Avaliação de controle de vocabulários em repositórios institucionais: uma proposta de aplicação metodológica intrínseca e extrínseca

Lizandra de Souza Santos Alves[a], *Roberta Cristina Dal'Evedove Tartarotti*[b] *e Mariângela Spotti Lopes Fujita*[c]

1. Introdução

Com o crescimento dos recursos informacionais em formato digital e com a evolução das Tecnologias de Informação e Comunicação (TIC), os repositórios institucionais se tornaram ambientes importantes para guarda, preservação e disseminação da produção informacional científica. Fujita (2018) evidencia que os repositórios digitais em universidades são, atualmente, recursos de grande relevância na organização e gestão do conhecimento gerado pela produção científica, tecnológica, artística e administrativa.

Nestes sistemas de recuperação da informação, o uso de vocabulários controlados possibilita melhor precisão na representação e recuperação por assuntos. Essas informações são

a Graduanda em Biblioteconomia pela Universidade Estadual Paulista – UNESP. E-mail: lizandra17alves@gmail.com.

b Doutora em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Bibliotecária na Universidade Estadual de Campinas – UNICAMP. E-mail: roberta_tartarotti@yahoo.com.br. ORCID: https://orcid.org/0000-0002-0246-4971.

c Doutora em Ciências da Comunicação pela Universidade de São Paulo – USP. Professora na Universidade Estadual Paulista – UNESP. E-mail: mariangelaslf57@gmail.com. ORCID: https://orcid.org/0000-0002-8239-7114.

armazenadas e preservadas, porém, para que haja eficácia na representação e disseminação desses recursos informacionais, é preciso verificar se há a disponibilização de vocabulário controlado para o uso do profissional indexador e analisar como se realiza o controle desses vocabulários em repositórios institucionais, pois a falta de controle de vocabulário pode gerar vários tipos de inconsistências determinantes para a imprecisão e irrelevância na recuperação de informações em qualquer sistema de busca (FUJITA; TOLARE, 2019, p. 94).

O uso de vocabulários controlados em repositórios institucionais possibilita melhor precisão na representação e recuperação por assuntos. No entanto, verifica-se ausência na literatura de procedimentos metodológicos voltados para a avaliação do controle de vocabulários nestes sistemas de recuperação da informação. Nesse sentido, o objetivo da pesquisa foi realizar a avaliação do controle de vocabulários controlados em repositórios institucionais, contribuindo com a melhoria da representação e recuperação por assuntos em repositórios institucionais.

2. Metodologias para avaliação de vocabulário controlado em repositórios institucionais

Leite, Amaro, Batista e Costa (2012) destacam que os repositórios institucionais levam em conta reunir, armazenar, organizar, preservar, recuperar e, sobretudo, a ampla disseminação da informação científica produzida na instituição acadêmica. No que tange à relevância de se utilizar um

vocabulário controlado em repositórios institucionais, Terra e Vasconcelos (2015) destacam que o controle de vocabulário permitirá a identificação dos documentos, visando à recuperação da informação por parte dos usuários de forma mais precisa.

Dahlberg (2006) acentua que a Organização do Conhecimento é uma ciência que cuidará de organizar e sistematizar os conceitos, sempre se preocupando com os valores dos conteúdos dos objetos ou assuntos que estão sendo organizados. Partindo disso, criam-se os Sistemas de Organização do Conhecimento, que servirão de auxílio na hora de realizar a interpretação do objeto e sistematizar seus conceitos para a representação. Para realizar essa representação de forma consistente, o uso de vocabulários controlados tem imensurável relevância para que esse processo seja realizado de forma padronizada conforme o domínio.

Para Harpring (2016), vocabulário controlado é um arranjo organizado de palavras e frases usadas para indexar e/ou recuperar conteúdos. Promovem consistência em termos preferidos e a atribuição dos mesmos termos a um conteúdo semelhante, tendo suas funções mais importantes de agrupamento de termos variantes e sinônimos em conceitos e a ligação dos conceitos em uma ordem lógica ou sua classificação em categorias. Kobashi (2008) salienta que o vocabulário controlado é uma linguagem artificial construída de termos que representarão um ou vários campos do conhecimento.

