DOI: 10.5433/1981-8920.2019v24n2p93

VOCABULÁRIOS CONTROLADOS NA REPRESENTAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO EM REPOSITÓRIOS BRASILEIROS[1]

CONTROLLED VOCABULARIES IN THE REPRESENTATION AND RETRIEVAL OF INFORMATION IN BRAZILIAN REPOSITORIES

Mariângela Spotti Lopes Fujita[a]
Jessica Beatriz Tolare[b]

RESUMO
Introdução: O objetivo do controle de vocabulário é a consistência na representação e na recuperação de modo a evitar a dispersão de recursos informacionais relacionados aos termos do vocabulário controlado. Repositórios digitais são, atualmente, recursos de fundamental importância na organização e gestão do conhecimento cujos recursos de informação, além de armazenados e preservados, precisam ser acessados e recuperados sem inconsistências. Para isso, é importante verificar a incorporação de tipos de vocabulário controlado em repositórios universitários nos processos de representação e recuperação. **Objetivo:** Esta pesquisa propõe a investigação sobre vocabulários controlados em repositórios com o objetivo de verificar recursos de interface de busca e navegação para identificar a incorporação de tipos de vocabulários controlados na representação e recuperação da informação. **Metodologia:** Análise documental, por meio de pesquisa eletrônica, para levantamento de repositórios e pesquisa exploratória para análise dos recursos de interface para identificação de tipos de vocabulários controlados utilizados na representação e recuperação da informação com base em parâmetros obtidos na fundamentação teórica. **Resultados:** A análise das interfaces de busca de 86 repositórios brasileiros revelou que os vocabulários controlados foram, de modo geral, incorporados à representação e recuperação quanto aos seguintes tipos: tesauros (6 repositórios), listas classificadas (8), taxonomias (7) e listas de termos em ordem alfabética (70). Os metadados incluem palavras-chave da linguagem natural e termos de vocabulários controlados em 56 repositórios. **Conclusões:** As listas de termos em ordem alfabética utilizadas para busca por assuntos nas interfaces de navegação necessitam incorporar a função de controle de vocabulário mediante a integração das palavras-chaves com os termos dos vocabulários controlados e obter consistência na recuperação.

[1] O artigo tem apoio do CNPq

[a] Doutora em Ciências da Comunicação pela Universidade de São Paulo (USP). Docente Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). E-mail: mariangelaslf57@gmail.com

[b] Mestranda do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). E-mail: jtolare@gmail.com

Descritores: Vocabulários controlados. Repositórios institucionais universitários Controle de vocabulário. Representação da informação.

1 INTRODUÇÃO

A linguagem natural é, sem dúvida, democrática, acessível e seu uso na indexação social, indiscutivelmente associado às redes sociais, blogs e sites diversos, é pertinente e necessário. Porém, em sistemas de recuperação da informação de universidades, por exemplo, a linguagem científica é inequivocamente associada ao vocabulário científico e tecnológico (BOCCATO; FUJITA; GIL LEIVA, 2011). Mesmo que haja um comportamento de busca com linguagem natural associado aos metabuscadores na web, a terminologia a ser utilizada como assunto em repositórios com uso de palavras-chave é mais especializada e mais precisa mesmo que seja da linguagem natural.

Com a filosofia do acesso aberto e softwares livres o arquivamento da produção científica, tecnológica, artística e administrativa com diferentes tipologias e suportes de recursos informacionais em ambiente digital foi incorporada por repositórios que, incialmente, tinha como objetivo principal a armazenagem, mas a política científica avançou rapidamente e, com isso, a necessidade de organização do conhecimento para facilitar a recuperação e consequente visibilidade para aplicações de métricas de avaliação acadêmica.

O maior problema na recuperação da informação é ainda associado à falta de controle de vocabulário (MONDOUX; SHIRI, 2009) que pode gerar vários tipos de inconsistências em qualquer sistema de busca. A falta de controle de vocabulário afeta também a visibilidade das publicações arquivadas no repositório bem como a interação do sistema com o usuário final.

A Norma ISO 25964-2 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p.16), sobre tesauros para a recuperação da informação, considera o controle de vocabulário essencial porque no discurso comum um termo pode ter mais de um significado e a escolha de um termo preferido para representar um conceito específico nunca é direta porque conceitos podem ser expressos de várias maneiras. Por isso, o tesauro, como vocabulário controlado, tem função importante na mediação entre os termos utilizados no discurso e aqueles que funcionam efetivamente para a recuperação da informação. Essa função implica o

usuário e o profissional da informação aceitar um grau de artificialidade no vocabulário controlado para o alcance de benefícios na recuperação.

A pesquisa levada a cabo por Weitzel (2019) para mapeamento dos repositórios institucionais brasileiros levantou resultados expressivos sobre número de itens arquivados, fomentos recebidos e infraestrutura institucional de apoio, e avaliação da visibilidade e impacto. Para o mapeamento a autora utilizou as seguintes fontes específicas: *Open DOAR*, ROARMAP, ROAR, Lista de repositórios do IBICT e resultado do edital da FINEP que resultou na identificação de 101 repositórios. Anteriormente, a pesquisa de Andrade e Cervantes (2012) sobre interoperabilidade semântica realizou investigação sobre controle de vocabulário com aplicação de questionário para instituições universitárias e constatou que em 26 repositórios nenhum repositório disponibiliza vocabulário controlado para os usuários consultarem durante a busca. Em âmbito canadense, Mondoux e Shiri (2009) realizaram investigação sobre os recursos de interface do usuário e incorporação de sistemas de organização do conhecimento (SOC) aos repositórios de instituições canadenses cujos resultados demonstram que poucos repositórios incorporaram SOCs mais complexos como vocabulário controlados.

Portanto, o interesse de investigação deste trabalho tem como objeto de análise os vocabulários controlados em repositórios com o objetivo de verificar recursos de interface do usuário para identificar tipos de vocabulários controlados na representação e recuperação da informação. O presente trabalho procurou determinar se os repositórios brasileiros incorporaram vocabulários controlados em metadados e interfaces de busca e navegação, para representação de assuntos e recuperação da informação. A importância de estudos exploratórios dessa natureza está na necessidade de observar, analisar, discutir, divulgar e compartilhar mudanças de aprimoramento de representação e recuperação de conteúdos digitais tal como a integração de vocabulários controlados e linguagem natural em repositórios, fenômeno absolutamente natural e observável em outros estudos internacionais o que, nos dá a vantagem de fazer comparações a partir da situação atual brasileira.

2 VOCABULÁRIOS CONTROLADOS EM REPOSITÓRIOS

Zeng e Chan (2004) relatam que o advento da internet possibilitou ir além das

capacidades dos tradicionais serviços de base de dados e de bibliotecas com o aparecimento dos primeiros repositórios, no final da década de 90. Diferentes comunidades construíram seus repositórios e desenvolveram novas diretrizes e arquiteturas que incluem novos padrões de metadados e sistemas de organização do conhecimento. Nesse sentido, tesauros pequenos e individuais de áreas pequenas e específicas foram construídos sem seguir padrões convencionais, porém, com objetivo de garantir a consistência na indexação e busca da informação.

Atualmente, repositórios digitais têm cumprido importante função na política do acesso aberto, principalmente, em universidades que necessitam realizar a comunicação científica com preservação e acesso a publicações e documentos digitais. A iniciativa de arquivos abertos ou *Open Archives Initiative* (OAI) permitiu o "uso do software livre, também chamado *open source*, auto-arquivamento, criação de repositórios de livre acesso, tanto institucionais quanto temáticos, e uso de padrões de preservação de objetos digitais" (KURAMOTO, 2005, p. 148).

Monteiro e Brascher (2007) distinguem os repositórios em temático e institucional pelas diferenças entre estruturas e características próprias. Para estabelecer as diferenças entre um e outro, consideram que "O repositório temático tem a preocupação de armazenar documentos com uma delimitação concisa de sua cobertura designada por um assunto, área do conhecimento ou temática específica" e que "O **repositório institucional** (RI) é a reunião de RTs sob a responsabilidade técnica e administrativa de uma instituição ou organismo" (MONTEIRO; BRASCHER, 2007, p.3).

Em universidades, os repositórios são institucionais e contêm toda a produção científica das áreas do conhecimento abrangidas pelas atividades acadêmicas à semelhança de um catálogo bibliográfico online. Essa circunstância é própria do ambiente universitário onde a produção científica abrange diferentes áreas do conhecimento e são especializadas com inovação científica e tecnológica. Ser precursora em geração de conhecimentos é condição natural da existência da universidade e significa que todos os seus sistemas de informação precisam acompanhar esta evolução na organização do conhecimento para sua gestão.

Os resultados de pesquisa sobre o uso de vocabulário controlado para dados científicos com participantes do sistema DataNet[1] dos Estados Unidos, indicam que são considerados valiosos e que o vocabulário controlado mais utilizado é a LCSH. Para os autores, Tudhope, Kock e Heery, 2006; Andrade e Cervantes, 2012; Zhang, et al, 2015, a explicação é que os vocabulários controlados ajudam na interoperabilidade semântica entre o conteúdo e os metadados e sobre isso, consideram que esta prática, historicamente comum em catálogos de bibliotecas, tem um grande potencial para repositórios de dados, onde sua aplicação não é tão comum (Zhang, et al, 2015, p.2).

Harper e Tillet (2009) discutem como vocabulários controlados, esquemas de classificação e tesauros, podem ajudar na construção da Web Semântica com uso compartilhado do controle de autoridade. A principal ajuda, segundo os autores, seria fazer uso efetivo das hierarquias, relações de equivalência e estruturas do conteúdo de vocabulários controlados contidos no registro de autoridade MARC. Acreditam que esse uso auxilia a busca na consulta de outros termos desconhecidos que fazem parte da estrutura hierárquica de um determinado termo ou que, também, seja equivalente ou associado.

