DOI: 10.5433/1981-8920.2020v25n1p229

METODOLOGIAS PARA REVISÃO E ATUALIZAÇÃO DE TESAUROS: MAPEAMENTO DA LITERATURA

METHODOLOGIES FOR REVIEWING AND UPDATING THESAURUS: LITERATURE MAPPING

Ana Carolina Ferreira[a]
Benildes Coura Moreira dos Santos Maculan[b]

RESUMO

Introdução: Os tesauros continuam desempenhando importante função na padronização terminológica para facilitar a comunicação da informação. Haja vista essa reconhecida importância, tais instrumentos não são revisados e atualizados segundo fundamentos metodológicos claros. **Objetivo:** Tendo em vista esse problema, objetivou-se diagnosticar, por meio do mapeamento da literatura, as principais metodologias empregadas para a revisão e a atualização de tesauros como subsídio ao desenvolvimento de estruturas conceituais mais consistentes. **Metodologia:** Para a realização desse mapeamento, utilizou-se o protocolo de revisão sistemática proposto por COMO... (2015). As buscas foram realizadas nas principais bases de dados da área de Ciência da Informação no contexto nacional e internacional. **Resultados:** Os resultados indicam distintos enfoques empregados na revisão e na atualização de tesauros, a exemplo de procedimentos para a modelagem conceitual intelectual, estudo de *logs* de pesquisa dos usuários, a adoção de *folsksonomias* e o emprego de tecnologias da *web* semântica. A partir dos relatos identificados, observou-se que a revisão e a atualização de tesauros envolvem três aspectos básicos: 1) aspectos terminológicos e conceituais gerais; 2) aspectos gerenciais e 3) estudos empíricos com o emprego de diferentes técnicas de atualização. **Considerações finais**: Ainda que se identifique uma multiplicidade de enfoques e técnicas para a atualização de tesauros, é preciso considerar a avaliação do contexto de uso, as necessidades de informação dos usuários e os propósitos da revisão para que os procedimentos adotados possam ser efetivos.

Descritores: Tesauro. Revisão. Atualização. Metodologia.

[a] Doutoranda pelo Programa de Pós-graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (PPG-GOC/ECI/UFMG). E-mail: anacarolinaf2004c@gmail.com

[b] Doutora pelo Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI/ECI/UFMG). Professora do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI/UFMG). E-mail: benildes@gmail.com

1 INTRODUÇÃO

Historicamente, os tesauros têm sido empregados na padronização terminológica para a representação do conhecimento e a recuperação da informação em bases de dados bibliográficas, catálogos e índices. Sua função principal refere-se à determinação de uma linguagem unívoca e sem ambiguidade para compartilhamento entre os diversos sujeitos do processo de comunicação da informação: bibliotecários, documentos, sistemas de recuperação da informação e usuários.

Maculan e Lima (2017, p. 55) apontam que o tesauro conceitual (ou tesauro terminológico) "é composto por uma terminologia (conjunto de descritores preferidos e não-preferidos) que tem o conceito como seu elemento nuclear, uma vez que este denota certo nível de abstração e de consenso dentro de um dado domínio". Nos tesauros, essa terminologia apresenta um conjunto de conceitos inter-relacionados por relações semânticas (de equivalência, hierárquicas e associativas), que permitem a construção do significado do termo em um dado contexto de aplicação.

Considerando-se a importância dos tesauros para a representação e a recuperação da informação, este artigo apresenta os resultados de uma revisão de literatura que objetivou investigar os métodos e as técnicas empregados na revisão e na atualização desses instrumentos para subsidiar o desenvolvimento de terminologias mais consistentes. É importante destacar que a apresentação dos tesauros, geralmente, expõe a necessidade de que se façam revisões e atualizações periódicas em sua terminologia, haja vista a dinamicidade da produção do conhecimento, mas observa-se que a maioria deles não detalham os procedimentos utilizados nesse processo de atualização (KIM; KIM, 1977; SIMÕES, 2008; MACULAN, 2015; SOUZA, 2017). O mapeamento da literatura focou no tipo de intervenção principal realizada no tesauro e nos resultados mais relevantes dos estudos experimentais, explorando o estado da arte do tema sobre metodologias para a revisão e a atualização de tesauros.

Nesse contexto, na seção 2 apresenta-se a metodologia de revisão sistemática utilizada para a elaboração da revisão; na seção 3 estão descritos,

sucintamente, os trabalhos identificados e as principais metodologias para a revisão e a atualização de tesauros; na seção 4, expõe-se uma análise desses documentos e, na seção 5, são tecidas as considerações finais.

2 METODOLOGIA PARA A REVISÃO DE LITERATURA

Para a condução da revisão, adotou-se o protocolo de revisão sistemática de COMO... (2015), sendo empregados sete aspectos básicos nessa atividade: 1) objetivo da revisão; 2) quais as questões devem ser respondidas com a revisão; 3) quais foram as fontes pesquisadas e como elas foram selecionadas; 4) quais foram as palavras-chave utilizadas na busca; 5) como as palavras-chave foram identificadas e definidas; 6) que tipos de documentos foram incluídos (exemplos: artigos, teses, capítulos de livros, anais) e em qual idioma; 7) quais foram os critérios de seleção, inclusão e exclusão dos trabalhos e como eles foram aplicados (COMO..., 2015).

Atendendo ao primeiro aspecto, como descrito anteriormente, o objetivo desta revisão de literatura foi o de identificar trabalhos sobre o tema da revisão e da atualização de tesauros para verificar quais metodologias foram adotadas nesses estudos e quais melhorias foram implementadas nos tesauros.

Para o segundo aspecto, definiram-se algumas indagações que deveriam ser respondidas com a revisão: quais foram as melhorias implementadas na revisão e na atualização de tesauros? Quais os procedimentos que foram adotados? Quais resultados foram obtidos nos estudos experimentais?