No entanto, para que haja consistência na representação e recuperação pelo uso desses vocabulários, é necessário realizar constante atualização e também aplicar metodologias avaliativas direcionadas tanto para a estrutura do vocabulário, quanto para a análise da precisão em relação à representação e recuperação. O vocabulário controlado é essencial para o momento da representação e recuperação de um documento, facilitando a comunicação entre a linguagem do indexador e dos usuários que realizam a busca. Todavia, para esse fim é importante a aplicação de avaliações dos vocabulários para que se tenha maior controle da padronização dos termos e assim os objetivos sejam alcançados de forma consistente. Sua avaliação é indispensável, pois, como um vocabulário controlado é composto por termos e o conhecimento está sempre em movimento, está sujeito a ficar desatualizado.

Essas avaliações precisam ser aplicadas visando-se o vocabulário controlado propriamente dito, buscando verificar se os elementos de sua estrutura estão cumprindo a função de auxiliar no momento da representação de maneira consistente; e avaliar como está sendo o seu impacto perante a recuperação da informação. Em vista disso, ao longo dos anos, mesmo que em pouca proporção, na literatura são apresentadas metodologias voltadas à avaliação de vocabulários controlados, podendo ser citados os estudos de Lancaster (2002), a Norma ANSI/NISO Z39.19 (2005), Souza e Campos (2007), Gil Leiva (2008) e Alves

(2021). Tomando-se como base estes estudos, foram elencados os procedimentos metodológicos, descritos a seguir.

3. Procedimentos metodológicos

Os procedimentos metodológicos foram realizados em duas etapas: (i) levantamento bibliográfico sobre a temática de pesquisa e; (ii) exploratório-explicativa com estudo de caso, sendo aplicados critérios avaliativos em seis repositórios institucionais que integram em seu sistema de vocabulários controlados estruturados, como os tesauros, em duas categorias de análise: avaliação intrínseca, contendo critérios voltados para o repositório propriamente dito e a avaliação extrínseca, com critérios voltados para a análise da representação e da recuperação da informação nesses repositórios.

Gil (2002, p. 41) salienta que a pesquisa exploratória proporciona maior familiaridade do pesquisador com o problema, propiciando ideias ou descobertas de intuições, sendo bastante flexível no que tange à possibilidade de consideração de diversos aspectos relativos ao fato estudado. Para a coleta de dados na primeira etapa, foi delineada a técnica de levantamento bibliográfico que tem por base os materiais já elaborados (GIL, 2008, p. 50). A revisão de literatura foi realizada a partir do corpus informacional advindo do levantamento bibliográfico em fontes bibliográficas nacionais e internacionais. Esse levantamento se deu a partir de livros, artigos, trabalhos apresentados em eventos de Ciência da Informação, teses e

dissertações em repositórios institucionais, bases de dados, periódicos e Internet.

A segunda etapa, exploratório-explicativa com estudo de caso, continuou a investigar o tema de pesquisa buscando identificar fatores que contribuam para a ocorrência de fenômenos. Na presente pesquisa, o controle de vocabulário em repositórios institucionais é considerado o fenômeno com o qual a pesquisa explicativa aprofunda o conhecimento da realidade (GIL, 2002, p. 42). O estudo de caso é caracterizado pelo estudo de um ou de poucos objetos, de maneira a permitir o seu conhecimento amplo e detalhado (GIL, 2008, p. 57-58).

A seleção dos repositórios institucionais que utilizam vocabulários controlados para a sua representação e recuperação da informação pautou-se em investigação de Fujita e Tolare (2019) voltada para a identificação de repositórios institucionais que empregam vocabulários controlados em seus sistemas e dos tipos e das principais características dos vocabulários controlados utilizados. Para o levantamento dos repositórios institucionais, as autoras realizaram pesquisa eletrônica na lista do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) composta com um total de 107 repositórios. Após o levantamento desses 107 repositórios, as autoras realizaram análise documental de cada um deles e a partir disso "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" (FUJITA; TOLARE, 2019, p. 113).