Em realidade, o controle de vocabulário em repositórios institucionais não é uma tarefa fácil para seus coordenadores. Em pesquisa sobre análise do controle de autoridade, Barrionuevo Almuzara, Alvite Díez e Rodríguez Bravo (2012, p.98) utilizaram dez critérios de análise em uma amostra de 35 repositórios espanhóis. Entre os dez critérios de análise estava o de ferramentas terminológicas usadas para o acesso temático: palavras-chaves, cabeçalhos de assuntos, descritores controlados, classificações e similares. Com os resultados obtidos da análise desse critério, os autores observaram que “[...] a função principal foi manipulada por listas de cabeçalhos de assunto e palavras-chave e, em menor grau, classificações, tesauros e listas de descritores.” e que, com a opção de autoarquivamento as pessoas que hospedam documentos podem determinar suas próprias palavras-chaves sem consulta aos assuntos que o sistema oferece. Concluem que “[...] o volume de termos

[1] O Consórcio da Federação DataNet é um projeto dentro da iniciativa NSF Office of Cyber-InfrastructureDataNet. O objetivo do Consórcio da Federação DataNet é reunir a infra-estrutura nacional de dados que permite a pesquisa colaborativa, através da federação de infra-estrutura de gerenciamento de dados existente. O DFC busca colaborações através de seis comunidades de prática que organizam conhecimentos nacionais. Disponível em: http://datafed.org/about/.

sem controle que podem ser incluídos em repositórios não é limitado, uma circunstância que parece exigir alguma forma de padronização."

Sistemas de organização do conhecimento (SOCs) são importantes porque realizam o controle de vocabulário em sistemas de informação que necessitam prover a recuperação da informação armazenada, seja de forma retrospectiva ou atualizada. Atuam diretamente na organização e representação do conhecimento e quando isso não ocorre, a recuperação da informação é realizada com a linguagem natural presente no discurso dos conteúdos informacionais.

No entanto, pela evolução das tecnologias de informação e no contexto da internet, surgiram sistemas de organização do conhecimento mais adequados às necessidades de organização de documentos nativos digitais que desempenham um importante papel no gerenciamento e aplicações de informações digitais em geral. Dotados de controle semântico, os sistemas de organização do conhecimento realizam funções importantes:

> [...] ferramentas para representação e indexação de informações e documentos, suporte baseado em conhecimento para pesquisadores de informação, roteiros semânticos para domínios e disciplinas, ferramentas de comunicação para fornecer estrutura conceitual, e uma base conceitual para sistemas baseados em conhecimento. (MAYR et. al., 2016)

Para entender a diversidade de tipologias de sistemas de organização do conhecimento em bibliotecas digitais, Hodge (2000) em estudo junto com *Networked Knowledge Organization Systems/Services Group*, foi pioneira em declarar que os sistemas de organização do conhecimento possuem características distintas para determinados contextos de aplicação e podem ser classificados em duas categorias:

- Classificação e categorização: sistemas de categorização para qualquer sistema organizado; sistemas de classificação bibliográficos; listas de cabeçalhos de assunto; taxonomia;

- Modelos de relacionamento: ontologias; redes semânticas; tesauros.

Em estudo mais abrangente, Hill *et al*. (2002), com base em Hodge (2000) incluem outros tipos mais específicos de sistemas de organização do conhecimento existentes em bibliotecas digitais e propõem classificação com quatro categorias mantendo duas categorias de Hodge (2000):

- Classificação e Categorização: Esquemas de categorização; Esquemas de classificação; Cabeçalhos de assunto; Taxonomias.

- Modelos semelhantes a metadados: Diretórios com listas de nomes e suas informações de contato associadas; Dicionários geoespaciais.

- Modelos de Relacionamento: Ontologias; Redes semânticas; Tesauro.

- Listas de Termos: Arquivos de autoridade; Dicionários; Glossários.

Zeng e Chan (2004) consideram as mesmas tipologias de sistemas de organização do conhecimento, porém as classificam em 3 categorias com denominações diferentes:

- Listas de termos: glossários, arquivos de autoridades, dicionários e dicionários geográficos;

- Esquemas de classificação e categorização: esquemas de classificação de bibliotecas, taxonomias e esquemas de categorias;

- Vocabulários relacionais: listas de cabeçalhos de assuntos, tesauros, redes semânticas e ontologias.

Em estudo mais avançado sobre as estruturas e características comuns de sistemas de organização do conhecimento, Zeng (2008) com base em Hodge (2000), Hill et al. (2002) e norma NISO (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010), mantém as denominações das quatro categorias e propõe ordenação crescente pelo tipo de estrutura da mais simples à mais complexa:

- Lista de termos: glossários, dicionários, listas de seleção, anéis de sinônimos;

- Modelos semelhantes a metadados: arquivos de autoridades, diretórios, dicionários geográficos;

- Classificação e categorização: listas de cabeçalhos de assunto**,** esquemas de classificação de bibliotecas, taxonomias e esquemas de categorias;

- Modelos de relacionamento: tesauros, redes semânticas e ontologias.

Em perspectiva mais abrangente e generalista com muito mais tipos de sistemas de organização do conhecimento, Souza, Tudhope e Almeida (2012) assumem abordagem *lato sensu* que considera “todos os tipos de esquemas para organizar a informação e promover a gestão do conhecimento” (HODGE, 2000) e propõem quatro categorias classificatórias expostas em um mapa conceitual com visualização gráfica de diversos tipos de SOCs: *Textos não estruturados* (resumos, concordâncias e substitutos); *Conceitos, relacionamentos e estruturas de layout* (por exemplo, mapas conceituais); *estruturas de conceitos e relacionamentos* (taxonomias,

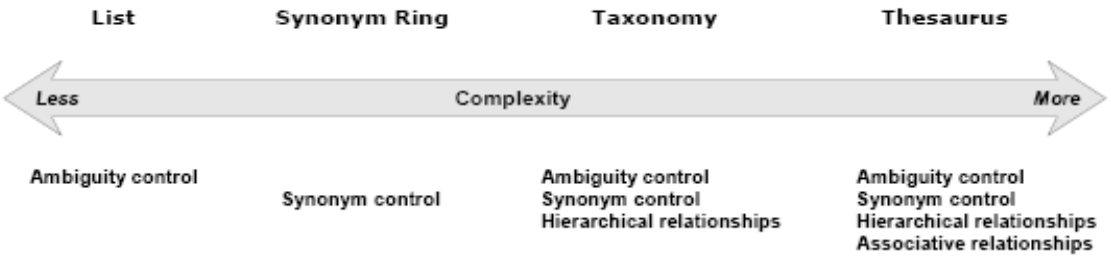
tesauros, taxonomias corporativas, taxonomias de *websites*, ontologias, vocabulários controlados, cabeçalhos de assuntos, esquemas de classificação, entre outros); e *Listas de termos e/ou conceitos* (*folksonomias*, dicionários, arquivos de autoridades, anéis de sinônimos, glossários, dicionários de sinônimos, dicionários geográficos e etc). A ideia dos autores foi abarcar os vários e diferentes tipos de SOCs que surgiram até então.

De fato, essa abordagem *lato sensu* é necessária em função de novos serviços e aplicativos de sistemas de organização do conhecimento em rede em atendimento a novos padrões em serviços da Web. Tudo isso é reflexo das abordagens centradas no usuário que estão produzindo modelos mais sofisticados de como os usuários buscam informações (TUDHOPE; KOCH, 2004; MAYR *et al*., 2016).

Tendo em vista que o vocabulário controlado foi incluído como um dos tipos de sistemas de organização do conhecimento, de acordo com Souza, Tudhope e Almeida (2012), levanta-se aqui uma discussão anteriormente realizada por Barité (2011) acerca das variações terminológicas entre sistemas de organização do conhecimento, vocabulário controlado e linguagem documentária. Para o autor, sistemas de organização do conhecimento é um termo geral e mais abarcativo dada a diversidade de tipos que atualmente existem e que, provavelmente, virão a existir com a finalidade de organizar o conhecimento.

Por outro lado, autores como Lancaster (2004), Hjorland (2008), Zeng (2005), Moreira e Moura (2006), Terra e Vasconcelos (2015) incluem a variação terminológica de vocabulário controlado em suas publicações para se referirem a diferentes tipos de estruturas. Nesse caso, é preciso considerar, também, a Norma ANSI/NISO Z39.19-2005 (R2010) que dedica-se às "*Guidelines for the construction, format, and management of monolingual controlled vocabularies*" (AMERICAN NATIONAL STANDARDS INSTITUTE, 2010) e classifica quatro diferentes tipos de vocabulários controlados: Lista, anel de sinônimos, taxonomia e tesauro. Entre esses vocabulários controlados existem diferentes estruturas construídas a partir de relacionamentos entre termos autorizados contidos em seus vocabulários. Portanto, os quatro tipos de vocabulários controlados foram determinados pelo menor e maior grau de complexidade de suas estruturas como demonstrado pela Figura 1:

Figura 1 - *Increasing structural complexity of controlled vocabullaries*



Fonte: AMERICAN NATIONAL STANDARDS INSTITUTE/NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p.17

Observa-se que em cada um dos tipos de vocabulários existe pelo menos um processo de controle de vocabulário e nos mais complexos mais de um processo. Se considerarmos a definição de controle de vocabulário da Norma ANSI/NISO Z39.19-2005 (R2010) veremos que os processos de controle de vocabulário são semânticos e a adição de mais de um processo torna a estrutura de termos do vocabulário mais complexa:

> O processo de organizar uma lista de termos (a) para indicar qual dos dois ou mais termos sinônimos é autorizado para uso; (b) distinguir entre homógrafos; e (c) indicar relações hierárquicas e associativas entre termos no contexto de um vocabulário controlado ou lista de cabeçalhos de assunto. (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p. 10)

Fundamentado em análise da literatura normativa sobre vocabulários controlados (BRITISH STANDARDS INSTITUTION, 2005; NATIONAL INFORMATION STANDARDS OGANIZATION, 2010; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011), Barité (2014), identifica cinco objetivos do controle de vocabulário: 1. Restrição com relação à quantidade de termos, extensão e alcance de cada termo, e, quantidade de relações entre termos; 2. Desambiguação entre termos homônimos e polissêmicos, variantes, sinônimos e/ou quasi-sinônimos; 3. Normalização para controle de autoridades; 4. Hierarquização; 5. Relação intratermos (sinônimos, quasi-sinônimos e variantes) e entre termos (relações hierárquicas e associativas). Esses objetivos esclarecem, também, o grau de complexidade das estruturas de vocabulários controlados.