No terceiro aspecto determinaram-se as seguintes fontes para a coleta de dados: *Library & Information Science Abstracts* (LISA); *Library, Information Science & Technology Abstracts* (LISTA-EBSCO); *Scopus*; *Web of Science*; *E-prints in Library and Information Science* (E-LIS); *Wiley Online Library* (Portal Capes); *Annual Review of Information Science and Technology* (ARIST) e Base de Dados em Ciência da Informação (BRAPCI), as quais foram selecionadas por serem as principais bases de dados da Ciência da Informação no contexto internacional e nacional; e também o Google Acadêmico foi amplamente explorado.

Para os aspectos quarto e quinto, que tratam das palavras-chave, aquelas

utilizadas na busca nas fontes selecionadas foram: *thesaurus*, *thesauri*, *controlled vocabulary*, *indexing language*, *construction*, *review*, *maintenance*, *update*, *updating*, *reformulation*, *restructuring*, *reengineering*, *enriching*, *conceptual modeling*, *semantic relationships*, *conceptual relationships*. Esclarece-se que a sintaxe das expressões de busca foi adaptada de acordo com as especificidades e a estrutura de cada base de dados. Esse conjunto de palavras-chave foi identificado e definido a partir do estudo de artigos, trabalhos técnicos e acadêmicos que abordavam o tema da revisão e da atualização de tesauros.

Em relação ao sexto aspecto, da tipologia de documentos, foram recuperados artigos científicos, trabalhos apresentados em congressos, teses, capítulos de livros e a apresentação dos próprios tesauros, em inglês, português, espanhol, castelhano e francês. Para o sétimo e último aspecto, dos critérios de seleção, inclusão e exclusão dos trabalhos recuperados, os critérios de seleção e inclusão abarcaram os estudos teóricos e práticos sobre o tema da atualização, revisão, manutenção, reformulação, aprimoramento, gestão e reengenharia de tesauros e vocabulários controlados. Já os critérios de exclusão foram os relacionados aos trabalhos que fogem ao escopo desta revisão, em especial aqueles sobre manutenção estritamente automática de tesauros, indexação automática com o uso de tesauros, estudos comparativos de uso do tesauro e de palavras-chave na recuperação da informação, integração da estrutura de vocabulário controlado como funcionalidade de pesquisa e navegação e estudos empíricos de interação do usuário com a interface de busca aprimorada por tesauro.

3 RESULTADOS DA REVISÃO DE LITERATURA

A partir dos critérios de seleção, inclusão e exclusão, foram recuperados 25 documentos. Esses trabalhos são apresentados por ordem crescente do ano de publicação e suas descrições procuram evidenciar os principais procedimentos metodológicos adotados na revisão e na atualização de tesauros, e também os principais resultados quando se tratar de estudos experimentais.

Kim (1973) apresenta dois diferentes métodos ou técnicas para o

processo de atualização de tesauros: 1) técnica de análise de conteúdo baseada em uma abordagem etimológica e 2) a teoria do campo semântico. Na primeira proposta, as técnicas adotadas são: mensuração quantitativa da taxa de adição de novas palavras com novos significados, a taxa de adição de novos significados para palavras antigas e a taxa de exclusão de palavras em virtude do desuso. A proposta de uso da teoria do campo semântico fundamenta-se no princípio de associação de palavras, a partir de similaridade, contiguidade e construção do significado das palavras de forma dependente de outras palavras. Assim, o significado da palavra é construído no contexto de suas redes associativas.

Soergel (1974) identifica três tipos de mudanças que podem ocorrer nos tesauros: 1) mudança em relação aos termos sinônimos e homônimos e em sua respectiva estrutura; 2) alteração em parte da estrutura classificatória; 3) mudanças na própria linguagem de indexação que afetam a recuperação da informação. Isso implica na necessidade de reindexar os documentos, tais como: a adição ou eliminação de um novo descritor; a subdivisão de um descritor existente em um conjunto de descritores específicos; a mudança na definição ou no uso de um descritor; adição ou eliminação de um relacionamento hierárquico, especialmente quando ocorre a alocação de um descritor em outro grupo na hierarquia; e a adição ou eliminação de um relacionamento.

Em IBICT, 1984 são apresentadas orientações com base na ISO 2788 que se aplicam também à atualização de tais instrumentos: 1) trabalho preliminar (comunicação de intenção; certificação sobre a existência de outros vocabulários semelhantes à área temática); 2) métodos de compilação (dedutivo e indutivo) e previsão de revisão dos relacionamentos e hierarquias a cada inclusão ou exclusão de termos; 3) registro dos termos em uma ficha; 4) revisão dos relacionamentos entre cada novo termo e os termos e as hierarquias já registradas; 5) especificidade: deve ser reservada ao núcleo central do assunto coberto; 6) admissão e exclusão de termos; 7) a frequência de uso dos termos e teste do tesauro por meio de uma experiência-piloto.

Batty (1989), estabelece que a manutenção de tesauros envolve quatro procedimentos básicos: 1) adição de novos termos; 2) exclusão de termos; 3)

modificação de termos; e 4) um processo de gestão. O autor sugere, também, que as mudanças significativas no escopo ou no uso dos termos existentes devam ser registradas na forma de notas históricas e comunicadas aos usuários.

Soergel *et al*. (2004) e Lauser *et al*. (2006) propõem um modelo de reengenharia de tesauro por meio da inserção de relacionamentos mais específicos, e experimentalmente aplicado em recortes do tesauro Agrovoc, da área da agropecuária. A reengenharia desse instrumento foi realizada por meio da separação da modelagem em cinco níveis de entidades: (1) conceito, que representa o significado; (2) termo ou lexicalização, que se refere à representação do conceito; (3) *string* ou variações linguísticas, como acrônimos e abreviaturas; (4) notas de escopo; e (5) relacionamentos. Esse modelo permitiu a explicitação dos relacionamentos em tesauro tradicional para transformá-lo em ontologia. Isso possibilitou estabelecer relações entre as distintas lexicalizações que representam um conceito e, consequentemente, maior semântica à estrutura do tesauro convertido.

