

AVALIAÇÃO DO USO DE VOCABULÁRIO CONTROLADO EM REPOSITÓRIOS INSTITUCIONAIS

EVALUATION OF CONTROLLED VOCABULARY USE IN INSTITUTIONAL REPOSITORIES

Lizandra de Souza Santos Alves^a

Roberta Cristina Dal'Evedove Tartarotti^b

Mariângela Spotti Lopes Fujita^c

RESUMO

Objetivo: Elaborar e aplicar metodologia de avaliação do uso de vocabulários controlados para o controle de vocabulário em repositórios institucionais. **Metodologia:** Os procedimentos metodológicos foram realizados em duas etapas: a) análise da literatura sobre avaliação do uso de vocabulário controlado para o controle de vocabulário em repositórios institucionais e; b) aplicação exploratório-explicativa com estudos de caso de seis repositórios institucionais que utilizam vocabulário controlado e em um repositório institucional que não integra o vocabulário controlado para a indexação e recuperação por assuntos. A segunda etapa consistiu na aplicação de critérios avaliativos 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. **Resultados:** Na maioria dos repositórios institucionais analisada que integram vocabulário controlado, a recuperação por assuntos apresenta boa precisão a partir de termos buscados, enquanto o repositório que não integra o vocabulário controlado necessita de maior controle terminológico dos termos correlacionados às pesquisas de teses e dissertações para maior precisão na recuperação dos recursos informacionais. **Conclusões:** Quanto mais especializado for o vocabulário controlado em relação ao domínio do repositório institucional, mais eficaz será a recuperação por assuntos. A falta de padronização na representação acarreta falhas na recuperação com inconsistências na interação do repositório institucional com seus usuários.

Descritores: Repositórios institucionais. Vocabulários controlados. Avaliação da

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

^b Doutora em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: roberta_tartarotti@yahoo.com.br

^c Doutora em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Docente no Programa de Pós-graduação em Ciência da Informação na Universidade Estadual Paulista – PPGCI/UNESP. E-mail: mariangelasf57@gmail.com

indexação. Recuperação por assuntos.

1 INTRODUÇÃO

A representação e disseminação de recursos informacionais é resultado de tratamento descritivo e temático a partir de processos e sistemas de organização e representação da informação. O vocabulário controlado é utilizado para a padronização terminológica na representação temática pelo indexador no momento da atribuição de descritores que representem o assunto do recurso informacional.

Os termos de um vocabulário controlado utilizados para descrever e recuperar os recursos informacionais são unidades terminológicas que realizam o controle de vocabulário para evitar variações terminológicas determinantes de problemas na recuperação da informação, tais como diferentes grafias de um mesmo termo, singular e plural, homonímia, polissemia ou palavras vazias, além de termos não específicos ou muito genéricos.

Portanto, o controle de vocabulário é um recurso essencial para a representação de assuntos na indexação e na estratégia de busca para a recuperação de informações. Pensando nisso, os vocabulários controlados foram criados para auxiliar na organização e representação de itens bibliográficos do acervo das bibliotecas, para que tivesse maior precisão no momento da recuperação da informação que o usuário estivesse buscando.

Em repositórios institucionais, a inserção e a representação dos recursos informacionais podem ser realizadas tanto pelo profissional da biblioteca responsável pela atividade quanto pelos autores, trazendo autonomia para essas atividades ao propiciar o processo de autoarquivamento, antes realizado somente pelo bibliotecário em catálogos de bibliotecas. No autoarquivamento de teses e dissertações ou demais publicações o próprio autor insere o recurso informacional no sistema e realiza o preenchimento de metadados descritivos e temáticos. Dessa forma, ele mesmo atribui palavras-chave para a representação temática do conteúdo de sua publicação visando sua representação e recuperação por assuntos. Entretanto, considerando-se que o uso da linguagem

natural na representação pode causar inconsistências na recuperação, a utilização de um vocabulário controlado pelo profissional indexador ou pelo usuário é de suma importância.

O uso de vocabulários controlados em repositórios institucionais possibilita melhor precisão na representação temática e recuperação por assuntos e evita as variações terminológicas que causam dispersão e impedem a identificação de vários conteúdos por um mesmo termo. Em investigação sobre vocabulários controlados em repositórios institucionais, Fujita e Tolare (2019) identificaram vários tipos de vocabulários controlados, tesouros, listas classificadas, taxonomias, mas a maioria utiliza lista de palavras em ordem alfabética provenientes de linguagem natural sem controle de vocabulário. 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.

Tal resultado revela que existe um problema a ser investigado quanto à necessidade de analisar o uso de vocabulários controlados para o controle de vocabulário em repositórios institucionais. Nesse sentido, o objetivo da pesquisa foi elaborar e aplicar metodologia de avaliação do uso de vocabulários controlados para o controle de vocabulário em repositórios institucionais, visando contribuir com a melhoria da representação e recuperação por assuntos nestes sistemas de recuperação da informação e com elementos de avaliação da indexação no campo de pesquisa teórico e prático da Organização e Representação do Conhecimento.

2 AVALIAÇÃO DE VOCABULÁRIO CONTROLADO E CONTROLE DE VOCABULÁRIO EM REPOSITÓRIOS

O vocabulário controlado é um instrumento dinâmico que precisa estar em consonância com o conhecimento científico, que vem se modificando e se atualizando constantemente, notadamente no âmbito acadêmico. É uma linguagem artificial construída de termos que irão representar um ou vários campos do conhecimento (KOBASHI, 2008). A fim de garantir a precisão e a eficácia na comunicação entre os usuários e os sistemas de informação,

[...] um vocabulário controlado surge para tentar solucionar problemas causados por características próprias da linguagem natural como ambiguidades, polissemia, sinonímia e homonímia, que quando não padronizados afetam os resultados dos processos de recuperação da informação quanto à precisão e relevância (SILVA, 2015, p. 21).