Neste trabalho adotaremos a expressão vocabulários controlados e passaremos a analisar com exemplos cada um dos quatro tipos de estruturas conforme classificação e definições de Zeng (2008) das quatro categorias: Lista de

termos, Modelos semelhantes à metadados, Classificação e categorização e Modelos de relacionamento.

Lista de termos – "[...] é um conjunto limitado de termos organizados como uma lista alfabética simples ou de algum outro modo logicamente evidente. " (ZENG, 2005; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010). Exemplos:

Quadro 1 - Listas de termos

Listas alfabéticas simples:		**Listas lógicas:**	
Acre Alagoas Amapá Ceará Espírito Santo Goiás	Abastecimento de alimentos Abrigo Banco central Base de dados Capacitação	Mercúrio Venus Terra Marte Júpiter Saturno	Educação Educação infantil Escola Ensino Aprendizagem Leitura Estágio supervisionado

Fonte: Criada pelos autores com base em Zeng (2005)

Os tipos de vocabulários controlados em listas de termos são: glossários, dicionários, listas de seleção, anéis de sinônimos

Glossários – listas alfabéticas de termos com definições com indicação de uso;

Exemplo:

Autofagia - Auto = por si / fagia = comer; usamos este termo para designar o ato de autodigestão. Ocorre em células ou tecidos que por liberarem enzimas digestivas dentro de suas estruturas acabam por fazer autodigestão.

Fonte: Glossário biológico[2]

Dicionários – listas alfabéticas de termos com indicações ortográficas, sintáticas e de significados.

Exemplo:

Significado de Coleção Substantivo feminino Reunião de objetos da mesma natureza: coleção de livros. Conjunto de objetos escolhidos por sua beleza, raridade, valor: coleção de selos, coleção de quadros. Fonte: Dicionário online de português.

Fonte: Dicionário Online de Português[3]

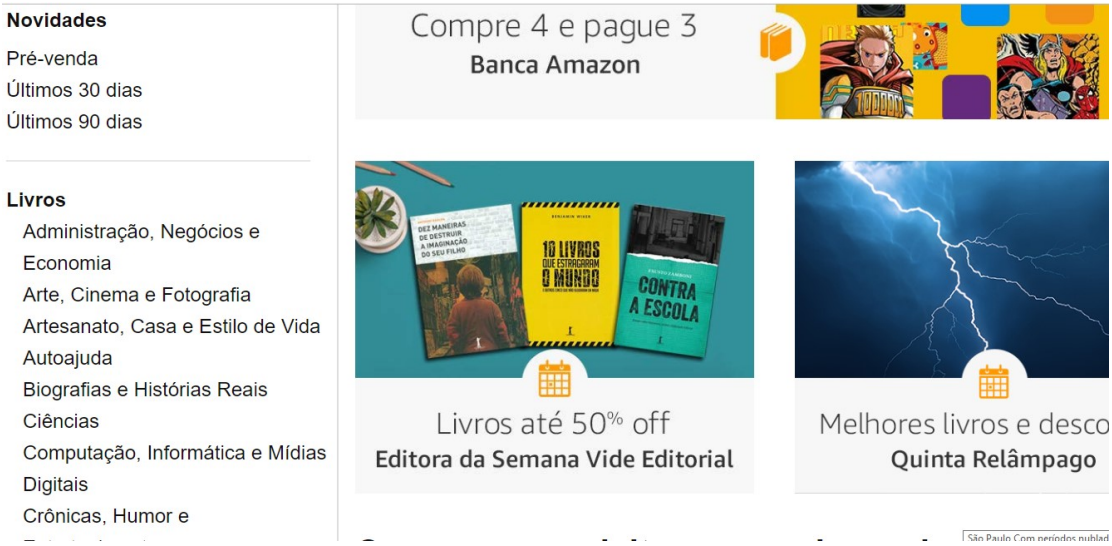
Listas de seleção – conjunto limitado de termos em ordem sequencial para auxílio à seleção durante a busca. São atualmente muito utilizadas em *websites* institucionais, governamentais, comerciais e outros para organização de coleções de itens documentais ou informacionais.

2 Disponível em: https://www.sobiologia.com.br/conteudos/Glossario/a.php

3 Disponível em: https://www.dicio.com.br/colecao/

Exemplo lista de seleção de um *website* comercial:

Figura 2 - Lista de seleção de um *website* comercial

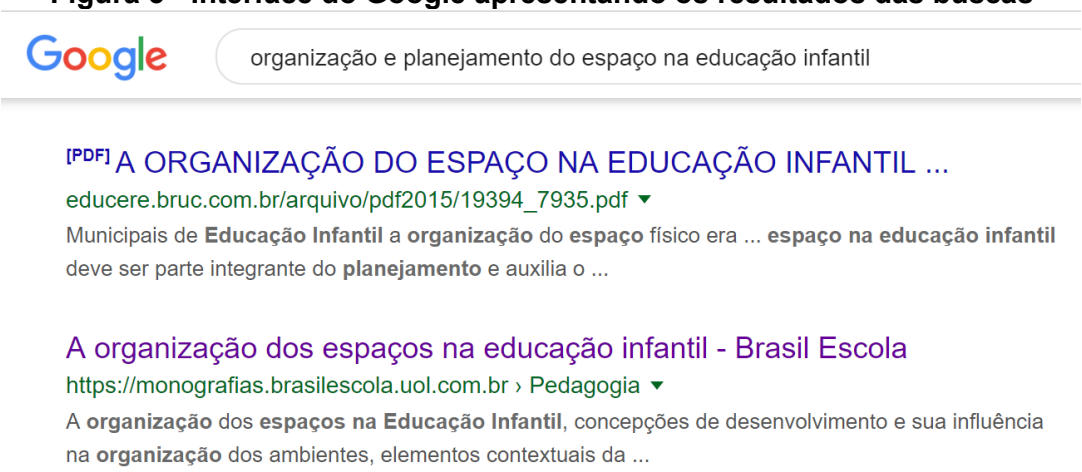


Fonte: https://www.amazon.com.br/Livros/b?ie=UTF8&node=6740748011

<u>Anéis de sinônimos</u> – conjunto de termos considerados equivalentes para fins somente de recuperação. Anéis de sinônimo são geralmente usados na interface em um sistema de informações eletrônicas e fornecem acesso a conteúdo representado em linguagem natural e não controlada. Não estabelecem hierarquias, sinônimos ou quase-sinônimos mas são intuitivos para o usuário estabelecer conexões cognitivas.

Exemplo 1: Resultados da interface de busca do Google com realce das palavras de busca localizadas.

Figura 3 - Interface do Google apresentando os resultados das buscas



Modelos semelhantes a metadados:

Arquivos de autoridades - listas de termos ou de nomes usadas para controle de variações de nomes de uma entidade ou pessoa ou de assuntos (ZENG, 2008). No caso de assuntos, são incluídos relacionamentos entre os termos além das variações. "As listas de autoridade de nomes devem identificar o objeto nomeado de forma exclusiva e fornecer acesso ao registro de entidade nomeado por nomes de variantes e formas de nome." (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p.81-82).

Exemplo 1 - Registro de autoridade de nome de Arquivo de autoridades de nomes.

Descrição	**Abdalad, Luciana Silva**
Remissiva Ver (US/UF)	Silva Abdalad, Luciana
Fonte positiva dos dados	Autora de: Mulheres e voô livre. c2005: (orelha: mestre em educação física pela UGF)
Fonte negativa dos dados	http://authorities.loc.gov 19/08/08
Fonte negativa dos dados	EDA 19/08/08
Fonte negativa dos dados	CNPq 19/08/08

Fonte: Registro de autoridade de nome do Catálogo de autoridades de nomes da Biblioteca Nacional do Brasil. Disponível em: http://acervo.bn.br/sophia_web/index.html

Exemplo 2 - Registro de autoridade de assuntos de Arquivo de autoridades de assuntos

Indexing

URI(s)
http://id.loc.gov/authorities/subjects/sh85064867
info:lc/authorities/sh85064867
http://id.loc.gov/authorities/sh85064867#concept
Instance Of
MADS/RDF Topic
MADS/RDF Authority
SKOS Concept
Scheme Membership(s)
Library of Congress Subject Headings
Collection Membership(s)
LCSH Collection - Authorized Headings
LCSH Collection - General Collection
LCSH Collection - May Subdivide Geographically
Variants
Books--Indexing
Index preparation

Preparation of indexes
Subject analysis
Broader Terms
Information organization
Narrower Terms
Automatic indexing
Chain indexing
Coordinate indexing
Punched card systems
Subject cataloging
Related Terms
Filing systems
Closely Matching Concepts from Other Schemes
Indexation (documentation)
Indexierung <Inhaltserschließung>
Indexing
indexing (function)
Inhaltsanalyse

Fonte: Registro de autoridade de assunto do arquivo de autoridade de assuntos da Library of Congress. Disponível em: http://id.loc.gov/authorities/subjects/sh85064867.html

Diretórios – Listas de nomes e suas informações de contatos.

Figura 4 - Diretório de nomes de empresas e profissionais no Brasil

Fonte: diretórioempresas.com.br – Empresas e profissionais no Brasil[4]

Dicionários geográficos (*Gazetters*) – Dicionários geoespaciais de locais nomeados e caracterizados (ZENG, 2008, p.161). Em definição da Wikipedia (https://pt.wikipedia.org/wiki/Gazetteer) , *gazetters* ou dicionários geográficos ou Índice de topónimos, é um dicionário geográfico ou diretório de lugares e nomes de lugar.