Douyère *et al*. (2004) descrevem alguns aprimoramentos no tesauro *Medical Subject Headings* (MeSH), adotado no Catálogo e Índice de Recursos de Saúde em Francês (CISMeF). Foram introduzidos dois novos elementos no tesauro: “tipos de recursos” e “metatermos”. Os primeiros descrevem a natureza do recurso no CISMeF, e correspondem a uma generalização de “tipos de publicação” da base de dados Medline. Já o metatermo é, geralmente, uma especialidade médica ou uma ciência biológica (exemplo: cardiologia ou bacteriologia). Ele foi criado para otimizar a recuperação da informação no CISMeF, a partir da expansão automática de outras árvores temáticas relacionadas ao assunto principal pesquisado. As melhorias possibilitaram o aprimoramento da recuperação da informação no CISMeF e a descoberta de recursos a partir da navegação nas árvores temáticas.

Na área jurídica, o *Projeto do Tesauro da RVBI* (BRASIL, 2004) objetivou reformular o Vocabulário Controlado Básico (VCB), adotado no catálogo coletivo da Rede Virtual de Bibliotecas (RVBI) do Congresso Nacional. Os subsídios para o estudo de novos termos incluíram: indicação do contexto do termo; definição; indicação do registro bibliográfico em que o termo foi utilizado; fontes

consultadas (dicionários, enciclopédias, tesauros) precedidas de informação sobre o resultado da pesquisa; proposta de estruturação do termo (termos genéricos, específicos, relacionados, remissivas, nota explicativa). O agrupamento dos conceitos ocorreu a partir das categorias: objeto, finalidade, processos, produtos, agentes, lugar, princípios, modo ou métodos, fenômenos, instrumentos e tempo. O projeto permitiu a sistematização dos procedimentos de atualização e maior consistência ao tesauro (BRASIL, 2004; VOCABULÁRIO, 2009).

A norma americana ANSI/NISO Z39.19-2005 (R2010) (NATIONAL..., 2010) aborda os aspectos referentes à manutenção de vocabulários controlados monolíngues, e sugere procedimentos para adição, modificação e exclusão de termos. A adição de novos termos ao vocabulário deve ser validada por um comitê de especialistas. Por sua vez, a modificação de termos ou de relacionamentos existentes deve fundamentar-se na documentação e nos critérios de seleção de termos. E os termos subutilizados na indexação devem ser considerados candidatos para modificação ou exclusão, já que há possibilidade de eles serem substituídos por dois ou mais termos de maior especificidade.

Lima *et al.* (2006) descrevem o processo de manutenção e aprimoramento do Vocabulário Controlado do Sistema de Bibliotecas da Universidade de São Paulo (SIBi/USP), implementado no Banco de Dados Bibliográficos (Dedalus)[1]. A atualização do SIBi/USP obedece a um fluxo de trabalho, formado por grupos representativos das 40 bibliotecas integrantes do sistema. Esses grupos detêm as seguintes atribuições: Grupo de Biblioteca: com base nas atividades de indexação e referência, sugere a atualização do vocabulário, por meio de inclusão, modificação ou exclusão de descritores e/ou qualificadores; Grupo de Manutenção: coleta e analisa as solicitações de mudança; Grupo de Coordenação de Conteúdo: valida o descritor e a sua localização na estrutura; Grupo de Coordenação de Processo: executa as decisões e atualiza o vocabulário. A adoção de uma ferramenta colaborativa (SIBIX) para a gestão do

[1] Trata-se do catálogo *online* da rede de bibliotecas da Universidade de São Paulo (USP).

vocabulário em rede otimiza os processos de trabalho, conferindo mais agilidade e visibilidade à atualização.

As publicações *Guia de Atualização do DeCS* (BIREME, 2007) e *Descritores em Ciências da Saúde* (2017) descrevem o processo de atualização do DeCS (Descritores em Ciências da Saúde), que é feito pela criação, modificação e eliminação de termos ou notas. As estratégias de atualização do DeCS seguem as diretrizes para atualização do MeSH e incluem também: estudo da literatura técnico-científica para identificação, coleta e estruturação de novos descritores, que são validados por especialistas em cada tema; alteração das relações, estruturas hierárquicas e categorias; determinação da correspondência do termo no MeSH e nos outros idiomas cobertos pelo instrumento.

No artigo de Cyrot e Preuss (2009) apresenta-se a experiência de reengenharia do tesauro da Agência Francesa de Desenvolvimento (AFD). Foram feitas adições de sinônimos e quase-sinônimos; adaptações semânticas entre termos diferentes que designam o mesmo conceito e adaptação da forma (singular/plural). Os 1.470 termos usados pela AFD foram analisados e comparados de forma intelectual e automática com os descritores ou termos equivalentes do Tesauro UNBIS da Unesco para estabelecer equivalência entre eles. Houve, ainda, um significativo incremento das relações de equivalência no novo tesauro, totalizando 1.713 não descritores. O processo de reengenharia possibilitou a incorporação de mais semântica ao tesauro, por meio do acréscimo e da especificação das relações entre os conceitos.

A norma internacional britânica *British Standards Institution* (BSI) ISO 25964-1 (INTERNATIONAL..., 2011) destaca a natureza das mudanças que podem ocorrer em um tesauro: adição de um novo conceito, representado por um termo preferido; adição de um novo termo não-preferido; alteração de um termo preferido ou não-preferido; exclusão de um termo preferido ou não-preferido; inversão da preferência entre um termo preferido e um termo não-preferido; adição ou modificação de uma nota de escopo; definição de nota histórica ou nota editorial; fusão de dois conceitos em um único conceito; divisão de um conceito em dois ou mais conceitos; alteração na estrutura hierárquica;

movimentação de uma parte da hierarquia de um lugar para o outro na estrutura classificatória; adição ou exclusão de relações associativas; adaptação das modificações anteriores para outras línguas, no caso de tesauros multilíngues.