Segundo Lancaster (2004, p. 19), “[...] um vocabulário controlado é essencialmente uma lista de termos autorizados” que irá realizar o controle de “sinônimos, diferenciar homógrafos e estabelecer as relações hierárquicas e não-hierárquicas”. Para Harpring (2016), vocabulário controlado é um arranjo organizado de palavras e frases usadas para indexar e/ou recuperar conteúdo. Promove consistência em termos preferidos e a atribuição dos mesmos termos a um conteúdo semelhante. Suas funções mais importantes são de agrupamento de termos variantes e sinônimos em conceitos e a ligação desses conceitos em uma ordem lógica ou classificação em categorias.

A norma ANSI/NISO Z39.19:2005 aponta que existem três relacionamentos utilizados em vocabulários controlados, cujas relações entre os termos são indicadas por vinculação semântica. Por sua vez, a ligação semântica abrange várias técnicas e convenções para indicar as relações entre os termos (ANSI/NISO Z39.19:2005, p. 42). Esses relacionamentos se dividem em: a) equivalência; b) hierarquia e c) associação.

Tradicionalmente, os vocabulários controlados foram criados a fim de padronizar a linguagem de indexação de recursos informacionais impressos em bibliotecas. Na contemporaneidade, com o advento dos sistemas de recuperação da informação como os repositórios institucionais, os vocabulários controlados possibilitam esta mesma função para os recursos informacionais digitais de forma mais dinâmica. No contexto acadêmico, os repositórios institucionais visam reunir, armazenar, organizar, preservar, recuperar e, sobretudo, a ampla disseminação da informação científica produzida na instituição acadêmica (LEITE *et al.*, 2012). Nestes sistemas de recuperação da informação, o vocabulário controlado adotado deve estar em consonância com a política de indexação definida pela instituição,

[...] intermediando o acesso à recuperação da informação e este, por outro lado, deve considerar a instituição onde se desenvolve; as expectativas e necessidades do usuário; as características do assunto tratado; os recursos humanos, físicos e financeiros; os

produtos e serviços visados e a relação custo/desempenho (BOCCATO; FUJITA, 2006, p. 268).

Considerando-se a relevância do uso de vocabulário controlado em repositórios institucionais, Terra e Vasconcelos (2015) destacam que o controle de vocabulário irá permitir a identificação dos recursos informacionais, visando à recuperação da informação por parte dos usuários de forma mais precisa. Isto é, para realizar essa representação de forma consistente, o uso de vocabulários controlados possibilita que esse processo seja realizado de forma padronizada, conforme o domínio.

No entanto, para que este nível de consistência na representação e recuperação por assuntos seja alcançado pelo uso desses vocabulários, é necessário realizar constante atualização e aplicar metodologias de avaliação em três principais aspectos: a) voltadas para a verificação do uso ou não de vocabulário controlado no repositório institucional (*avaliação de controle de vocabulário controlado*); b) direcionadas para a estrutura do vocabulário, isto é, se os elementos de sua estrutura estão cumprindo a função de auxiliar no momento da representação de maneira consistente; ou c) para a análise da precisão em relação à representação e recuperação por assuntos (*avaliação de vocabulário controlado*), isto é, verificar como está sendo o seu impacto na recuperação da informação.

A aplicação de avaliações nos vocabulários controlados utilizados pelos repositórios institucionais possibilita maior controle da padronização dos termos dos recursos informacionais, considerando-se que no âmbito acadêmico um vocabulário controlado é composto por termos advindos do conhecimento científico, que está sempre em movimento. Ou seja, a atualização e a avaliação do vocabulário controlado em repositórios institucionais visam uma maior consistência na representação e na recuperação por assuntos dos recursos informacionais. Porém, não apenas o vocabulário propriamente dito precisa ser avaliado, mas também o seu uso nos repositórios institucionais, pois alguns repositórios utilizam vocabulários controlados advindos de outras instituições e não vocabulários próprios que podem ter controle terminológico direto pelos profissionais da instituição responsáveis pela sua atualização.

A seguir, são apresentadas algumas metodologias de avaliação do uso

de vocabulário controlado em repositórios institucionais.

2.1 METODOLOGIAS DE AVALIAÇÃO DO USO DE VOCABULÁRIO CONTROLADO EM REPOSITÓRIOS INSTITUCIONAIS

Na Ciência da Informação, existem distintos tipos de avaliação, incluindo “[...] métodos com abordagens objetivas, subjetivas, automáticas ou manuais. Podendo enfatizar usabilidade, estrutura, relevância e a qualidade geral de um sistema de organização” No entanto, embora a literatura aponte alguns métodos de avaliação, sua aplicação ainda é incipiente na prática profissional de bibliotecários que gerenciam sistemas de organização e recuperação do conhecimento (SOUSA, 2019, p. 72).

Como metodologias de avaliação de vocabulários controlados podem ser citados os estudos de Lancaster (2002), a Norma ANSI/NISO Z39.19 (2005), Souza e Campos (2007) e Gil Leiva (2008).

A avaliação de um serviço de informação pode ser subjetiva ou objetiva. Estudos subjetivos, baseados em opiniões, não deixam de ser úteis, pois é importante saber o que as pessoas sentem em relação ao serviço. Mas a avaliação terá sua utilidade máxima se for analítica e diagnóstica, procurando descobrir como o serviço poderia melhorar, e é difícil basear este tipo de estudo apenas em opiniões. Em geral, portanto, devem-se adotar critérios e procedimentos objetivos (LANCASTER, 2004, p. 16).