[4] Disponível em: http://www.diretorioempresas.com.br/empresas/bombeiros_60

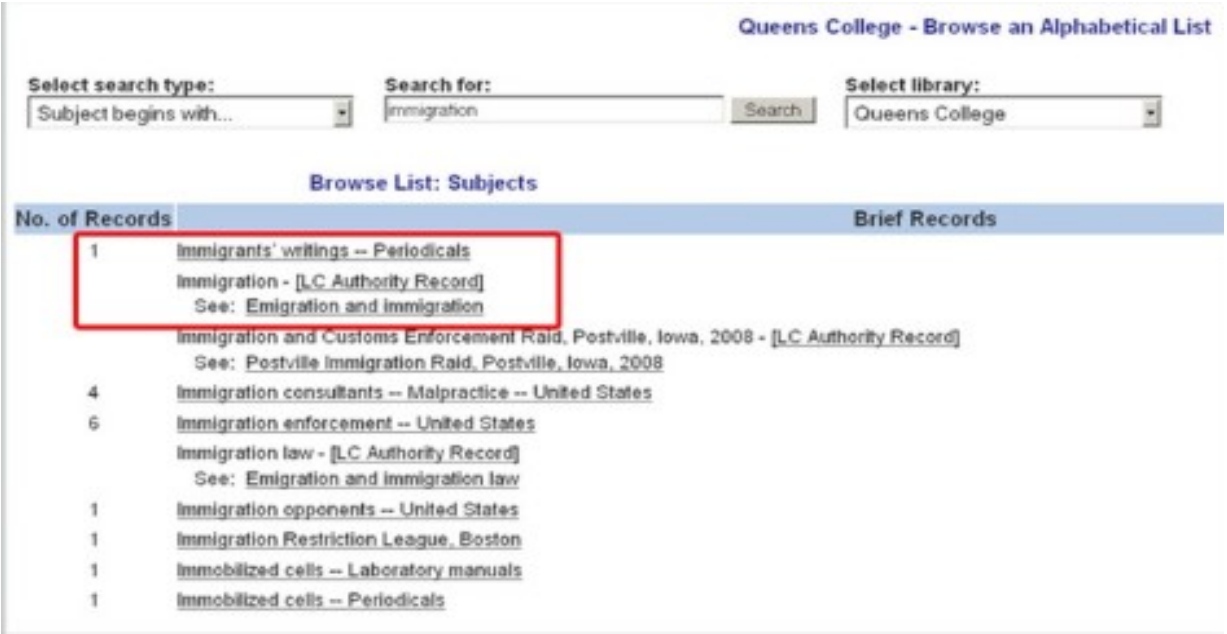
Figura 5 - Website de busca com índice toponímico (parte de cima)

Fonte: Guia Geográfico: Mapas do Brasil e do mundo. Disponível em: https://www.guiageo.com/

Classificação e categorização:

Listas de cabeçalhos de assunto: "são sistemas basicamente classificatórios, formados por um conjunto de termos para representar os assuntos dos itens de uma coleção de acordo com regras estabelecidas de combinação de termos na constituição de cabeçalhos compostos. " (BOCCATO, 2011, p.167).

Figura 6 - Lista de cabeçalho de assunto



Fonte:https://slideplayer.com.br/slide/10351500/)

Esquemas de classificação de bibliotecas: é considerado como o processo de reunir entidades, ideias ou seres em grupos, de acordo com a sua semelhança. Ao longo da História foram criados diversos tipos de classificação a classificação Baconiana; Classificação de Locke, Classificação de Comte; Classificação de Perason, Classificação de Brown, etc. Atualmente, as classificações mais utilizadas pelas bibliotecas são a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU) (SANTOS, 2018).

Exemplo: Categorias Divididas no sistema de Classificação Decimal de Dewey (CDD) e no sistema de Classificação Decimal Universal (CDU).

Figura 7 - Campos de classificações gerais da CDD e da CDU

CDD	CDU
000 GENERALIDADES	0 GENERALIDADES
100 FILOSOFIA	1 FILOSOFIA
200 RELIGIÃO	2 RELIGIÃO
300 CIÊNCIAS SOCIAIS	3 CIÊNCIAS SOCIAIS
400 LINGUISTICA	4 VAGA
500 CIÊNCIAS PURAS	5 CIENCIAS PURAS
600 CIÊNCIAS APLICADAS	6 CIÊNCIAS APLICADAS
700 ARTES	7 ARTES, RECREAÇÃO. DIVERSÃO. ESPORTES
800 LITERATURA	8 LINGUISTICA. LITERATURA
900 HISTORIA. GEOGRAFIA. BIOGRAFIA	9 HISTORIA. GEOGRAFIA. BIOGRAFIA

FONTE: SILVA, 2015, p.7

Destaca-se a aplicação de sistemas de classificação para negócios, considerado conjunto de termos organizados de forma hierárquica, a partir da superordenação e subordinação existente entre si. Possui o intuito de tratar, classificar, controlar informações geradas por uma corporação. Sua forma de classificação pode ser através da cronologia, assuntos gerias organizados de forma alfabética, tipo do negócio. (BOCCATO, 2011, p. 178).

Taxonomias: “[...] são divisões de itens ordenados em grupos ou categorias baseando-se em suas características específicas. ” (BOCCATO, 2011). A taxonomia possui uma forma tradicionalmente hierárquica, contudo, pode ser representa por listas, árvores, hierarquias, sistemas de mapa. (LAMBE, 2007).

Figura 8 - Taxonomia simples e expandida

Fonte: Figueiredo, 2007

Na literatura é apresentado alguns tipos de taxonomias. Sendo elas:

Taxonomia canônica de unidades sistemáticas: considerada uma classificação binária (dicotômica) que revela relações de família, gênero, espécie. Não permitem a agregação de novos assuntos depois de sua construção. (CAMPOS; GOMES, 2008):

Taxonomia de domínio: considerada policotômica, voltada para a representação de um domínio do conhecimento. Dessa forma, possui diversas divisões possíveis. Normalmente, seu recorte é feito por categorias e determinadas pelas características da organização em que estão inseridos, possibilitando a subdivisão. Possibilita uma organização que representa classe de conceitos com um princípio de divisão. (CAMPOS; GOMES, 2008).

Figura 9 - Taxonomia de domínio na área da Geologia

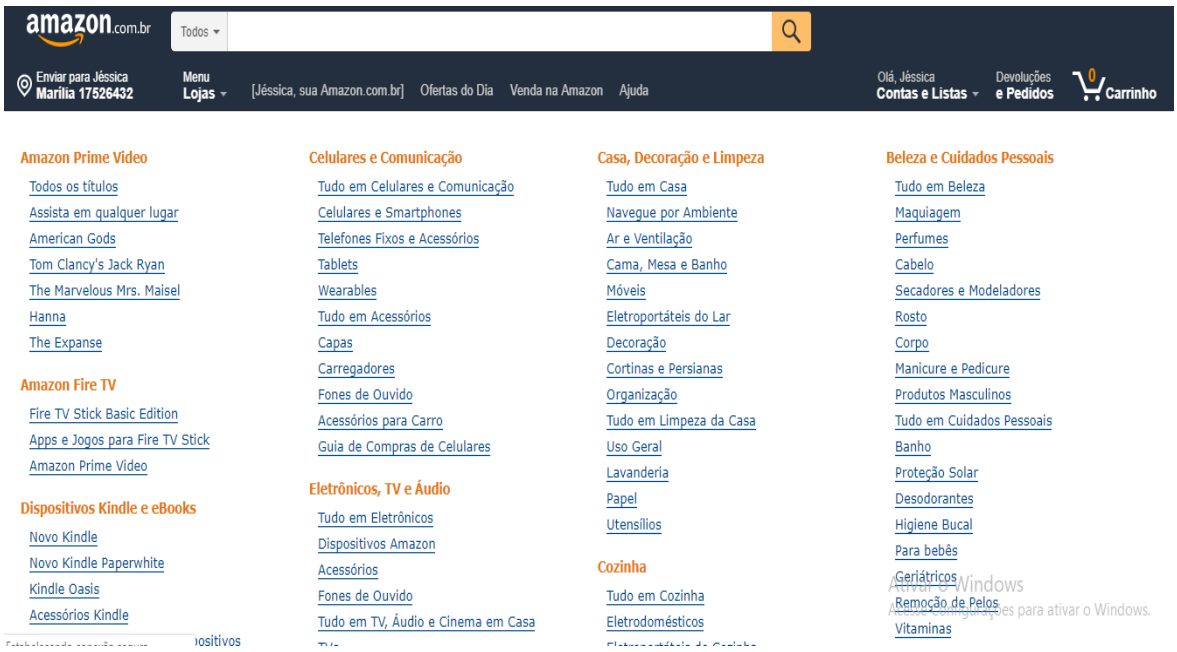
Personalidade	Domínio do Saber	Energia	Espaço	Tempo
*água (hidrologia) ** água das chuvas **água doce **água termal *argila *bacia **bacia hidrográfica ** bacia sedimentar *estrutura da terra **hidrosfera **litosfera (por partes) ***crosta terrestre ***rocha (por seus tipos) ****rocha eruptiva ****rocha sedimentar ****rocha metamórfica ****calcário ****granito ****xisto betuminoso	*geociência (por partes) **cristalografia **estratigrafia **geodésia **geofisica ***geomagnetismo **espeleologia **geografia (por partes) ***geografia fisica **hidrografia **geologia (por partes) ***vulcanologia ***sedimentologia ***petrologia ***geologia estratigráfica ***geologia fisica (por partes) ****geomorfologia	*catástofre **furacão **maremoto **seca **terremoto *conservação do solo *erosão *previsão do tempo *translação continental	*garimpo *mina(mineralogia) *manancial *mangue(geologia)	*era geológica (por partes) **cenozóico **mesozóico **paleozóico

Fonte: (SALES; MOTTA, 2010)

Esquemas de categorias: qualquer sistema que possui uma organização. (BOCCATO, 2011).