Caracciolo *et al*. (2012) expõem procedimentos para a atualização do tesauro multilíngue Agrovoc, realizado pela Organização das Nações Unidas para Alimentação e Agricultura (FAO). Esse estudo de caso envolveu a implementação do Agrovoc no modelo de dados *Simple Knowledge Organization System* (SKOS), a referenciação de seus conceitos por meio de URIs e a estruturação dos conceitos no modelo de dados *Resource Description Framework* (RDF). Para a manutenção do tesauro, foi desenvolvida a ferramenta AGROVOC *Concept Server Workbench*, em plataforma web: foram incluídas funcionalidades sofisticadas com suporte para edição distribuída e colaborativa, dada a sua natureza multilíngue, validação de alterações dentro de um fluxo de trabalho editorial formalizado e estabelecimento de funções do usuário e direitos de edição nos diferentes idiomas. Esse experimento possibilitou a integração e a correspondência de pares de conceitos do Agrovoc com outros dez sistemas de organização do conhecimento do domínio da agricultura disponíveis em SKOS-RDF. Após essa primeira implementação, a plataforma evoluiu para um modelo mais geral, renomeado *VocBench*, que suporta um fluxo de trabalho formalizado, por função de usuário e por idioma, além de funcionalidades relacionadas ao rastreamento de alterações, tradução de nomes de elementos e pesquisa entre/dentro de idiomas.

Narukawa *et al*. (2012) descrevem os procedimentos adotados para a revisão do *Vocabulário Controlado da Universidade de São Paulo (USP) da área jurídica*, que foi realizada em duas etapas: 1) identificação e coleta de termos na doutrina e na legislação para cada subárea temática e 2) revisão do vocabulário controlado. A decisão sobre alteração, exclusão ou inclusão de termos considera os critérios: a definição do conceito em dicionários jurídicos, a sua representação no tesauro do Senado Federal, a quantidade e a natureza dos registros indexados na base de dados Dedalus e a análise da doutrina e da legislação. Os resultados parciais indicam a complexidade da linguagem técnica jurídica, em virtude da característica interdisciplinar de seus conceitos e das mudanças

conceituais que acompanham as transformações sociais.

Os espanhóis Baños-Moreno, Pastor-Sánchez e Martínez-Béjar (2013) relatam o trabalho de atualização dos tesauros da UNESCO e da União Europeia (EUROVOC), realizado a partir da extração de termos dos títulos de 320 notícias de divulgação científico-tecnológica publicadas em periódicos de informação geral. As conclusões dos autores indicam a potencialidade das notícias de divulgação científico-tecnológica para a atualização e a inclusão de novos termos nos tesauros e para a redefinição das relações entre eles.

Laipelt (2015) propõe uma metodologia para a seleção de termos equivalentes e descritores em tesauro a partir da utilização dos *logs* de buscas dos usuários do Portal LexML, especializado em informação jurídica e legislativa. Os *logs*, coletados durante um período de 15 dias, possibilitaram a identificação de 2.617 termos e variantes utilizados pelos usuários para a busca de informação relativa ao Direito do Trabalho e ao Direito Previdenciário. A pesquisa trabalhou com um *corpus* textual, composto por parte da jurisprudência e da legislação trabalhista e previdenciária; um *corpus* de referência, constituído pelo Vocabulário Controlado Básico (VCB) do Senado Federal, utilizado na indexação de documentos no portal LexML; e a *corpora* do estudo foi constituído pelas *logs* de pesquisa identificadas no portal. Verificou-se que, no âmbito do Direito do Trabalho e Previdenciário, as variações terminológicas de maior ocorrência foram os recursos não-verbais, especialmente lexias numéricas e alfanuméricas relacionadas ao nome da lei e ao seu respectivo número e data.

Maculan (2015) aplicou o modelo de Soergel *et al*. (2004) e Lauser *et al*. (2006) e propôs uma reformulação do tesauro brasileiro Thesagro, do domínio da agropecuária, no recorte temático da *Intensificação Agropecuária*. O desenvolvimento da remodelagem do Thesagro foi composto de três etapas: 1) definição da estrutura do tesauro; 2) coleta de terminologia e 3) edição do tesauro. A primeira etapa incluiu a análise das características e da estrutura original do Thesagro, assim como a modelagem da estrutura conceitual da temática trabalhada, que foi organizada por um conjunto de nove facetas semânticas. Na segunda etapa, a coleta de terminologia foi realizada a partir de fontes de informação do domínio da agricultura. Na última etapa, a edição do

tesauro foi realizada na ferramenta *e*-Termos, e contou com os seguintes procedimentos: compilação de trechos explicativos ou definitórios sobre o domínio modelado; elaboração de glossário de definições; confecção das fichas terminológicas, determinação das notas de escopo e construção do sistema de conceitos. Os resultados indicaram que a reformulação do Thesagro imprimiu maior nível semântico à estrutura do tesauro e que a explicitação do significado das relações entre termos e conceitos tornou o sistema de relacionamentos mais enriquecido.

Os espanhóis Vállez *et al*. (2015) apresentam um modelo semiautomático de atualização de vocabulários controlados fundamentado na análise de *logs* de busca dos usuários e em um *corpus* de 100 (cem) artigos da revista acadêmica espanhola *BiD Textos Universitaris de Biblioteconomia i Documentación*, indexada no portal *Temaria*. O modelo foi desenvolvido em três etapas: 1) processamento e análise dos dados obtidos a partir das *logs* de consulta, cujas palavras-chave ou frases utilizadas foram comparadas com os descritores utilizados para indexá-lo, a fim de identificar as que permitiram aos usuários acessar os documentos pelos mecanismos de busca; 2) reprocessamento dos dados usando ferramentas linguísticas e estatísticas para isolar exclusivamente aquelas palavras-chave que poderiam ser incluídas para atualizar o vocabulário controlado; e 3) aplicação de uma fórmula de relevância para classificar essas palavras-chave para obter uma lista de termos candidatos mais relevantes quanto a sua importância terminológica. Os resultados da pesquisa indicaram que o modelo foi capaz de processar as consultas para identificar termos candidatos para inclusão em um vocabulário controlado.

No contexto brasileiro, Caribé e Oliveira (2015) relatam a experiência de construção e avaliação do Tesauro de Contas do Tribunal de Contas do Distrito Federal (TCDF). Na atualização do tesauro, a seguinte sistematização é adotada: o termo proposto pelo usuário é encaminhado à Comissão responsável, que criou um formulário próprio para o estudo do termo, o que é feito a partir da investigação da literatura especializada, da legislação, jurisprudência e de outras fontes, como dicionários especializados e outros tesauros. Ao final do processo, o tesauro é atualizado no próprio *e*-TCDF e comunicam-se aos usuários sobre

os resultados do estudo dos termos.