Lancaster (2002) apresenta um ponto histórico sobre a avaliação, destacando que testes estatísticos foram aplicados a tesouros por meio de uma razão de relacionamento e uma medida de acessibilidade, onde:

A razão de relacionamento é a proporção de termos com referência cruzada (ou seja, termos unidos por outro termo, por exemplo, com TG, TE ou TR) e o número total de termos no vocabulário. A medida de acessibilidade é a média de referências aos descritores do vocabulário. Por exemplo, uma medida de acessibilidade de 2,923 indica que cada termo de um vocabulário tem uma média de encaminhamento de aproximadamente três outros termos. Essas medidas indicam a amplitude das uniões (ou seja, referências cruzadas) entre os termos de um vocabulário (LANCASTER, 2002, p. 173).

A Norma ANSI/NISO Z39.19 (2005, p. 94-95) indica duas principais razões para testar e avaliar vocabulários controlados: a) para determinar se o

vocabulário controlado usado para descrever objetos de conteúdo fornece resultados de pesquisa adequados (ou seja, alta relevância e *recall*) para a maioria dos usuários, e; b) para estabelecer se o vocabulário controlado corresponde às expectativas dos usuários quanto aos termos integrados. Por exemplo, se o vocabulário controlado não for suficientemente rico, os usuários poderão não descobrir os termos desejados. A Norma apresenta os seguintes tipos de avaliação que podem ser considerados quando confrontados com o desafio de medir a qualidade e/ou a eficácia de um vocabulário controlado:

- **Avaliação heurística:** realizada por um ou vários especialistas. Pode ser informal e qualitativa (por exemplo, cada especialista compartilhando suas reações e opiniões) ou formal e quantitativa (por exemplo, todos os especialistas atribuindo pontuações de acordo com uma lista compartilhada de critérios);

- **Modelagem de afinidade:** uma amostra representativa dos usuários é solicitada a classificar uma coleção de cartões, com termos do vocabulário em grupos. Os resultados são analisados em relação à hierarquia existente de termos. Os usuários também podem ser convidados a pontuar termos equivalentes e relacionados de acordo com o nível de similaridade;

- **Teste de usabilidade:** a avaliação holística do sistema de informação pode fornecer informações sobre a eficácia do vocabulário controlado.

Souza e Campos (2007, p. 7-8) apresentam uma metodologia como um instrumento para avaliação de linguagens baseado nos seguintes aspectos:

- **Forma de apresentação:** análise da origem da linguagem, se é adaptação de outro instrumento, se apresenta introdução explicativa com informações como histórico para sua contextualização; que domínio abrange; objetivos; condições de atualização e uso; número de termos; quais os fundamentos teóricos que nortearam sua elaboração; forma de apresentação (se possui parte alfabética e/ou parte sistemática); aspectos tipográficos e estéticos se apresentam notação e fonte diferenciada para destacar o termo geral;
- **Campo de abrangência:** análise da(s) área(s) de conhecimento compreendida(s) pela linguagem de indexação, se os termos da linguagem indicam período histórico e se os assuntos são compostos por identificadores geográficos e cronológicos;
- **Estrutura dos termos:** análise do relacionamento entre os termos, ou seja, a “estrutura do vocabulário”. Lancaster (2004) usa a expressão “nível de coordenação dos termos” para designar termos constituídos por frases ou expressões;

- **Forma das palavras:** abordagem da unidade linguística adotada no vocabulário. Na revisão de literatura há trabalhos que analisam se a linguagem emprega plural ou singular, se apresenta entrada direta ou inversa, o idioma adotado, se as palavras são compostas ou simples, se existem variações ortográficas e se são adotadas abreviaturas;
- **Estrutura semântica:** análise do vocabulário sob o ponto de vista conceitual, as relações entre os termos, a articulação dos assuntos e se adota notas de aplicação para excluir as ambiguidades;
- **Relações de equivalência:** análise do plano linguístico do vocabulário, observando se há o controle de sinonímia;
- **Software adotado pelo sistema de informação:** análise das consequências diretas que a seleção de determinado software tem na escolha do tipo de linguagem de indexação a ser usada em um acervo de documentos e sua organização interna.

Gil Leiva (2008, p. 213) propõe duas metodologias de avaliação: tomando como objeto de análise o próprio vocabulário controlado (*avaliação intrínseca*) ou estudando o papel do vocabulário controlado na recuperação da informação (*avaliação extrínseca*).

A *avaliação intrínseca* estuda o interior de algo e para um vocabulário controlado refere-se à análise de todos ou alguns de seus elementos por especialistas do assunto (GIL LEIVA, 2008, p. 213). Esse tipo de avaliação pode ser de dois tipos:

- a) **Avaliação intrínseca qualitativa:** visa revisar e avaliar um vocabulário controlado para extrair conclusões sobre todos ou qualquer um de seus elementos contemplados. Para tal, segue recomendações:
- **Introdução:** verificar se ele contém os dados mínimos recomendados pelas normas e práticas habituais;
 - **Fontes básicas:** verificar as fontes de informação usadas para a extração da terminologia;
 - **Campos ou facetas temáticas:** analisar se as áreas principais e secundárias mais importantes do domínio coberto estão representadas;
 - **Abreviações, símbolos e tipografias convencionais:** verificar se essas convenções sugeridas pela norma ou pela prática foram usadas e se estão incluídas na introdução, no caso de serem diferentes.
 - **Notas de explicação:** examinar a existência ou não, a redação ou a utilidade das notas explicativas que acompanham os descritores;
 - **Códigos de classificação:** verificar a presença ou ausência desses códigos de classificação escolhidos, bem como a implantação correspondente. Da mesma forma, a correspondência dos códigos atribuídos aos campos temáticos ou facetas principais

e secundárias. Na apresentação sistemática deve ser revisada com os códigos que acompanham cada descritor na apresentação alfabética e no índice KWIC ou KWOC;