Exemplo: Organização das categorias no site do Amazon

Figura 10 - Categorias do Amazon

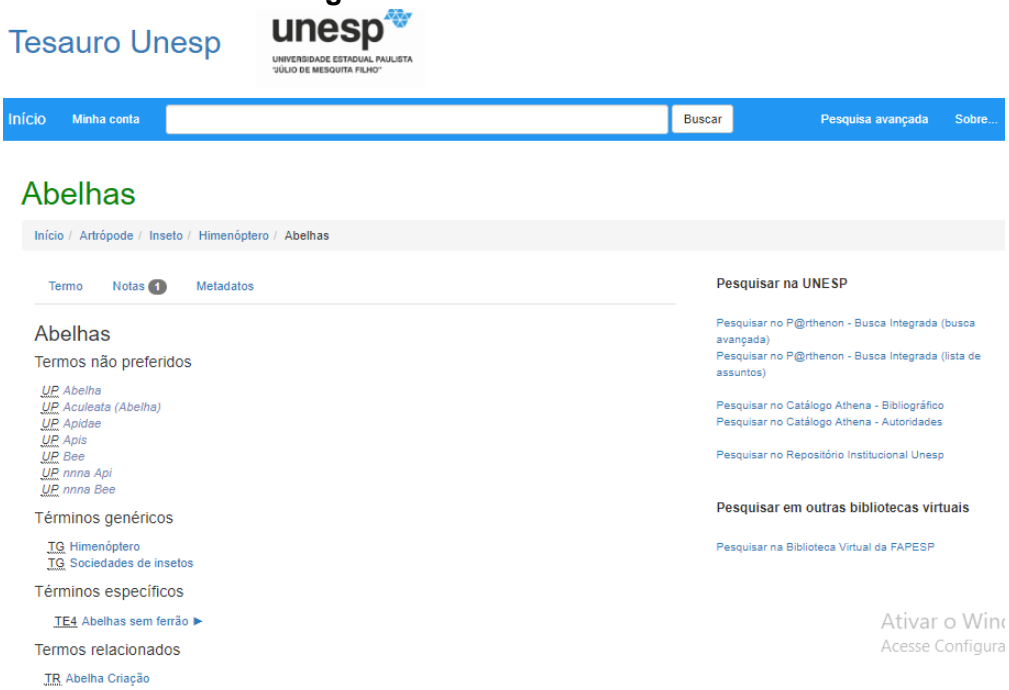


Fonte: Disponível em: https://www.amazon.com.br/gp/site-directory?ref_=nav_shopall_btn

Modelos de relacionamento

Tesauros: "[...] são vocabulários controlados formados por termos descritores semanticamente relacionados, atuando como instrumentos de controle terminológico" (SALES; CAFÉ, 2009, p. 102). Os tesauros podem possuir dois tipos de estruturas: hierarquicamente (gênero-espécie) e por associação semântica, utilizados para realizar a indexação e recuperar as informações através de seu conteúdo (SALES; CAFÉ, 2009).

Figura 11 - Tesauro da UNESP



FONTE: https://www.biblioteca.unesp.br/tesauro/vocab/index.php?tema=13283&/abelhas

<u>Redes semânticas</u> - Conjunto de termos representados como nós interconectados em uma notação gráfica. São úteis para sistemas automatizados de inferências sobre o conhecimento (ZENG, 2008; BOCCATO, 2011).

Figura 12 - Rede semântica dos subdomínios do *Unified Medical Language System (UMLS)*

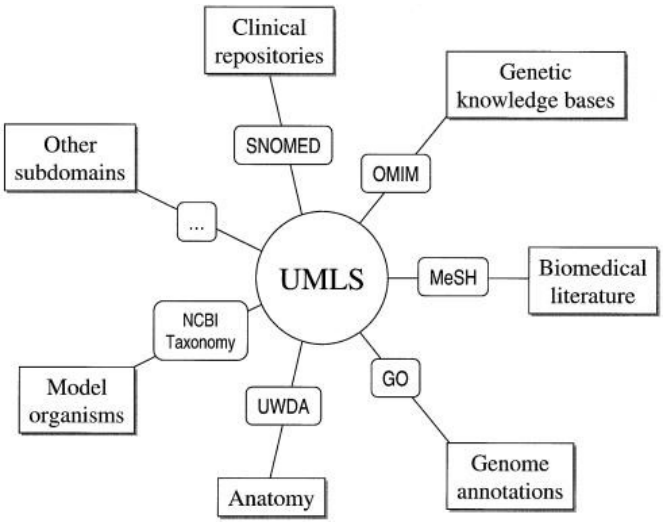


Figura: *The various subdomains integrated in the UMLS*.
Fonte: (BODENREIDER, 2004, p.268)

Ontologias – "Modelos conceituais específicos representando relações complexas entre objetos" (ZENG, 2008, p.162). Contém entidades e elementos que se relacionam entre si em ambientes digitais. Diferencia-se das redes semânticas ao incluir as regras e axiomas (BARITÉ, 2011).

Figura 13 - Representação dos relacionamentos entre entidades e elementos com regras e axiomas em uma ontologia.

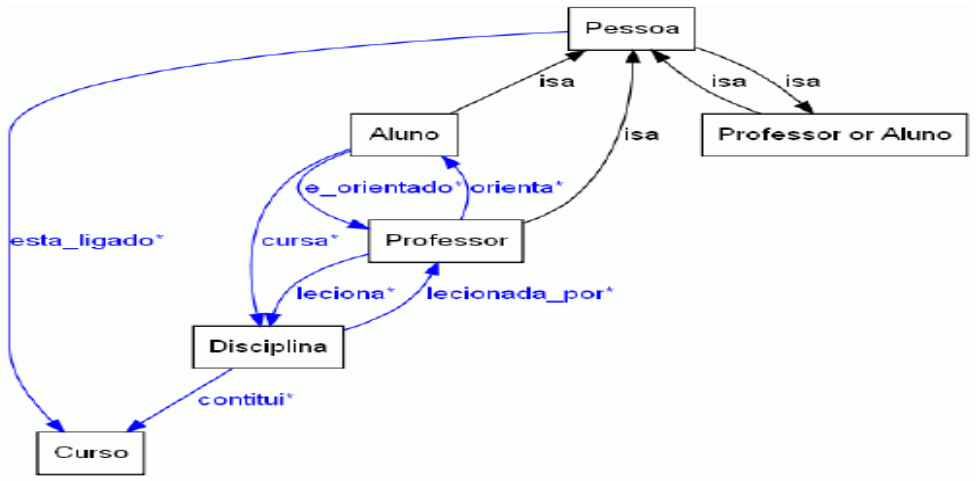


FIGURA 2 – Relacionamentos de uma ontologia

Fonte: RAMALHO, 2009, p.9

De fato, sistemas de organização do conhecimento ou vocabulários controlados apresentam estruturas mais ou menos complexas conforme necessidade de representação de áreas do conhecimento e todos possuem processo de controle de vocabulário em maior ou menor grau de complexidade.

3 METODOLOGIA

Esta pesquisa é exploratória para realizar observação e análise dos recursos de interface de busca e tipos de vocabulários controlados para a representação e recuperação da informação. Em metodologia da pesquisa, Gil (2010) destaca a necessidade e importância da pesquisa exploratória para a familiarização com o objeto da pesquisa a fim de torna-lo mais explícito para o levantamento dos problemas de pesquisa. Para isso, a coleta de dados utilizada nesta pesquisa foi a análise dos recursos de interface para identificação de tipos de vocabulários controlados utilizados na representação e recuperação da informação em repositórios brasileiros com base

em parâmetros obtidos na fundamentação teórica sobre vocabulários controlados. A investigação compõe-se de quatro etapas de identificação:

- Identificar os repositórios.
- Identificar e analisar recursos de interface de pesquisa e navegação em Repositórios.
- Identificar quais repositórios incorporaram vocabulários controlados e como eles são incorporados em suas interfaces de pesquisa e navegação.
- Identificar os tipos e as principais características de vocabulários controlados usados em repositórios.

O levantamento e composição da amostra de repositórios brasileiros foi realizado por meio de pesquisa eletrônica na lista[5] do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) composta com um total de 107. Em seguida, foi realizada análise a partir de informações obtidas pelo conteúdo das páginas disponíveis na internet sobre suas principais características com a finalidade de observar diferenças e semelhanças, principais repercussões, características metodológicas, formais e de conteúdo.

A identificação dessas características selecionou uma amostra representativa para aplicação da análise de controle de vocabulário nos processos de representação e recuperação da informação. Antes, porém, foi realizada observação e análise das interfaces de pesquisa e navegação oferecidas pelos Repositórios aos usuários.

A análise de controle de vocabulário foi aplicada aos repositórios que utilizam vocabulário controlado e realizou os seguintes procedimentos:

- Identificação do tipo de vocabulário controlado mediante lista de sistemas de organização do conhecimento disponibilizada pela fundamentação teórica com base em Zeng (2008), Boccato (2011), Barité (2011);
- Verificação do controle de vocabulário a partir de simulações de buscas por assunto que propiciaram a observação do uso vocabulário controlado nos processos de representação e recuperação mediante comparativo dos termos utilizados nos campos de assuntos de metadados e vocabulário controlado.

Posteriormente à apresentação dos resultados foi realizada análise e discussão com base na fundamentação teórica e resultados da investigação sobre repositórios

[5] http://www.ibict.br/informacao-para-a-pesquisa/repositorios-digitais

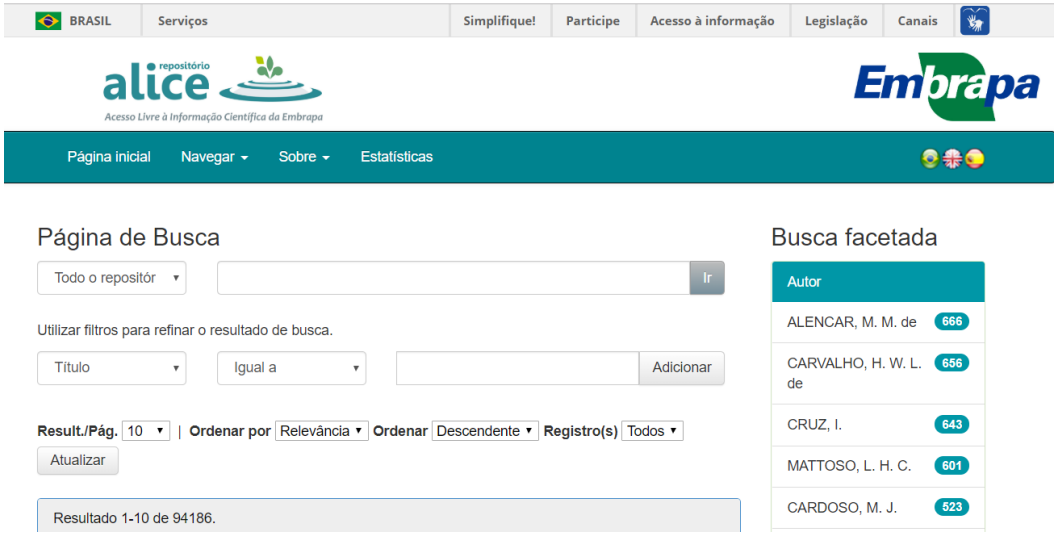
canadenses.