Ribas e Parra (2016) apresentam a experiência de construção e atualização da linguagem de indexação adotada pela rede de bibliotecas universitárias da Universidade Estadual Paulista (Unesp). A atualização do instrumento terminológico incluiu a compatibilização dos registros de autoridade presentes no catálogo das bibliotecas da Unesp com as linguagens da Biblioteca Nacional (BN) brasileira, com a lista de cabeçalhos de assuntos da Biblioteca do Congresso Americano e com os Cabeçalhos de Assuntos Médicos (MeSH). Aplicou-se o método terminográfico para a criação de novos termos, o que incluiu a elaboração de uma ficha terminológica com os dados do termo, com as informações sobre a análise e as pesquisas em outras bases de dados e fontes de informação. A atualização do tesauro foi articulada com a construção de uma política de indexação para as bibliotecas da Unesp.

Su *et al*. (2016) relatam a experiência de reengenharia do MeSH para a melhoria da recuperação da literatura médica sobre células-tronco na base de dados PubMed. Os procedimentos de reengenharia do MeSH incluíram: 1) os termos do MeSH foram filtrados e classificados a partir dos critérios de frequência e especificidade; 2) exclusão de termos amplos; 3) avaliação da terminologia por especialistas; 4) realização de experimento de recuperação da informação, por meio de testes de precisão e revocação. Os resultados indicaram altas taxas de revocação e precisão e a correlação significativa entre os termos do novo tesauro e os resultados da recuperação. Para os autores, a melhoria da especificidade dos termos contribui para a descoberta de novos recursos de informação e a seleção dos termos mais apropriados para a busca.

Mochón Bezares, Méndez Rodríguez e Sorli Rojo (2017) relatam o estudo de atualização do *Tesauro de Biblioteconomía e Documentación*, a partir da comparação dos termos identificados na base de dados ISOC-BD Biblioteconomía e Documentación com as *tags*, em linguagem natural (*folksonomias*), extraídas de *blogs* especializados da área. Foi feito um levantamento de 130 termos na base de dados ISOC-BD e de 186 termos dos 127 *blogs* especializados, selecionados a partir do emprego da tecnologia *web scraping* para a extração automática das *tags* dos *blogs*. A comparação entre as

palavras-chave empregadas nos campos dos descritores da base de dados pelos documentalistas, e as etiquetas atribuídas aos *blogs* pelos próprios usuários demonstra que este último processo aporta uma terminologia mais variada, de maneira mais rápida e atualizada, em relação ao primeiro, o que conferiu às *tags* um importante papel na atualização do Tesauro.

Wu (2018), assim como Soergel *et al*. (2004), apresenta um método para o enriquecimento semântico de tesauros. Ele propõe a especificação das relações associativas, como ferramenta auxiliar na recuperação da informação, tendo em vista que os termos relacionados são os melhores candidatos para uma expansão interativa da busca. O autor aplicou a sua metodologia em uma parte do tesauro *Education Resources Information Center* (ERIC), especificando as relações associativas existentes e expandindo o estabelecimento de mais relações associativas. O processo foi feito com o 1) enriquecimento das relações associativas existentes e com a 2) expansão de relações associativas. Os resultados obtidos pelo autor sugerem que, com base nas percepções dos usuários, um tesauro enriquecido com relações associativas explicitadas e expandidas é uma ferramenta mais útil na recuperação da informação do que um tesauro não enriquecido.

No âmbito dos tribunais de contas, a atualização do Vocabulário de Controle Externo (VCE), pelo Tribunal de Contas da União (TCU) teve início em setembro de 2011, quando foram definidas categorias temáticas para a classificação dos termos, tendo como fonte o recorte temático correspondente na Classificação Decimal Universal (CDU). A atualização do VCE também foi estruturada de acordo com as funções e subfunções de governo definidas pela Portaria n. 42, de 14 de abril de 1999, do Ministério do Orçamento e Gestão. Os descritores de assunto, siglas e nomenclaturas presentes na versão atualizada foram definidos segundo o critério de uso consagrado do termo no TCU. A seleção dos termos preferidos privilegiou, também, a validação anterior já existente no tesauro e a categoria dos descritores classificados como entidades foi construída com base na clientela do Tribunal, nas listas de unidades jurisdicionadas e nas entidades de fiscalização superiores associadas à Organização Internacional de Entidades Fiscalizadoras Superiores (INTOSAI).

O trabalho permitiu a uniformização da terminologia técnica utilizada pela instituição e a padronização do tratamento da informação (BRASIL, 2018).

4 ANÁLISE CRÍTICA DA REVISÃO DE LITERATURA

O mapeamento da literatura sobre revisão e atualização de tesauros identificou 25 (vinte e cinco) trabalhos, assim classificados: 11 artigos; 3 manuais de apresentação de tesauros; 2 teses; 1 capítulo de livro; 2 livros; 2 normas internacionais e 4 trabalhos apresentados em eventos científicos. Esses trabalhos distribuem-se, temporalmente, conforme indicado na Figura 1.

Figura 1 - Quantidade de trabalhos por período

Fonte: Dados da pesquisa (2019)

O gráfico indica que os primeiros estudos sobre revisão e atualização de tesauros foram publicados na década de 1970, período no qual, acredita-se, a adoção dos tesauros em bibliotecas e centros de documentação foi mais difundido. Observa-se também uma maior concentração de trabalhos nos anos de 2004 (3 trabalhos) e 2015 (4 trabalhos), sem que tenha sido identificado na literatura algum movimento que pudesse explicar tal fato.

Retomando-se o objetivo da revisão de literatura de mapear os principais métodos, técnicas e resultados experimentais identificados na revisão e na atualização de tesauros, e considerando-se as perguntas a serem respondidas por esse trabalho (*quais foram as melhorias implementadas na revisão e na atualização de tesauros? Quais os procedimentos que foram adotados? Quais*

resultados foram obtidos nos estudos experimentais?), apresentam-se as análises expostas a seguir.

No Quadro 1 apresenta-se uma síntese dos principais aspectos metodológicos verificados nos trabalhos sobre revisão e atualização de tesauros.