- **Indicadores de classificação:** verificar a existência ou não desses indicadores de classificação na apresentação sistemática para indicar o princípio lógico da organização;
- **Acrônimos e acrônimos:** determinar se os acrônimos e acrônimos usados como descritores ou termos preferenciais são suficientemente conhecidos pela comunidade à qual o tesouro é direcionado;
- **Relações semânticas entre os termos:** verificação da reciprocidade dos relacionamentos existentes, mas principalmente a adequação dos relacionamentos poli hierárquicos e associativos;
- **Grau de especificidade:** estudar o nível de especificidade terminológica em relação ao objetivo do vocabulário controlado e seus usuários em potencial;
- **Apresentações:** analisar a reciprocidade entre as diferentes apresentações selecionadas para o arranjo terminológico;
- **Formato (papel, eletrônico, web-):** no caso de um vocabulário controlado em CD-ROM, deve-se prestar atenção especial às personalidades que navegam no vocabulário controlado;
- **Interoperabilidade, formatos e protocolos:** conhecer o grau de comunicação com outros vocabulários em um ambiente digital;
- **A eficácia do vocabulário controlado para a indexação e recuperação:** coleta de dados, ensaios (tanto na consistência na indexação quanto na integridade e precisão na recuperação) e impressões de profissionais dedicados à indexação, como aqueles que realizaram pesquisas, bem como dos usuários finais para verificar a utilidade do dicionário de sinônimos.

- b) **Avaliação quantitativa ou estatística intrínseca:** é obtida a partir de cálculos matemáticos aplicados a diferentes elementos presentes no vocabulário controlado.

A segunda avaliação proposta por Gil Leiva (2008) é a avaliação extrínseca, que explica ser um estudo do impacto do vocabulário controlado nos sistemas de informação que utilizam a linguagem padronizada para realizar o processo de indexação e auxiliar na recuperação. O autor salienta que desta forma há espaço para qualquer teste de consistência na indexação e cálculo da eficácia da recuperação em que um vocabulário está envolvido.

3 PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos foram realizados em duas etapas: a)

análise da literatura sobre avaliação do uso de vocabulário controlado para o controle de vocabulário em repositórios institucionais e; b) aplicação exploratório-explicativa com estudos de caso de seis repositórios institucionais que utilizam vocabulário controlado e em um repositório institucional que não utiliza vocabulário controlado para a indexação e recuperação por assuntos.

Para os estudos de caso, a elaboração dos critérios de avaliação considerou as categorias de análise de Gil Leiva (2008, p. 213): *avaliação intrínseca*, contendo critérios voltados para o repositório institucional propriamente dito e a *avaliação extrínseca*, que estuda o papel do vocabulário controlado na representação e recuperação da informação no repositório institucional.

3.1 PROCEDIMENTOS DE AVALIAÇÃO

I Categoria de análise com enfoque na avaliação intrínseca

Considerando-se que a avaliação intrínseca estuda o interior de algo, essa categoria irá trazer critérios de avaliação voltados para o repositório institucional propriamente dito:

- 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 controlado correspondente ao domínio:** se o vocabulário controlado empregado confere ao domínio contemplado pelo repositório institucional (ALVES, 2021).
- c) Vocabulário controlado para o 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).

II Categoria de análise com enfoque na avaliação extrínseca

A avaliação extrínseca volta-se para o impacto do vocabulário controlado nos repositórios institucionais que utilizam a linguagem padronizada para realizar o processo de indexação e auxiliar na recuperação:

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).

3.2 UNIVERSO DE PESQUISA

Foram definidos seis repositórios institucionais que utilizam diferentes vocabulários controlados, identificados por Fujita e Tolare (2019), e um repositório institucional que não utiliza vocabulário controlado. Cada repositório institucional é identificado pela letra “R” e um número (Quadro 1):

Quadro 1 - Amostra de repositórios institucionais para aplicação da metodologia de avaliação

Código	Nome do Repositório	Vocabulário controlado
R1	ARES – Acervo de Recursos Educacionais em Saúde – UNA.SUS	DeCS – Descritores em Ciências da Saúde
R2	Repositório ALICE - EMBRAPA	Thesagro e NAL Thesaurus
R3	ReP - Repositório USP	VocaUSP
R4	Biblioteca Digital do Senado Federal	Tesouro do Senado Federal
R5	Biblioteca Virtual em Saúde – Ministério da Saúde	Terminologia da Saúde
R6	Infoteca-e – Informação Tecnológica em Agricultura	Thesagro e NAL Thesaurus
R7	Repositório Unesp	Não utiliza

Fonte: Elaborado pelas autoras

4 RESULTADOS

Os resultados foram pautados nas categorias de análise *avaliação intrínseca* (que contêm 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 com e sem vocabulário controlado, 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 foi feita por toda a interface de cada repositório institucional (Quadro 2):

Quadro 2 - Resultados do critério “Disponibilidade de link para o vocabulário controlado”

Repositório Institucional	Resultado
R1	É apresentado na interface do repositório um ícone de lupa que direciona para uma lista em ordem alfabética de palavras-chaves do DeCS
R2	Não localizado
R3	Não localizado
R4	Não localizado
R5	Não localizado
R6	Não localizado
R7	Não localizado

Fonte: Elaborado pelas autoras

Dos seis repositórios com vocabulário controlado 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 para uma possível pesquisa pelos usuários sem necessidade de contato com os gestores.