4 RESULTADOS

Após análise documental dos 107 repositórios presentes na lista do IBICT foram selecionados 86 em função de haver 13 repositórios com *websites* indisponíveis, 7 bibliotecas digitais diferentes de repositórios e um diretório de repositórios não selecionados. A pesquisa exploratória foi realizada em amostra de 86 repositórios brasileiros disponíveis na lista do IBICT. A observação e análise qualitativa concentrou-se nas modalidades de busca das interfaces de pesquisa e na incorporação dos tipos de vocabulários controlados na representação de assuntos e recuperação da informação.

Após localização dos diferentes repositórios brasileiros o próximo passo foi determinar se havia disponibilização de vocabulário controlado para facilitar a recuperação e o gerenciamento de informações e para isso analisamos, inicialmente, as modalidades de busca nas interfaces de pesquisa dos repositórios porque seus parâmetros incluem modalidades de busca orientadas por tipos de vocabulários controlados. A observação dos recursos de interface de busca revelou que todos os repositórios oferecem opções de busca simples e avançada, conforme exemplo da Figura 14.

Figura 14 - Opções de busca simples e avançada



Fonte: Disponível em: https://www.alice.cnptia.embrapa.br/alice/simple-search?query=

A busca simples é realizada em texto completo de todos os itens da coleção digital. Nas buscas avançadas é possível pesquisar de forma combinada e com lógica booleana por campos específicos dos metadados como autor, título, assunto, tipo de documento, instituição, data, publicador e etc. conforme exemplo da Figura 15:

Figura 15 - Busca avançada com pesquisa em campos específicos do metadados

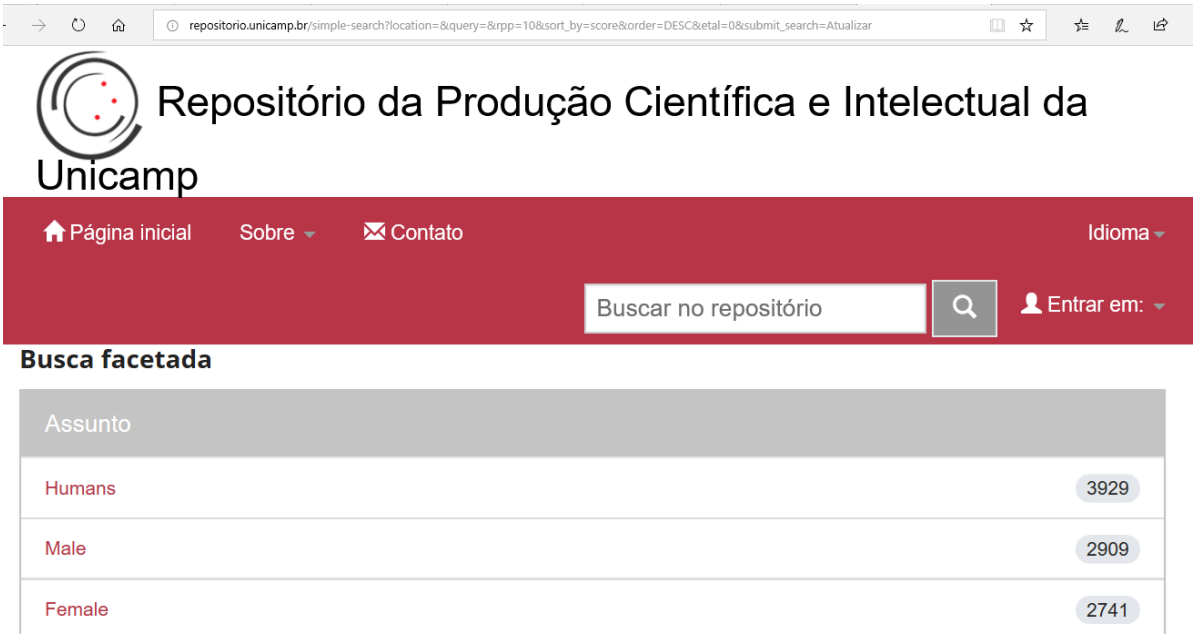
Fonte: Disponível em: www2.senado.leg.br/bdsf/

Algumas plataformas apresentam, além da interface de busca simples e avançada, a opção de uma busca denominada “facetada”, que aparece como uma lista de seleção de termos que representam os assuntos significativos em uma coluna da direita precedido pela quantidade de documentos classificados naquelas facetas de assunto ou, ainda, como uma lista de assuntos em ordem alfabética que leva diretamente aos documentos relacionados. Estão disponíveis, também, outras modalidades de buscas facetadas por listas alfabéticas de autor, título, instituição, ano de publicação, tipo de documento, áreas do conhecimento. O termo “busca facetada” não é estranho na área de Organização do Conhecimento, pelo contrário, é bem conhecido pelos estudos de Ranganathan e atualmente seus princípios são aplicados para: campos novos e emergentes; áreas interdisciplinares; vocabulários em que múltiplas hierarquias são necessárias; classificar documentos eletrônicos, objetos de conteúdo e também para metadados que representam objetos de conteúdo (AMERICAN NATIONAL STANDARD INSTITUTION, 2010).

Além disso, conforme a norma ANSI/NISO Z39.19-2005 (R2010), atributos podem ser selecionados como facetas para organizar objetos de conteúdo: tópico (assunto do conteúdo), formato do material, público alvo, tipo de conteúdo e, dessa forma,

> Sistemas de recuperação de informações podem usar facetas para representar pelo menos parte da estrutura do banco de dados subjacente. Ou seja, facetas podem ser definidas para corresponder a alguns dos atributos dos objetos de conteúdo. Em uma ferramenta impressa ou on-line, como um catálogo, por exemplo, as facetas podem aparecer como os termos genéricos usados para organizar as informações mais específicas (AMERICAN NATIONAL STANDARD INSTITUTION, 2010, p.15).

Figura 16 - Busca facetada em repositório

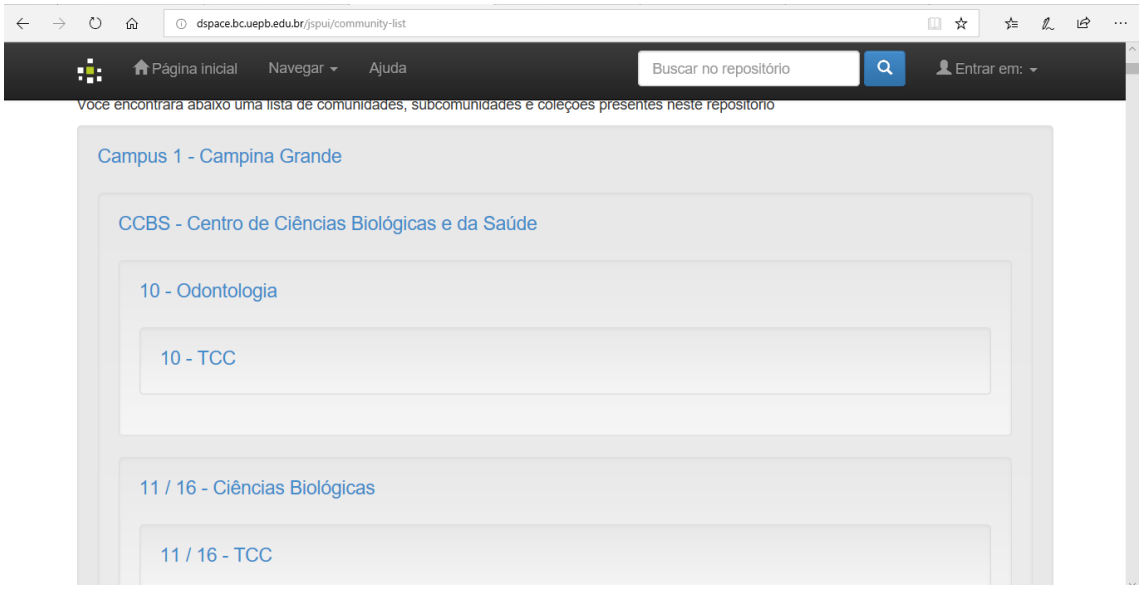


Fonte: Disponível em: repositorio.unicamp.br/simple-search?location=&query=&rpp=10&sort_by=score&order=DESC&etal=0&submit_search=Atualizar

Figura 17 - Busca em listas de assuntos em ordem alfabética

Em plataformas do DSPACE as coleções são, também, organizadas em comunidades e coleções que representam um sistema de classificação:

Figura 18 - Busca por comunidades e coleções em repositórios



Fonte: http://dspace.bc.uepb.edu.br/jspui/community-list

Os vocabulários controlados foram, de modo geral, incorporados à representação e recuperação quanto aos seguintes tipos:

Quadro 2 - Tipos de vocabulários controlados incorporados às modalidades de busca

TIPOS	QUANTIDADE DE REPOSITÓRIOS[6]
Tesauros (Thesagro, NAL Thesaurus, Tesauro do Senado Federal, VocaUSP, DeCS, Vocabulário WEBTHES)	6
Listas de termos em ordem alfabética	70
Listas de áreas do conhecimento ou listas classificadas	8
Taxonomias	7

Fonte: elaborado pelos autores com base na observação e análise dos repositórios

A análise demonstra que a incorporação desses tipos de vocabulários controlados na recuperação foi realizada de modo primitivo em alguns casos como, por exemplo, nas coleções e comunidades que podem ser vistas como uma classificação cujos assuntos foram atribuídos aos itens como, também, nas listas de termos em ordem alfabética e na classificação da informação disponível no *website*. Entretanto, ao analisar as listas de termos em ordem alfabética disponível para a navegação ou busca por assunto, supõe-se que os termos não são de vocabulários controlados, mas de palavras-chave de linguagem natural. Destacam-se alguns repositórios que utilizam o vocabulário controlado "Descritores em Ciências da Saúde" (DeCS) e que citam essa fonte para a disponibilização dos descritores nas listas alfabéticas como é o caso do repositório "Acervo de recursos educacionais em saúde - ARES" e, além disso, é possível averiguar que a representação dos termos na indexação dos itens é realizada com o mesmo vocabulário controlado conforme citação abaixo:

> O ARES adota o vocabulário de Descritores em Ciências da Saúde (DeCS). Seu uso possibilita a uniformidade dos termos de indexação e a recuperação da informação. O DeCS foi desenvolvido pela BIREME, a partir do Medical Subject Headings (MeSH), para servir como uma linguagem única para a indexação de conteúdo na área de Saúde (UNIVERSIDADE ABERTA DO SUS).[7]

A observação e análise das interfaces de buscas revelou que os repositórios utilizam mais de uma opção de busca ao mesmo tempo, simples e avançada com várias opções a serem combinadas e ainda agregam listas para buscas facetadas com diferentes opções como representado no Quadro 3:

[6] Existem repositórios que incorporaram mais de um tipo de vocabulário controlado

[7] Disponível em: https://ares.unasus.gov.br/acervo/browse?type=decs

Quadro 3 - Recuperação da informação na interface de busca e navegação

MODALIDADES DE BUSCA E NAVEGAÇÃO	QUANTIDADE DE REPOSITÓRIOS
Título	68
Resumo	56
Palavras-chave	61
Assunto	44
Palavras-chave de vocabulário controlado	3

Fonte: elaborado pelos autores com base na observação e análise dos repositórios

Portanto, 11 repositórios incorporam a lista classificada como mais um tipo de vocabulário controlado além das listas de termos alfabéticas simples, incorporadas por 85 repositórios. Em seis repositórios são utilizados tesauros cujos descritores são incorporados nas listas de termos em ordem alfabética e nos campos de assunto dos metadados. O que se nota é o pouco uso de vocabulários controlados com maior complexidade como tesauros.