Quadro 1 – Síntese dos aspectos metodológicos na revisão e atualização de tesauros

Métodos para a revisão e a atualização de tesauros	Autores
Adição, modificação e exclusão de termos	KIM (1973) SOERGEL (1974) BATTY (1989) ANSI/NISO Z39.19-2005 (NATIONAL ..., 2010 LIMA *et al*. (2006) BIBLIOTECA... (2017) ISO 25964-1 (INTERNATIONAL..., 2011) NARUKAWA *et al*. (2012)
Revisão dos relacionamentos e hierarquias	INSTITUTO… (1984) NARUKAWA *et al*. (2012)
Registro das mudanças quanto ao escopo e uso em notas históricas	BATTY (1989) ANSI/NISO Z39.19-2005 (NATIONAL..., 2010) ISO 25964-1 (INTERNATIONAL... 2011)
Especificação das relações associativas	SOERGEL (2004) LAUSER et al. (2006) CYROT; PREUSS (2009) MACULAN (2015) WU (2018)
Especificação das relações de equivalência	SOERGEL (2004) LAUSER *et al*. (2006) MACULAN (2015)
Especificação das relações hierárquicas	SOERGEL (2004) LAUSER *et al*. (2006) MACULAN (2015)
Incremento das relações de equivalência	SOERGEL (1974) DOUYÈRE *et al*. (2004) CYROT; PREUSS (2009)
Garantia literária	ANSI/NISO Z39.19-2005 (NATIONAL..., 2010) BIBLIOTECA... (2017) NARUKAWA *et al*. (2012) CARIBÉ; OLIVEIRA (2015) MACULAN (2015) SU *et al*. (2016)
Ficha terminológica	INSTITUTO... (1984) BRASIL (2004) BIBLIOTECA... (2017) NARUKAWA *et al*. (2012) MACULAN (2015) RIBAS; PARRA (2016)

Métodos para a revisão e a atualização de tesauros	Autores
Categorias para o agrupamento dos conceitos	BRASIL (2004) MACULAN (2015) BRASIL (2018)
Gerenciamento e controle terminológico	BATTY (1989) LIMA *et al*. (2006)
Formação da equipe de trabalho	BATTY (1989) BRASIL (2004) ANSI/NISO Z39.19-2005 (NATIONAL..., 2010) LIMA *et al*. (2006) ISO 25964-1 (INTERNATIONAL... 2011) NARUKAWA *et al*. (2012) MACULAN (2015) RIBAS; PARRA (2016)
Atualização da indexação na base de dados	SOERGEL (1974) BRASIL (2004) ANSI/NISO Z39.19-2005 (NATIONAL..., 2010) CYROT; PREUSS (2009)
Registro da memória do grupo de trabalho	BRASIL (2004)
Identificação e coleta de termos, e revisão do vocabulário controlado	NARUKAWA *et al*. (2012) MACULAN (2015)
Adoção de *folksonomias*	MOCHÓN BEZARES; MÉNDEZ RODRÍGUEZ; SORLI ROJO (2017)
Logs de busca dos usuários	ISO 25964-1 (INTERNATIONAL... 2011) LAIPELT (2015) VÁLLEZ *et al*. (2015)
SKOS-RDF e tecnologias da *web* semântica	CARACCIOLO *et al*. (2012)
Abordagem etimológica e teoria do campo semântico	KIM (1973)
Frequência de uso dos termos	IBICT (1984) BRASIL (2004) BIBLIOTECA... (2017) ISO 25964-1 (INTERNATIONAL... 2011) NARUKAWA *et al*. (2012) SU *et al*. (2016)
Monitoramento terminólogico (*Google Trends*/*Google Analytics*)	VÁLLEZ *et al*. (2015)
Extração de termos de notícias de divulgação técnico-científica	BAÑOS-MORENO; PASTOR-SÁNCHEZ; MARTÍNEZ-BÉJAR (2013)
Compatibilização de linguagens	BIBLIOTECA... (2017) CYROT; PREUSS (2009) RIBAS; PARRA (2016)
Validação dos termos pelo especialista do domínio	BATTY (1989) BIBLIOTECA… (2017) NARUKAWA *et al*. (2012) MACULAN (2015) SU *et al*. (2016)
Teste do tesauro pelo usuário final	WU (2018)

Fonte: Dados da pesquisa (2019).

Pela análise do Quadro 1, observa-se que a literatura apresenta uma diversidade de métodos para a revisão e a atualização de tesauros, perpassando pelas técnicas mais tradicionais, mas não menos importantes, de modelagem conceitual intelectual (KIM, 1973; SOERGEL, 1974; INSTITUTO..., 1984; BATTY, 1989; BRASIL, 2004; LIMA *et al*., 2006; NARUKAWA *et al.*, 2012; CARIBÉ; OLIVEIRA, 2015; RIBAS; PARRA, 2016; BRASIL, 2018), até os aspectos mais atuais, relativos ao emprego das tecnologias da *web* semântica (SOERGEL *et al.,* 2004; CARACCIOLO *et al*., 2012; MACULAN, 2015; WU, 2018), as quais possibilitam a interligação de dados e recursos e o refinamento e a explicitação das relações semânticas entre os conceitos e termos.

A título de uma categorização dos trabalhos identificados, sugere-se que os métodos relativos à revisão e à atualização de tesauros possam ser subdivididos em três grupos, não mutuamente exclusivos, de acordo com o enfoque principal dos estudos: 1) *aspectos terminológicos e conceituais gerais* (KIM, 1973; SOERGEL, 1974; INSTITUTO..., 1984; BATTY, 1989; ANSI/NISO Z39.19-2005 - R2010 (NATIONAL..., 2010); ISO 25964-1 (INTERNATIONAL..., 2011); 2) *aspectos gerenciais* (LIMA *et al*., 2006; CARIBÉ; OLIVEIRA, 2015) e 3) *estudo de caso de revisão de tesauro, com o emprego de diferentes técnicas* (SOERGEL *et al.,* 2004; BRASIL, 2004; DOUYÈRE *et al*., 2004; BIBLIOTECA... , 2017; CYROT; PREUSS, 2009; CARACCIOLO *et al*., 2012; NARUKAWA *et al*., 2012; BAÑOS-MORENO; PASTOR-SÁNCHEZ; MARTÍNEZ-BÉJAR, 2013; LAIPELT, 2015; MACULAN, 2015; VÁLLEZ *et al*., 2015; CARIBÉ; OLIVEIRA, 2015; RIBAS; PARRA, 2016; SU *et al*., 2016; MOCHÓN BEZARES; MÉNDEZ RODRÍGUEZ; SORLI ROJO, 2017; WU, 2018 e BRASIL, 2018). O primeiro aspecto inclui as atividades referentes à manipulação e ao desenvolvimento da terminologia, o segundo inclui elementos mais amplos sobre a logística dos processos de revisão e atualização e o último exemplifica alguns tipos de técnicas que foram adotadas nos estudos de caso dos tesauros atualizados.