O **R7** não possui *link* em sua interface que direcione os usuários para um vocabulário controlado ou apresenta alguma informação do uso de um vocabulário para a representação e recuperação dos recursos informacionais,

embora a instituição possua tesouro próprio (Tesouro UNESP). Destaca-se seu uso como vocabulário controlado no catálogo online Athena, enquanto no repositório institucional da mesma instituição não há a adoção desse vocabulário controlado para a representação e recuperação por assuntos dos recursos informacionais.

b) Vocabulário correspondente ao domínio

Para esse critério, foi analisado o domínio do vocabulário controlado, caso utilizado pelo repositório institucional, a fim de verificar se é correspondente ao domínio de conhecimento abrangente. Os resultados (Quadro 3) demonstraram ampla integração de vocabulários controlados correspondentes ao domínio que o repositório institucional abarca:

Quadro 3 - Resultados do critério “vocabulário controlado correspondente ao domínio”

Repositório Institucional	Resultado
R1	Sim
R2	Sim
R3	Sim
R4	Sim
R5	Sim
R6	Sim
R7	Não tem

Fonte: Elaborado pelas autoras

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.

Conforme mencionado, não há nenhuma informação que relate o uso de um vocabulário controlado no **R7** para realizar o controle dos termos utilizados no momento do autoarquivamento as teses e dissertações, nem para outros tipos de recursos informacionais. Portanto, nesse critério não foi possível analisar se o vocabulário controlado corresponde ao domínio em que estão inseridas.

c) Vocabulário controlado para o autoarquivamento

Os resultados desse critério (Quadro 4) foram coletados a partir da análise na interface do repositório institucional e de seus menus em busca da existência de orientação para o uso do vocabulário controlado durante o autoarquivamento e manual de direcionamento de como realizá-lo. Foi utilizada a palavra “indisponível” quando nenhum acesso foi localizado para o uso do vocabulário controlado no autoarquivamento.

Quadro 4 - Resultados do critério “vocabulário controlado para o autoarquivamento”

Repositório Institucional	Autoarquivamento	Disponibilização de vocabulário controlado
R1	É possível realizar o autoarquivamento, porém, é necessário ser autorizado	Não foi encontrada orientação para o uso do vocabulário controlado. Há um menu que direciona para alguns manuais, entretanto, a grande maioria não abria e outros não correspondiam
R2	Não há acesso para autoarquivamento	Indisponível
R3	Não há acesso para autoarquivamento	Indisponível
R4	Não há acesso para autoarquivamento	Indisponível
R5	Não há acesso para autoarquivamento	Indisponível
R6	Não há acesso para autoarquivamento	Indisponível
R7	É realizado o autoarquivamento de teses e dissertações pelos autores-alunos da pós-graduação das diversas áreas oferecidas pela Universidade	Na interface do repositório há um <i>link</i> com a disponibilização de informações e manual sobre autoarquivamento

Fonte: Elaborado pelas autoras

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, é necessária autorização. Não foi possível o acesso ao 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.

No **R7** é realizado o autoarquivamento de teses e dissertações pelos

autores-alunos da pós-graduação das diversas áreas oferecidas pela Universidade. Na interface do repositório institucional há um *link* com a disponibilização de informações sobre autoarquivamento, contendo um breve conceito do que é esse processo, prazos para o arquivamento, informações sobre a versão impressa e informações de como realizá-lo. Essas informações estão disponíveis em um tutorial textual, um tutorial em vídeo, uma portaria de 2015, um folder explicativo e um fluxograma de autoarquivamento.

Após realizar login no *link* disponível exclusivamente para o autoarquivamento de teses e dissertações, o autor-aluno realiza a inserção dos metadados que correspondem ao seu trabalho, sendo compostos por: título e subtítulo (em português e em inglês), nome completo, nome do orientador e coorientador (se houver), data da defesa, resumo (em português e em inglês), palavras-chave, idioma, tipo de trabalho (dissertação ou tese), unidade universitária, Programa de pós-graduação, área de concentração e linha de pesquisa. Em seguida, a escolha da versão do trabalho (parcial ou integral) para o *upload* de trabalho em formato PDF.

No entanto, a disponibilização de um vocabulário controlado para a inserção de termos que correspondam ao trabalho é inexistente, fazendo com que os trabalhos sejam representados no repositório institucional pela linguagem natural, isto é, pelas palavras-chave dos autores-alunos.

4.2 AVALIAÇÃO EXTRÍNSECA

Os critérios dessa categoria abordam o impacto do vocabulário controlado na representação e recuperação dos recursos informacionais. Foi definido um termo (palavra-chave) que correspondesse ao domínio do repositório institucional para realizar uma busca simples e analisar o título das teses e dissertações que sejam apresentadas nos dez primeiros recursos informacionais recuperados. Observou-se como o termo de busca foi apresentado na recuperação comparado com a leitura dos títulos dos recursos informacionais recuperados, buscando ambiguidades, sinonímia, polissemia, homonímia ou se foi recuperado exatamente como posto na busca.

a) Controle terminológico

Referente ao controle terminológico (Quadro 5), verificou-se o correto controle de quatro repositórios, sendo eles: **R2, R3, R4 e R6**.

