/2018