Observa-se, ademais, que a revisão e a atualização de tesauros perpassam pelos macroprocessos de gestão e de controle terminológico (BATTY, 1989; LIMA *et al*., 2006), e pelos seguintes procedimentos: adição, modificação e exclusão de termos, de forma conjunta com a revisão de seus

relacionamentos e hierarquias. Além disso, a literatura ressalta a importante função do registro, em notas históricas, das modificações quanto ao escopo ou uso dos termos. Esses aspectos são recorrentes em vários trabalhos (KIM, 1973; SOERGEL, 1974; INSTITUTO..., 1984; BATTY, 1989; LIMA *et al*., 2006; BIBLIOTECA..., 2017; NARUKAWA *et al*., 2012) e também nas duas normas internacionais sobre tesauros (ANSI/NISO 2005; ISO 25964-1 (INTERNATIONAL..., 2011)).

Por sua vez, diversos autores, inclusive institucionais (INSTITUTO..., 1984; BRASIL, 2004; BIBLIOTECA..., 2007; NARUKAWA *et al*., 2012; MACULAN, 2015; RIBAS; PARRA, 2016), ressaltam a importância da ficha terminológica para o registro da memória do termo. É importante destacar que as modificações na terminologia, em especial no contexto jurídico, fundamentam-se na garantia literária, a partir do estudo da legislação e da doutrina (BRASIL, 2004; NARUKAWA *et al*., 2012; LAIPELT, 2015; BRASIL, 2018) e também na garantia do usuário (BRASIL, 2004; LAIPELT, 2015; BRASIL, 2018).

Em relação aos aspectos semânticos, verifica-se que as relações associativas (SOERGEL, 2004; CYROT; PREUSS, 2009; MACULAN, 2015; WU, 2018) e de equivalência (SOERGEL 1974; DOUYÈRE *et al*., 2004; CYROT; PREUSS, 2009; MACULAN, 2015) são priorizadas no processo de revisão e atualização dos tesauros. Segundo os autores, isso acontece porque essas relações apresentam potencial de expansão interativa da busca pelo usuário. As relações associativas proporcionam um contexto semântico para a pesquisa, apresentando ao usuário termos relacionados ao assunto, e as relações de equivalências incrementam a rede de remissivas e de sinônimos que podem ser empregados na busca. Destaca-se que as relações associativas são as mais complexas de se determinar, o que implica na necessidade de maiores estudos sobre a sua modelagem.

Mochón Bezares, Méndez Rodríguez e Sorli Rojo (2017) enfatizam o papel dos usuários na etiquetagem de recursos digitais, fenômeno conhecido como *folksonomias*. De acordo com esses autores, a revisão e a atualização de vocabulários controlados é enriquecida com a verificação dos termos que os

usuários utilizam quando descrevem recursos na rede.

O mapeamento e estudo dos *logs* de pesquisa dos usuários na busca nos sistemas de recuperação da informação também subsidia o processo de revisão e atualização de tesauros (ISO 25964-1, (INTERNATIONAL..., 2011); LAIPELT, 2015; VÁLLEZ *et al*., 2015). Essa técnica permite a identificação das variantes linguísticas para um mesmo conceito e do perfil e comportamento de busca do usuário, o que melhora o processo de indexação e de manutenção dos tesauros. Haja vista a crescente demanda por informação digital e o fato de o atendimento presencial nas unidades de informação ser cada vez menor, essa técnica permite, também, o diagnóstico das demandas de informação dos usuários.

Outros elementos importantes no processo de revisão e atualização de tesauros incluem, conforme os trabalhos descritos, a categorização para o agrupamento dos conceitos. A categorização adotada no Vocabulário de Controle Externo, pelo Tribunal de Contas da União (TCU), seguiu a classificação das funções e subfunções do governo federal definidas pela Portaria n. 42, de 14 de abril de 1999, do Ministério do Orçamento e Gestão. Já no Vocabulário Controlado Básico (VCB), do Senado Federal, as categorias para o agrupamento dos conceitos foram definidas *a priori* em: objeto, finalidade, processos, produtos, agentes, lugar, princípios, métodos, fenômenos, instrumentos e tempo.

Entende-se que o trabalho de Wu (2018) se diferencia dos demais por apresentar o processo de revisão do tesauro – realizado a partir do enriquecimento das relações associativas entre os termos, com a sua especificação e expansão – combinado a um estudo de validação, pelo usuário, da utilidade do instrumento na recuperação da informação.

Por fim, identificam-se aspectos da revisão e da atualização de tesauros que também estão presentes no processo de construção desses instrumentos, quais sejam: identificação e coleta de termos, e revisão do vocabulário controlado; formação de uma equipe de trabalho composta por bibliotecários, usuários e especialista no domínio de assunto coberto pelo tesauro; validação da terminologia pelo especialista no domínio; diagnóstico da frequência de uso do termo na indexação; revisão dos assuntos nas bases de dados no caso de

alteração do termo e teste do tesauro pelo usuário final.

5 CONSIDERAÇÕES FINAIS

Os tesauros continuam desempenhando importante função na gestão, representação e recuperação da informação em contextos diversificados nos quais seja necessário o estabelecimento de uma linguagem uniforme para a comunicação da informação. Entretanto, ainda que haja essa reconhecida importância, essas ferramentas carecem da definição de procedimentos sistemáticos de revisão e atualização de sua terminologia. Tendo em vista esse problema, este trabalho de revisão objetivou identificar os principais métodos, técnicas e resultados dos estudos experimentais que possam subsidiar o processo de revisão e atualização de tesauros.