Quadro 5 - Resultados do critério “controle terminológico”

Repositório Institucional	Termo	Resultado
R1	Hipertensão	4 recursos informacionais apresentaram o termo de busca, enquanto que 6 apresentaram o termo “hipertensos”
R2	Plantio	1 recurso informacional não apresentou o termo de busca, enquanto que em 9 o termo estava de acordo
R3	Jornalismo	Todos os recursos informacionais recuperados estavam de acordo com o termo de busca
R4	Direito	Todos os recursos informacionais recuperados estavam de acordo com o termo de busca
R5	Psicoterapia	7 recursos informacionais apresentaram o termo de busca; 2 não apresentaram e 1 apresentou os termos “psicologia” e “psicólogos”
R6	Irrigação	Todos os recursos informacionais recuperados estavam de acordo com o termo de busca
R7	Aquisição	Dissertações: 6 recursos informacionais apresentaram o termo de busca no título (sendo que destes 4 sobre o assunto tratado); 4 não apresentaram o termo de busca Teses: 6 recursos informacionais apresentaram o termo de busca no título (sendo que destes 5 sobre o assunto tratado); 4 não apresentaram o termo de busca

Fonte: Elaborado pelas autoras

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”.

No **R7**, o termo escolhido foi “aquisição”, utilizado, por exemplo, nas teses e dissertações pertencentes ao campo da Ciência da Informação. Para este resultado, foi necessário realizar a busca diretamente por tipo de material, pois, na busca simples pelo campo disponível na interface não foi recuperado nenhuma tese ou dissertação nos dez primeiros resultados. Na busca realizada em “dissertação de mestrado”, o termo “aquisição” foi apresentado no título de seis dos dez recursos

informacionais recuperados. Mesmo havendo o termo no título, só quatro dos seis recursos informacionais realmente eram referentes à aquisição de algo e entre os quatro que não apresentaram o termo no título não correspondiam ao assunto tratado. Já em relação ao tipo “tese de doutorado” foram recuperados seis dos dez primeiros recursos informacionais com o termo “aquisição” no título. Desses seis, cinco correspondiam ao assunto tratado e entre os quatro que não apresentaram o termo no título não tinham correlação direta em suas pesquisas.

b) Forma dos termos

Quanto à forma como os termos são apresentados (Quadro 6), foi verificado se o idioma do termo de busca apresentava singularidade ou pluralidade, se as palavras-chave são compostas ou simples, se são adotadas abreviaturas, se os recursos informacionais eram recuperados sozinhos ou se apresentavam alguma composição ou abreviações do termo.

Quadro 6 - Resultados do critério “forma dos termos”

Repositório Institucional	Termo	Resultado
R1	Hipertensão	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular e no plural; apresentaram composição como “hipertensão gestacional”; não constaram abreviações
R2	Plantio	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular ou não apresentaram o termo de busca; apresentaram composição como “sistema plantio direto”; não constaram abreviações
R3	Jornalismo	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular e no plural; apresentaram composição como “jornalismo internacional”; não constaram abreviações
R4	Direito	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular; apresentaram composição como “direito civil”; não constaram abreviações
R5	Psicoterapia	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular; apresentaram composição como “psicoterapia infantil”; não constaram abreviações
R6	Irrigação	Todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular; apresentaram composição como “irrigação localizada”; não constaram abreviações

R7	Aquisição	Dissertações: todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular; apresentaram simples ou com composição, como, “aquisição de dados”; não constaram abreviações Teses: todos os recursos informacionais recuperados em português; apresentaram o termo de busca no singular; apresentaram o termo de busca no singular; apresentaram simples ou com composição, como, “aquisição volumétrica”; não apresentaram abreviações
----	-----------	--

Fonte: Elaborado pelas autoras

Quanto 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.

Em “dissertações de mestrado” no **R7** foram recuperados termos apenas no singular, em português e com a apresentação simples e composta, como, “aquisição de dados”, e não houve nenhuma abreviatura. No tipo “tese de doutorado”, foram recuperados recursos informacionais que apresentavam o termo no título apenas no singular e em português, havendo apresentação simples e composta, como, “aquisição volumétrica” e não apresentou abreviações.

d) Grau de especificidade

Para esse critério, observou-se se os recursos informacionais recuperados estavam de acordo com a área de domínio dos repositórios institucionais (Quadro 7).

Quadro 7 - Resultados do critério “grau de especificidade”

Repositório Institucional	Termo	Resultado
R1	Hipertensão	Correspondem à área de domínio
R2	Plantio	Correspondem à área de domínio
R3	Jornalismo	Correspondem à área de domínio
R4	Direito	Correspondem à área de domínio
R5	Psicoterapia	Correspondem à área de domínio
R6	Irrigação	Correspondem à área de domínio
R7	Aquisição	Correspondem à área de domínio

Fonte: Elaborado pelas autoras

Os seis repositórios institucionais atenderam ao grau de especificidade na recuperação dos recursos informacionais. No **R7** a busca recuperou recursos informacionais no tipo “dissertações de mestrado” voltados para as áreas que a universidade abrange. Um facilitador para isso é a seleção de critérios para o arquivamento e a disponibilização de autoarquivamento. O mesmo resultado corresponde para as recuperações no tipo “teses de doutorado”.

4.3 ANÁLISE DOS RESULTADOS DA AVALIAÇÃO INTRÍNSECA E EXTRÍNSECA

Em uma visão geral dos resultados obtidos com os critérios de avaliação intrínseca e extrínseca (Quadro 8), observa-se que o uso do vocabulário controlado nos sete repositórios analisados é possivelmente restrito à representação dos recursos informacionais por profissionais da informação e indisponível no autoarquivamento e na estratégia de busca para recuperação. Nenhum dos repositórios analisados disponibiliza o vocabulário controlado para uso de autores no autoarquivamento ou na busca. O efeito do uso de vocabulários controlados é visível no controle terminológico de quatro dos sete repositórios com 100% de compatibilidade entre o termo de busca e o conteúdo dos 10 primeiros documentos recuperados.