Posteriormente à localização dos diferentes repositórios, as referidas autoras passaram para o processo de verificar a disponibilização de vocabulário controlado nesses repositórios. Os vocabulários controlados foram empregados à representação e recuperação a partir dos seguintes tipos: tesauros em 6 repositórios; listas de termos em ordem alfabética (utilizados em 70 repositórios); listas de áreas do conhecimento ou listas classificadas utilizadas por 8; e taxonomias utilizadas por 7 repositórios. Para a presente pesquisa, utilizou-se do resultado dos 6 repositórios institucionais que utilizam 5 diferentes tesauros: DeCS, Thesagro, NAL Thesaurus, VocaUSP e Tesauro do Senado Federal. No que compete aos repositórios institucionais utilizados como universo de pesquisa, cada repositório é referenciado pela letra "R" e um número, da seguinte maneira: R1: DeCS; R2: Thesagro e NAL Thesaurus; R3: VocaUSP; R4: Tesauro do Senado Federal; R5: Tesauro do Senado Federal; R6: Thesagro e NAL Thesaurus.

Para que fosse realizada a aplicação da metodologia de avaliação de controle de vocabulário em repositórios institucionais, foram separadas duas categorias de análise contendo critérios avaliativos, cada uma delas, para que fossem aplicados nos repositórios institucionais selecionados.

Gil Leiva (2008, p. 213) aponta duas metodologias de avaliação: uma tomando como objeto de análise o próprio

vocabulário controlado, chamada de avaliação intrínseca, e a outra que estuda o papel do vocabulário controlado na representação e recuperação de informações, denominada de avaliação extrínseca. A partir dessas duas metodologias propostas pelo referido autor, foram criadas as duas categorias de análise e critérios avaliativos que vão compor essas categorias, tendo como embasamento os pontos de metodologias apresentadas na literatura. Cabe ressaltar que os critérios foram adaptados para serem aplicados nos repositórios institucionais.

3.1 Categoria de análise com enfoque na avaliação intrínseca

Considerando-se que a avaliação intrínseca estuda o interior de algo, essa categoria trará critérios de avaliação voltados para o repositório institucional propriamente dito. Nesta categoria, foram delineados os seguintes critérios avaliativos:

a) Disponibilidade de link para o vocabulário controlado: corresponde ao fornecimento pelo repositório institucional, em sua interface, de um link que direcione o usuário prontamente para o vocabulário controlado (ALVES, 2021).
b) Vocabulário correspondente ao domínio: se o vocabulário controlado empregado confere ao domínio contemplado pelo repositório institucional (ALVES, 2021).
c) Autoarquivamento: Disponibilização do autoarquivamento. Em caso positivo, analisar se é utilizada a linguagem natural ou se há o direcionamento

para o vocabulário controlado para realizar a escolha de termos advindos da linguagem padronizada correspondente ao recurso informacional. Explorar no repositório institucional se está disponível algum manual de autoarquivamento (ALVES, 2021).

3.2 Categoria de análise com enfoque na avaliação extrínseca

A avaliação extrínseca volta-se para o impacto do vocabulário controlado nos sistemas de informação que utilizam a linguagem padronizada para realizar o processo de indexação e auxiliar na recuperação. Os critérios avaliativos serão pautados exclusivamente à recuperação da informação nos repositórios institucionais:

a) Controle terminológico: voltado para a análise de imprecisões advindas de ambiguidade, sinonímia, polissemia e homonímia na recuperação dos recursos informacionais.
b) Forma dos termos: corresponde à forma com que os termos foram apresentados na recuperação. Se aparecem em plural ou singular, se apresenta entrada direta ou inversa, o idioma adotado, se os termos são simples ou compostos e se são adotadas abreviaturas (SOUZA; CAMPOS, 2007).
c) Grau de especificidade: análise do nível de especificidade terminológica no momento da recuperação dos recursos informacionais, isto é, verificar se a especificidade da

recuperação está de acordo com os usuários potenciais (GIL LEIVA, 2008).