Por outro lado, na representação para indexação, a observação dos metadados dos itens recuperados durante simulações de busca demonstra que os repositórios utilizam, em sua maioria, termos de vocabulário controlado e palavras-chave da linguagem natural no mesmo metadados conforme dados quantitativos do Quadro 5. Estes resultados revelam uma tendência de integração de vocabulários controlados e linguagem natural em repositórios comentada por Barité (2014) e verificada por Mondoux e Shiri (2009) que a consideram vantajosa tendo em vista as possibilidades de mútua atualização e enriquecimento de vocabulário.

Quadro 4 - A representação na indexação: vocabulários controlados e linguagem natural

INDEXAÇÃO EM METADADOS	QUANTIDADE DE RI
Por palavras-chave	23
Por termos do vocabulário controlado	7
Por palavras-chave e termos do vocabulário controlado	56

Fonte: elaborado pelos autores com base na observação e análise dos repositórios

Essa discussão sobre a integração é necessária tendo em vista o aprimoramento da recuperação para o usuário. Problemas simples de representação e recuperação ocasionado pela ortografia poderiam ser evitados com essa integração. Como exemplo disso, observa-se que na lista de assuntos da Figura 19 o termo “Macaco prego” aparece cinco vezes com diferentes grafias e itens informacionais diferentes são associados a cada palavra, o que obriga o usuário a ir e voltar na busca

para obter todos os documentos sobre “Macaco prego” que representa o mesmo assunto:

Figura 19 - Lista de assuntos com problema de controle de vocabulário na grafia dos termos.

< Anterior	Mostrando resultados 53366 a 53385 de 98263	Próximo >
Macacauba		1
Macaco Prego		2
Macaco prego		2
Macaco Prego.		4
Macaco-prego		4
Macaco-prego.		1
Macacos		1
Macacu		1
Macadâmia		2

Fonte: Disponível em: https://www.alice.cnptia.embrapa.br/alice/browse?type=subject&order=ASC&rpp=20&offset=53365

Os repositórios que disponibilizam termos somente em português são 27 e 59 em português e em inglês, o que representa um aproveitamento de *keywords* e possibilita maior visibilidade internacional. Nesses 59 repositórios existe a possibilidade de mudar a língua do *website*.

5 CONCLUSÕES

A investigação realizou análises das interfaces de busca e navegação de repositórios brasileiros em seus sistemas de recuperação da informação de suas coleções de recursos de informação e documentos armazenados. As interfaces de busca analisadas possuem recursos de navegação por várias modalidades de busca que permitem o acesso aos recursos e documentos armazenados. Foi possível confirmar que a busca por assuntos, na maioria dos repositórios, incorpora a lista de termos em ordem alfabética como um dos tipos de vocabulários controlados, no entanto e ao que parece, sem controle de vocabulário para conversão da linguagem natural de autores, indexadores e usuários e uniformização dos formatos dos termos para obter consistência entre representação e recuperação da informação. A

eliminação de ambiguidades é uma função básica das listas de termos e, assim, outras vantagens do controle de vocabulário poderiam acontecer.

Um dos principais objetivos do vocabulário controlado é servir como auxiliar de pesquisa na localização de objetos de conteúdo (AMERICAN NATIONAL INFORMATION STANDARD ORGANIZATION, 2010). As listas de termos são vocabulários controlados de menor complexidade que os tesauros e poderiam ser aprimoradas com a possibilidade de uso dos termos dos tesauros porque, a rigor, passaram pelo controle de vocabulário e poderiam servir com mais consistência na recuperação por meio das inferfaces de busca, assim como na representação em metadados.

A representação dos objetos de conteúdo em metadados observada revelou uma tendência de integração de termos de vocabulários controlados com palavras-chave da linguagem natural atribuídas pelo autor. Essa integração é importante e necessária para a atualização dos vocabulários porque as listas de termos utilizadas para buscas por assunto incorporam os termos e as palavras-chave. Além disso, o quantitativo de termos e palavras-chave atribuídos aos metadados ampliam a visibilidade da produção científica armazenada. Entretanto, seria necessário pensar na consistência entre termos e palavras-chave para obter o máximo de revocação na recuperação e evitar a dispersão.

A construção e atualização de vocabulários controlados têm, atualmente, estreita relação com a web semântica e o movimento de dados abertos. A Norma ISO 25964, Partes 1 (2011) e 2 (2013) que trata de tesauros e a interoperabilidade com outros vocabulários, tem relação com os dados abertos e web semântica. Essa relação, principalmente voltada para a construção de tesauros, é inovadora e precisa ser utilizada para que traga melhorias ao trabalho dos profissionais e mais acessibilidade à comunidade usuária.

Com a sistematização teórica e metodológica sobre controle de vocabulário e tipos de vocabulários controlados, entendemos ser imprescindível, do ponto de vista pedagógico, que os gestores e profissionais responsáveis por repositórios compreendam e valorizem os processos e sistemas de organização do conhecimento na representação e recuperação da informação de domínios de assuntos especializados. Nesse sentido, consideramos que a pesquisa fornecerá subsídios para que o vocabulário controlado seja articulado ao ambiente de repositórios

universitários para integrar-se à linguagem natural e cumprir seu melhor papel: o de controle de vocabulário sempre atualizado.

REFERÊNCIAS

AMERICAN NATIONAL STANDARDS INSTITUTE/ NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Z39.19-2005 (R2010). Bethesda, Maryland: NISO Press, 2010. Disponível em: https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Acesso em: 23 maio 2019.

ANDRADE, M. C. de, CERVANTES, B. M. N. A contribuição da organização do conhecimento para a interoperabilidade semântica: alternativas para repositórios institucionais. **Inf. Prof.***,* Londrina, v. 1, n. 1/2, p. 152 –170, jul./dez. 2012. Disponível em: http:www.uel.br/revistas/infoprof/.

BARITÉ, M. El control de vocabulario en la era digital: revisión conceptual. **Scire**, v.20, n.1, p.99-108, 2014. Disponível em: https://www.ibersid.eu/ojs/index.php/scire/article/view/4196 . Acesso em: 25 jun de 2019.

BARITÉ, M. Sistemas de organización del conocimiento: una tipología actualizada. **Inf. & Inf.**, Londrina, v.16, n.3, p.122-139, 2011. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/9952 . Acesso em: 3 maio 2019.

BARRIONUEVO ALMUZARA, L., ALVITE DÍEZ, M. L., RODRÍGUEZ BRAVO, B. A study of authority control in Spanish university repositories. **Knowledge Organization***,* 39 (2), pp.95-103, 2012. Disponível em: http://www.ergon-verlag.de/isko_ko/downloads/ko_39_2012_2_e.pdf.

BOCCATO, V.R.C. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. **InCID: R. Ci. Inf. e Doc***.,* Ribeirão Preto, v. 2, n. 1, p. 165-192, jan./jun. 2011.

BOCCATO, V. R. C.; FUJITA, M. S. L.; GIL LEIVA, I. Avaliação comparada do uso de linguagens de indexação em catálogos coletivos de bibliotecas universitárias para recuperação por assunto. ***Scire***, Zaragoza, v.17, n. 2, p. 55-64, jul-dic. 2011.

BODENREIDER, O. The unified medical language system (UMLS): integrating biomedical terminology. **Nucleic Acids Res**., v.32, n.1, p.267-70, 2004. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308795/ . Acesso em: 30 jun. 2019.

CAMPOS, M. L. de A.; GOMES, H. E. Taxonomia e classificação: o princípio de categorização. **DataGrama Zero**, v. 9, n. 4, p. 2008.

BRITISH STANDARDS INSTITUTION. **Structured vocabularies for Information retrieval:** BS 8723-1. [SI].:BSI, 2005.

FIGUEIREDO, S. P. **Taxonomia**. Cursos Miniweb, 2007 Disponível em: http://www.miniweb.com.br/atualidade/tecnologia/artigos/taxonomia.html. Acesso em: 26 de novembro de 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. Ed. São Paulo: Atlas, 2010. 184p.

HARPER, C. A., TILLETT, B. B. Library of Congress controlled vocabularies and their application to the Semantic Web. **Cataloging & Classification Quarterly**, v.43, n.3-4, p.47-68, 2009. Doi: 10.1300/J104v43n03_03.

HILL, L.; JANEÉ, G.; BUCHEL, O.; ZENG, M.L. Integration of knowledge organization systems into digital library architectures: Position Paper for 13th ASIS&T SIG/CR Workshop, "Reconceptualizing Classification Research". **Advances of classification reserch**, v.13, 2002. Disponível em: https://www.researchgate.net/publication/228689093_Integration_of_Knowledge_Organization_Systems_into_Digital_Library_Architectures_Position_Paper_for_13th_ASIST_SIGCR_Workshop_Reconceptualizing. Acesso em: 27 jun. 2019.

HJORLAND, B. What is Knowledge Organization (KO)? **Knowl. Org.,** v.35, n.2/3, p.86-101, 2008.