Quadro 8 - Síntese dos critérios avaliativos do uso de vocabulários controlados (VC) pelos repositórios institucionais

Critérios avaliativos	R1 Com VC	R2 Com VC	R3 Com VC	R4 Com VC	R5 Com VC	R6 Com VC	R7 Sem VC
AVALIAÇÃO INTRÍNSECA							
Disponibilidade de link para o VC	É encontrado o link para o VC DeCS.	Não encontrado.	Não encontrado.	Não encontrado.	Não encontrado.	Não encontrado.	Não encontrado.
VC correspondente ao domínio	Sim	Sim	Sim	Sim	Sim	Sim	Não há
VC para o autoarquivamento	Autorquivamento: com autorização; VC: indisponível	Autorquivamento: sem acesso; VC: indisponível	Autorquivamento: sem acesso; VC: indisponível	Autorquivamento: sem acesso; VC: indisponível	Autorquivamento: sem acesso; VC: indisponível	Autorquivamento: sem acesso; VC: indisponível	Autorquivamento: acessível; VC: indisponível
AVALIAÇÃO EXTRÍNSECA							

Controle terminológico	40% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	90% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	100% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	100% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	70% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	100% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca	60% de compatibilidade do conteúdo dos documentos recuperados com o termo de busca
Forma dos termos	Termos em português; termos no singular e plural; termos compostos no singular; sem abreviaturas	Termos em português; termos no singular; composição de termos; sem abreviaturas	Termos em português; termos no singular e plural; poucos termos compostos; sem abreviaturas	Termos em português; termos no singular; todos os termos compostos; sem abreviaturas	Termos em português; termos no singular e plural; termos compostos; sem abreviaturas	Termos em português; termos no singular; termos compostos; sem abreviaturas	Termos em português; termos no singular; termos compostos; sem abreviaturas
Grau de especificidade	Correspondem à área de domínio	Correspondem à área de domínio	Correspondem à área de domínio	Correspondem à área de domínio	Correspondem à área de domínio	Correspondem à área de domínio	Correspondem à área de domínio

Fonte: Elaborado pelas autoras

Referente ao primeiro critério da categoria de **avaliação intrínseca**, que teve como foco analisar o repositório propriamente dito, foi possível destacar a falta do direcionamento ao vocabulário integrado ao sistema para conhecimento dos usuários e para uma possível pesquisa por algum pesquisador que queira saber sem precisar entrar em contato com os gestores. Dos seis repositórios apenas o **R1** fornecia acesso direto ao vocabulário utilizado. Os resultados do segundo critério demonstraram ampla integração de vocabulários controlados correspondentes ao domínio dos seis primeiros repositórios com exceção do **R7** que não utiliza vocabulário controlado. Esse resultado é de grande relevância, pois, quanto mais especializado o vocabulário for em relação ao domínio, maior será a precisão no momento da representação e recuperação do conhecimento ali guardado e disseminado (KOBASHI, 2008; TERRA, VASCONCELOS, 2015).

O terceiro critério correspondeu ao autoarquivamento da produção acadêmica pelos autores e disponibilização de vocabulário controlado. O

resultado foi parecido com o resultado do primeiro critério acrescido do **R7**, além do **R1**. Para que seja realizado o autoarquivamento no **R1** é necessária autorização. Embora fosse apresentado no menu uma aba que direcionasse para vários manuais, o manual que corresponde ao preenchimento dos dados não podia ser aberto. Nesse caso, não foi possível constatar se a representação dos recursos informacionais é realizada por linguagem natural ou se os autores são direcionados para o vocabulário controlado integrado ao repositório institucional. Para intermediar o processo de representação de modo a garantir o acesso à recuperação da informação, é preciso que o vocabulário controlado esteja disponível e acessível e que seja acompanhado de orientações ao seu uso em conformidade com a política de indexação do repositório institucional (BOCCATO; FUJITA, 2006).

No **R7**, os resultados obtidos a partir dos critérios das categorias de análise permitiram observar um descontrole da linguagem no momento da recuperação dos recursos informacionais, demonstrando falha na representação deles. Por conta da falta de controle de vocabulário no momento da representação, há excessos de recursos informacionais na recuperação, apresentando pesquisas que não são correlacionadas com o termo de busca. Esse fator pode estar relacionado com o uso da linguagem natural no momento da inserção dos metadados no autoarquivamento das teses e dissertações pelos autores-alunos. O fato de não haver indexação dos termos que representem esses materiais de forma padronizada com o domínio e que sejam prontamente correspondentes à pesquisa ocasionam inconsistências na recuperação dos recursos informacionais, como ressaltado em pesquisas de Terra e Vasconcelos (2015), Silva (2015), Fujita e Tolare (2019).

É possível salientar que pelo fato de a referida universidade possuir o seu próprio vocabulário controlado, isto é, o Tesouro Unesp, o mesmo poderia estar integrado ao repositório institucional de forma que estivesse disponível para conhecimento dos usuários e utilizado no momento do autoarquivamento pelos autores, visando melhor padronização na representação e recuperação por assuntos neste sistema de recuperação da informação. Porém, essa questão é complexa, pois, ao mesmo tempo em que poderia facilitar, poderia causar

dificuldades, visto que muitos autores-alunos não possuem conhecimento da existência do vocabulário controlado e não sabem utilizá-lo. Para resolver tal lacuna, seria necessária a disponibilidade de tutoriais de seu uso e treinamentos, preferencialmente no início do mestrado/doutorado. Além disso, a inserção de um novo campo de metadados direcionado para os bibliotecários que realizam a validação dos metadados inseridos pelos autores-alunos possibilitaria um melhor controle e padronização da linguagem, tanto na representação quanto na recuperação por assuntos.