4. Resultados e discussões

Os resultados e discussão foram pautados nas categorias de análise avaliação intrínseca (que contem critérios voltados para o repositório propriamente dito) e avaliação extrínseca (cujos critérios foram voltados para a análise da representação e da recuperação da informação) nos repositórios institucionais, conforme apresentados nos procedimentos metodológicos, obtendo-se os seguintes resultados:

4.1 Avaliação intrínseca

a) Disponibilidade de link para o vocabulário controlado

Nesse critério a busca pelo resultado foi feita por toda a interface do repositório institucional. Dos seis repositórios analisados, apenas o R1 fornece acesso direto ao vocabulário utilizado. Foi possível destacar a falta do direcionamento ao vocabulário controlado integrado ao sistema para conhecimento dos usuários e também para uma possível pesquisa por algum pesquisador interessado, sem necessidade de contato com os gestores.

b) Vocabulário correspondente ao domínio

Os resultados do segundo critério demonstraram ampla integração de vocabulários controlados correspondentes ao domínio que o repositório institucional é pautado. Esse resultado

é relevante, pois quanto mais especializado o vocabulário controlado for em relação ao domínio de especialidade, maior será a precisão no momento da representação e recuperação por assuntos dos recursos informacionais que compõem a coleção.

c) Autoarquivamento

Em relação ao autoarquivamento da produção acadêmica pelos autores, o resultado foi parecido com o resultado do primeiro critério, pois apenas o R1 possibilita aos autores realizarem o autoarquivamento de seus próprios materiais. Porém, para que seja realizado o autoarquivamento, é necessária autorização. Não foi possível o acesso o manual que instrui sobre o preenchimento dos dados, não podendo ser constatada se a representação dos recursos informacionais é realizada via linguagem natural ou se os autores são direcionados para o vocabulário controlado integrado ao repositório institucional.

4.2 Avaliação extrínseca

a) Controle terminológico

Referente ao controle terminológico, verificou-se o correto controle de quatro repositórios, sendo eles: R2, R3, R4 e R6. Os outros dois repositórios apresentaram variações terminológicas em relação ao termo de busca. O R1 teve "hipertensão" como o termo escolhido e obteve quatro resultados por "hipertensão" e seis por "hipertensos". Já no R5, o termo escolhido foi "psicoterapia" e no momento da busca houve a recuperação de 7 recursos informacionais pelo termo escolhido, 2 recursos

informacionais não tinham nenhuma relação com o termo e um estava descrito como "psicologia" e "psicólogos".

b) Forma dos termos

Quanto à forma como os termos são apresentados, foi verificado se o idioma do termo de busca apresentava singularidade ou pluralidade, se os recursos informacionais eram recuperados sozinhos ou se apresentavam alguma composição ou abreviações do termo. No que tange à recuperação com o idioma proposto, todos recuperaram recursos informacionais em português sem necessidade de busca avançada. Os repositórios R1 e R3 apresentaram singularidade e pluralidade. Os demais apenas recuperaram materiais na singularidade do termo. Em relação à composição do termo escolhido com outros termos, o R1 só apresentou composição ao termo quando o mesmo estava no singular e os demais apresentaram composição independentemente de singular ou plural e nenhum dos repositórios institucionais apresentaram abreviações na recuperação do termo.

c) Grau de especificidade

Quanto à verificação se recursos informacionais recuperados abrangiam a especificidade em relação aos usuários potenciais do sistema de recuperação, todos os repositórios institucionais atenderam ao grau de especificidade na recuperação dos recursos informacionais.