HODGE, G. **Systems of knowledge organization for digital libraries**: beyond traditional authorities files. Washington, DC: CLIR, 2000. Disponível em: http://www.clir.org/pubs/reports/pub91/contents.htm . Acesso em: 24 maio 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Information and documentation** - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. ISO 25964-1:2011. Geneva: International Organization for Standardization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Information and documentation** - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies. ISO 25964-1:2011. Geneva: International Organization for Standardization, 2013.

KURAMOTO, H. Ferramentas de software livre para bibliotecas digitais. *In*: MARCONDES, C.H.; KURAMOTO, H.; TOUTAIN, L.B.; SAYÃO, L.F. (Orgs.). **Bibliotecas digitais:** saberes e práticas. Salvador/Brasília: UFBA/IBICT, 2005. p. 293-309.

LAMBE, P. **Organising knowledge***:* taxonomies, knowledge and organizational effectivess. Oxford, England: Chandos, 2007.

LANCASTER, F. W. **Indexação e resumos**: teoría e prática. 2.ed. Trad. de Antônio Agenor de Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 2004. 452p. (Título original: Indexing and abstracting in theory and practice).

MAYR, P.; TUDHOPE, D.; CLARKE, S.D.; ZENG, M.L.; LIN, X. Recent applications of knowledge organization systems: introduction to a special issue. **International Journal of Digital Library**, v.17, n.1, p.1-4, 2016. DOI 10.1007/s00799-015-0167-x.

MONDOUX, J.; SHIRI, A. Institutional repositories in Canadian post-secondary institutions: user interface features and knowledge organization systems. **Aslib Proceedings: New Information Perspectives**, v.61, n.5, p. 436-458, 2009. Disponível em: https://doi.org/10.1108/00012530910989607.

MONTEIRO, F., BRASCHER, M. Organização da informação em repositórios temáticos: o uso da modelagem conceitual. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais** [...]. Salvador: UFBA/ANCIB, 2007. 12p. Disponível em: http://www.enancib.ppgci.ufba.br/artigos/GT2--261.pdf.

MOREIRA, M. P.; MOURA, M. A. Construindo tesauros a partir de tesauros existentes: a experiência do TCI: Tesauro em Ciência da Informação. **DataGramaZero**. v. 7, n. 4, 2006. Disponível em: http://www.brapci.inf.br/index.php/article/view/0000007757/27466a615ff97c8ea30886db7bc3b9f4 . Acesso em: 06 maio 2019.

RAMALHO, R. A. S. Representação do Conhecimento e ontologias: reflexões interdisciplinares. *In*: Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), 10, 2009, João Pessoa. **Anais** [...] João Pessoa: Ideia/Universitária, 2009. p. 714-727. Disponível em: http://enancib.ibict.br/index.php/enancib/xenancib/paper/viewFile/3175/2301. Acesso em: 3 jun. 2019.

SALES, R. de.; CAFÉ, L. Diferenças entre tesauros e ontologias. **Perspectivas em Ciência da Informação,** v. 14, n. 1, p. 99-116, jan./abr. 2009. Disponível: http://www.scielo.br/pdf/pci/v14n1/v14n1a08.pdf. Acesso em: 03 jul. 2019.

SALES, L. F.; MOTTA, D. F. da. **Base teórica da Ciência da Informação para construção de taxonomias consistentes**. 2010. Disponível em: http://www.casaruibarbosa.gov.br/dados/DOC/palestras/memo_info/mi_2010/FCRB_MI_Construcao_de_taxonomias_consistentes.pdf. Acesso em: 28 jun. 2019.

SANTOS, C. A. C. M. dos. **Classificação em bibliotecas***:* os esquemas, 2018. Disponível em: https://slideplayer.com.br/slide/12540959/. Acesso em: 28 jun. 2019.

SILVA, Y. de S. **Apostila auxiliar de biblioteca***.* 2015. Disponível em: https://www.ebah.com.br/content/ABAAAgkTMAJ/apostila-auxiliar-biblioteca. Acesso em: 28 jun. 2019.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a taxonomy of KOS: dimensions for classifying knowledge systems. **Knowledge Organization**, v.39, n.3, p.179-92, 2012.

TERRA, G. de M.; VASCONCELOS, T. R. Vocabulário controlado em repositórios de instituições de ensino superior brasileira: estudo comparativo. **RACIn***,* João Pessoa, v. 3, n. 1, p. 32-51, 2015. Disponível em: http://racin.arquivologiauepb.com.br/edicoes/v3_n1/racin_v3_n1_artigo03.pdf . Acesso em: 23 abr. 2019.

TUDHOPE, D.; KOCH, T. New applications of knowledge organization systems: introduction to a special issue. **Journal of Digital Information**, v.4, n.4, Article No. 286, 2004. Disponível em: https://journals.tdl.org/jodi/index.php/jodi/article/view/109/108. Acesso em: 30 maio 2019.

TUDHOPE, D.; KOCH, T.; HEERY, R. **Terminology services and technology**. JISC state of the art review. Bath, UK: UKLON, 2006. Disponível em: http://www.ukoln.ac.uk/terminology/ . Acesso em: 30 maio 2019.

UNIVERSIDADE ABERTA DO SUS. **Acervo de recursos educacionais em saúde** – ARES. Disponível em: https://ares.unasus.gov.br/acervo/browse?type=decs . Acesso em: 30 jun. 2019.

WEITZEL, S. da R. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. **Encontros Bibli**, v.24, n.54, p.105-23, 2019. DOI: 10.5007/1518-2924.2019v24n54p105

ZENG, M. L. Knowledge orgnization systems (KOS). **Knowledge Organization**, v.35, n.2/3, p.160-82, 2008. Disponível em: https://www.academia.edu/26672820/Knowledge_Organization_Systems_KOS_ . Acesso em: 06 maio 2019.

ZENG, M. L. **Construction of controlled vocabularies**: a primer *(based on Z39.19*). 2005. Disponível em: http://www.slis.kent.edu/~mzeng/Z3919/index.htm. Acesso em: 02 maio 2019.

ZENG, M. L.; CHAN, L. M. Trends and issues in establishing interoperability among knwoledge organization systems. **Journal of the American Society for Information Science and Technology**, v. 55, n. 5, p. 377-395, 2004. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/asi.10387. Acesso em: 30 maio 2019;

ZHANG, Y.; OGLETREE, A.; GREENBERG, J.; ROWELL, C. Controlled Vocabularies for Scientific Data: Users and Desired Functionalities. *In*: Proceedings of the 78th ASIS&T Annual Meeting,November 6-10, 2015. St. Louis: **ASIST**, v.52, 2015. Disponível em: https://www.asist.org/files/meetings/am15/proceedings/frontmatter/titlepage15.html

CONTROLLED VOCABULARIES IN THE REPRESENTATION AND RETRIEVAL OF INFORMATION IN BRAZILIAN REPOSITORIES

ABSTRACT
Introduction: The purpose of vocabulary control is consistency in representation and retrieval in order to avoid the dispersion of informational resources related to the terms of the controlled vocabulary. Digital repositories are currently of fundamental importance in the organization and management of knowledge whose information resources, besides stored and preserved, need to be accessed and retrieved without inconsistencies. For this, it is important to verify the incorporation of types of controlled vocabulary in university repositories in the processes of

representation and retrieval. **Objective**: This research proposes research on controlled vocabularies in repositories with the purpose of verifying search and navigation interface resources to identify the incorporation of types of controlled vocabularies in the representation and retrieval of information. **Methodology**: Documentary analysis, through electronic research, for the collection of repositories and exploratory research for the analysis of interface resources to identify types of controlled vocabularies used in representation and retrieval of information based on parameters obtained in the theoretical basis. **Results**: The analysis of the search interfaces of 86 Brazilian repositories revealed that controlled vocabularies were generally incorporated into representation and retrieval for the following types: thesauri (6 repositories), classified lists (8), taxonomies (7), and lists of terms in alphabetical order (70). Metadata includes natural language keywords and vocabulary terms controlled in 56 repositories. **Conclusions**: Lists of terms in alphabetical order used to search for subjects in navigation interfaces need to incorporate the vocabulary control function by integrating the keywords with the terms of the controlled vocabularies and obtaining consistency in the retrieval.

Descriptors: Vocabularies controlled. Institutional university repositories Vocabulary control. Representation of information.

VOCABULARIOS CONTROLADOS EN LA REPRESENTACIÓN Y RECUPERACIÓN DE INFORMACIÓN EN REPOSITORIOS BRASILEÑOS

Resumen:
Introducción: El propósito del control de vocabulario es la consistencia en la representación y recuperación para evitar la dispersión de recursos informativos relacionados con los términos del vocabulario controlado. Los repositorios digitales tienen actualmente una importancia fundamental en la organización y gestión del conocimiento, cuyos recursos de información, además de almacenados y conservados, deben ser accedidos y recuperados sin incoherencias. Para ello, es importante verificar la incorporación de tipos de vocabulario controlado en repositorios universitarios en los procesos de representación y recuperación. **Objetivo:** esta investigación propone investigaciones sobre vocabularios controlados en repositorios con el fin de verificar los recursos de la interfaz de búsqueda y navegación para identificar la incorporación de tipos de vocabularios controlados en la representación y recuperación de información. **Metodología**: análisis documental, a través de la investigación electrónica, para la recopilación de repositorios e investigación exploratoria para el análisis de los recursos de interfaz para identificar los tipos de vocabularios controlados utilizados en la representación y recuperación de información basada en parámetros obtenidos en la base teórica. **Resultados:** El análisis de las interfaces de búsqueda de 86 repositorios brasileños reveló que los vocabularios controlados generalmente se incorporaron en la representación y recuperación para los siguientes tipos: tesauros (6 repositorios), listas clasificadas (8), taxonomías (7) y Listas de términos en orden alfabético (70). Los metadatos incluyen palabras clave en lenguaje natural y términos de vocabulario controlados en 56 repositorios. **Conclusiones:** Las listas de términos en orden alfabético que se usan para buscar sujetos en las interfaces de navegación necesitan incorporar la función de control de vocabulario al integrar las palabras clave con los términos de los vocabularios controlados y obtener consistencia en la recuperación.

Descriptores: Vocabularios controlados. Instituciones universitarias repositorios de control de vocabulario. Representación de la información.