Já o primeiro critério de avaliação da categoria de **avaliação extrínseca**, que teve enfoque na representação e recuperação da informação, resultou na observação de que o controle terminológico estava correto em quatro repositórios. Os outros três repositórios apresentaram variações terminológicas em relação ao termo colocado na busca. Esse resultado dos quatro repositórios coincide com o uso de vocabulário controlado adequado ao domínio e o fato de que a representação dos recursos informacionais seja validada pelo mesmo vocabulário controlado.

No que tange à recuperação com o idioma proposto, todos recuperaram documentos em português sem precisar realizar busca avançada. Os repositórios **R1** e **R3** apresentaram singularidade e pluralidade, já 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 ele estava no singular e os demais apresentaram composição independentemente de singular ou plural e nenhum dos repositórios institucionais apresentou abreviações na recuperação do termo.

Em relação ao terceiro critério, que era analisar se os documentos recuperados abrangiam a especificidade em relação aos usuários potenciais do repositório institucional, todos os repositórios institucionais atenderam ao grau de especificidade na recuperação dos documentos.

No geral, a aplicação de critérios avaliativos no **R7** ressalta 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.

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 ou padronizando essa comunicação, será possível um maior nível de precisão, diminuindo ruídos na recuperação dos recursos informacionais.

Para que se tenha consistência na representação e na recuperação por assuntos dos recursos informacionais em repositórios institucionais, 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.

AGRADECIMENTOS

A pesquisa científica contou com o apoio do CNPq com bolsa de Iniciação Científica e bolsa de Produtividade em Pesquisa.

REFERÊNCIAS

- ALVES, L. 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, Universidade Estadual Paulista, Marília, 2021.
- ANSI/NISO. Z39.19 – 2005. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda: NISO Press, 2005.
- BOCCATO, V. R. C; FUJITA, M. S. L. Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p. 267-281, 2006.
- 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 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.
- LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos/livros, 2004.
- LEITE, F. C. L.; AMARO, B.; BATISTA, T.; COSTA, M. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012.
- SILVA, B. **Vocabulário de termos livres e controlados para a coleção de teses e dissertações da Universidade Federal de Viçosa**. 2015. Dissertação (Mestrado em Biblioteconomia) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015.
- SOUSA, J. L. **Avaliação do padrão Simple Knowledge Organization System (SKOS) para a representação de vocabulários controlados**. 2019. 100 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2019.

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 (ENANCIB)*, 8., 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.

EVALUATION OF CONTROLLED VOCABULARY USE IN INSTITUTIONAL REPOSITORIES

ABSTRACT

Objective: To develop and apply a methodology for evaluating the use of controlled vocabularies for vocabulary control in institutional repositories. **Methodology:** The methodological procedures were carried out in two stages: a) analysis of the literature on the evaluation of the use of controlled vocabulary to control vocabulary in institutional repositories and; b) exploratory-explanatory application with case studies of six institutional repositories that use controlled vocabulary and in an institutional repository that does not integrate the controlled vocabulary for indexing and retrieval by subject. The second stage consisted of applying evaluative criteria in two categories of analysis: intrinsic evaluation, containing criteria aimed at the repository itself, and extrinsic evaluation, with criteria aimed at the analysis of representation and information retrieval. **Results:** In most of the analyzed institutional repositories that integrate controlled vocabulary, the retrieval by subjects presents good precision from searched terms, while the repository that does not integrate the controlled vocabulary needs greater terminological control of the terms correlated to the theses and dissertations searches to greater precision in the retrieval of informational resources. **Conclusions:** The more specialized the vocabulary controlled in relation to the domain of the institutional repository, the more effective the subject retrieval will be. The lack of standardization in the representation causes failures in the recovery with inconsistencies in the interaction of the institutional repository with its users.

Descriptors: Institutional repositories. Controlled vocabulary. Indexing evaluation. Subject retrieval.

EVALUACIÓN DEL USO DE VOCABULARIO CONTROLADO EN REPOSITARIOS INSTITUCIONALES

RESUMEN

Objetivo: Desarrollar y aplicar una metodología para evaluar el uso de vocabularios controlados para el control de vocabulario en repositorios institucionales. **Metodología:** Los procedimientos metodológicos se realizaron en dos etapas: a) análisis de la literatura sobre la evaluación del uso de vocabulario controlado para controlar el vocabulario en repositorios institucionales y; b) aplicación exploratoria-explicativa con estudios de caso de seis repositorios institucionales que utilizan vocabulario controlado y en uno repositorio institucional que no integra el vocabulario controlado para indización

y recuperación por materias. La segunda etapa consistió en aplicar criterios evaluativos en dos categorías de análisis: *evaluación intrínseca*, con criterios dirigidos al propio repositorio, y *evaluación extrínseca*, con criterios dirigidos al análisis de la representación y recuperación de información. Resultados: En la mayoría de los repositorios institucionales analizada que integran vocabulario controlado, la recuperación por materias presenta buena precisión a partir de los términos buscados, mientras que el repositorio que no integra el vocabulario controlado necesita mayor control terminológico de los términos correlacionados con las búsquedas de tesis y disertaciones para mayor precisión en la recuperación de los recursos informacionales.

Conclusiones: Cuanto más especializado sea el vocabulario controlado en relación al dominio del repositorio, más efectiva será la recuperación del tema. La falta de estandarización en la representación conduce a fallas en la recuperación con inconsistencias en la interacción del repositorio institucional con sus usuarios.

Descriptores: Repositorios institucionales. Vocabulario controlado. Evaluación de la indización. Recuperación por materias.

Recebido em: 07.12.2021

Aceito em: 28.03.2022