De forma geral, com a aplicação dos critérios avaliativos nos repositórios institucionais, observou-se que o repositório classificado como R1 apresenta melhor representação e recuperação da informação, mesmo que todos os outros tenham apresentado resultados muito semelhantes.

5. Considerações finais

Os vocabulários controlados são uma linguagem em construção com estrutura relacional, que tem como função recuperar recursos informacionais e controlar a linguagem natural. Além de serem úteis para o momento da recuperação por assuntos, os vocabulários controlados são utilizados para a realização da representação do recurso informacional por meio do processo de indexação para melhorar a comunicação entre sistema e usuário. Controlando essa comunicação, será possível um maior nível de precisão na recuperação, diminuindo ruídos na recuperação dos recursos informacionais.

Para que se tenha consistência na representação e na recuperação por assuntos dos recursos informacionais são indispensáveis a atualização e a avaliação do vocabulário controlado, tendo em vista que é um instrumento dinâmico e que precisa estar em consonância com o conhecimento científico que vem se modificando e se atualizando constantemente. Neste contexto, é necessária a avaliação e atualizações constantes do vocabulário controlado, por meio de uma metodologia de avaliação de controle de vocabulário nos repositórios

institucionais para que a representação e a recuperação dos recursos informacionais sejam consistentes.

A aplicação de critérios avaliativos nos repositórios institucionais analisados corrobora a importância da utilização de um vocabulário controlado na representação dos recursos informacionais para melhorar a precisão na recuperação por assuntos no momento da busca pelos usuários e da integração ao repositório institucional de um vocabulário controlado que integre termos de seu domínio. Quanto mais especializado o vocabulário controlado, maior a precisão na representação e recuperação por assuntos dos recursos informacionais.

Referências

BRASIL. **Constituição da República Federativa do Brasil**. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 05 fev. 2021.

ANSI/NISO. Z39.19 – 2005. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda: NISO Press, 2005. 184 p.

ALVES, L. de S. S. **Avaliação de controle de vocabulários em repositórios institucionais**: uma proposta de aplicação metodológica intrínseca e extrínseca. 2021. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) - Faculdade de Filosofia e Ciências, Unesp, Marília, 2021.

DAHLBERG, I. Knowledge Organization: A New Science? **Knowledge Organization**, v.33, n.1, p. 11-19, 2006.

FUJITA, M. S. L. **Controle de vocabulário na representação e recuperação da informação em repositórios universitários.** 2017. Projeto de Pesquisa (Bolsa de Produtividade em Pesquisa) - CNPq, Programa de Pós-Graduação em Ciência da Informação, Unesp, Marília, 2018.

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, Londrina, v. 24, n. 2, p. 93-125, 2019.

GIL, A. C. **Como elaborar projetos de pesquisa.** 4. ed. São Paulo: Editora Atlas, 2002.

GIL, A. C. **Métodos e técnicas de pesquisa social.** 6. ed. São Paulo: Editora Atlas, 2008.

GIL LEIVA, I. **Manual de indización:** teoría y práctica. Gijón: Trea, 2008.

HARPRING, P. **Introdução aos vocabulários controlados**: Terminologia para arte, arquitetura e outras obras culturais. São Paulo: Secretaria da Cultura do Estado de São Paulo, Pinacoteca de São Paulo, ACAM Portinari, 2016.

KOBASHI, N. Y. **Vocabulário controlado**: estrutura e utilização. Brasília, Escola Nacional de Administração Pública, 2008.

LANCASTER, F. W. **El control del vocabulário em la recuperación de información.** 2. ed. Espanha: Universitat de València, 2002.

LEITE, F. C. L.; AMARO, B.; BATISTA, T.; COSTA, M. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: Ibict, 2012.

SOUZA, J. C. C. E.; CAMPOS, M. L. A. Avaliação de linguagem de indexação aplicada à informação jornalística: estudo de caso. *In:* **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 18, 2007, Salvador. **Anais** […]. Salvador: 2007. p. 1-17.

TERRA, G. 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, jan./jun. 2015.