Observou-se, a partir da década de 1970, a existência de métodos de atualização de tesauros de diferenciados níveis de escopo e complexidade, que perpassam por procedimentos de modelagem intelectual, evoluindo para o emprego de tecnologias da *web* semântica. A partir dos relatos identificados, nota-se que a revisão e a atualização de tesauros envolvem três aspectos básicos: 1) aspectos terminológicos e conceituais gerais; 2) aspectos gerenciais e 3) estudos empíricos com o emprego de diferentes técnicas de atualização.

No desenvolvimento do tesauro, há de destacar, porém, a necessidade de se avaliar o contexto de uso, as necessidades de informação dos usuários e os propósitos da revisão para que os procedimentos adotados possam ser efetivos.

REFERÊNCIAS

BAÑOS-MORENO, M-J.; PASTOR-SÁNCHEZ, J-A.; MARTÍNEZ-BÉJAR, R. Propuesta de actualización de macro-tesauros a partir de noticias de divulgación científico-tecnológica. *In:* CONGRESSO ISKO ESPANHA E PORTUGAL, 1., 2013; CONGRESSO ISKO ESPANÃ, 11, Portugal. **Atas** [...]. Porto: Faculdade de Letras da Universidade do Porto, 2013. p. 99-112. Disponível em: http://eprints.rclis.org/20684/. Acesso em: 20 nov. 2018.

BATTY, D. Thesaurus: construction and maintenance: a survival kit. **Database**, v. 12, n. 1, p. 13-20, fev. 1989.

BIBLIOTECA VIRTUAL EM SAÚDE. DeCS: Descritores em Ciências da Saúde. São Paulo: BIREME / OPAS / OMS, 2017. Disponível em: http://decs.bvsalud.org. Acesso em 14 de nov. 2018.

BRASIL. **Projeto do tesauro da RVBI**. Brasília: Senado Federal, 2004. Disponível em:https://www2.senado.leg.br/bdsf/bitstream/handle/id/70375/rvbi_projeto_vcb.pdf?sequence=1. Acesso em: 18 set. 2018.

BRASIL. **VCE**: Vocabulário de Controle Externo do Tribunal de Contas da União. Brasília: Tribunal de Contas da União, 2018. Disponível em: https://portal.tcu.gov.br/vocabulario-de-controle-externo/. Acesso em: 16 nov. 2018.

CARACCIOLO, C.; MORSHED, A.; STELLATO, A.; JOHANNSEN, G.; JAQUES, Y.; KEIZER, J. Thesaurus maintenance, alignment and publication as linked data: the Agrovoc use case. *In*: METADATA AND SEMANTICS RESEARCH CONFERENCE, 6., 2012, Cádiz, Spain, **Anais** [...]. Cádiz, Spain: Springer, 2012. Disponível em: http://eprints.rclis.org/17734/. Acesso: 5 mar. 2019.

CARIBÉ, R. de C. do V.; OLIVEIRA, E. S. de. Avaliação do Tesauro do Tribunal de Contas do Distrito Federal a partir dos usuários. **Rev. Digit. Bibliotecon. Cienc. Inf.**, Campinas, v. 13, n.1, p. 103-119, jan./abr. 2015. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1583. Acesso em: 14 nov. 2018.

CIROT, C.; PREUSS, C. Réingénierie de thesaurus: une étude de cas. **Documentaliste-Sciences de l'Information,** Paris, v. 46, n. 3, p. 4-13, 2009. Disponível em: https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2009-3-page-4.htm. Acesso em: 22 nov. 2018.

COMO fazer revisão sistemática. Produção LARP MAE/USP. São Paulo: LARP MAE/USP, 2015. Vídeo de YouTube (2h04min). Palestrante Fátima de Loudes dos Santos Nunes. Disponível em: https://www.youtube.com/watch?v=Wgaw97mTKWM. Acesso em: 6 dez. 2018.

DOUYÈRE, M.; SOUALMIA, L. F.; NÉVÉOL, A.; ROGOZAN, A.; DAHAMNA, B.; LEROY, J-P.; THIRION, B.; DARMONI, S. J. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. **Health Information and Libraries Journal**, v. 21, n. 4, p. 253–261, 2004.

BIREME. GUIA de atualização do DeCS. São Paulo: BIREME / OPAS / OMS, 2007. 30p. Disponível em: http://red.bvsalud.org/decs/wp-content/uploads/sites/4/2016/11/DeCS-GuiaAtualizacao.pdf. Acesso em 14 de nov. 2018.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Diretrizes para elaboração de tesauros monolíngues**. Brasília: IBICT, 1984.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964-1:2011**: thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval. Geneve: International Standard Organization, 2011.

KIM, C. Theoretical foundations of thesaurus construction and some methodological considerations for thesaurus updating. **Journal of the American Society for Information Science**, v. 24, n. 2, p. 148-156, mar./abr. 1973.

KIM, C.; KIM, S. D. Consensus vs frequency: an empirical investigation of theories for identifying descriptors in designing retrieval thesauri. **Information Processing Management**, London, v. 13, n. 4, p. 253-258, 1977.

LAIPELT, R. do C. F. **Metodologia para seleção de termos equivalentes e descritores de tesauros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. 213 f. Tese (Doutorado em Linguística Aplicada) – Escola da Indústria Criativa: Comunicação, Design e Linguagens, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, 2015.

LAUSER, B.; SINI, M.; LIANG, A.; KEIZER, J.; KATZ, S. From AGROVOC to the Agricultural Ontology Service: Concept Server an OWL model for creating ontologies in the agricultural domain. *In:* INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. **Proceedings**… México: DCMI, 2006.

LIMA, V. M. A.; KOBASHI, N. Y.; SANTOS, C. A. C. M. dos; AMARAL, M. C.; TOKAREVICZ, S.; TORRE, S. R. S. D.; GUERRA, S. R. Y.; BOCCATO, V. R. C.; BARCELLOS, J. C. H. Estudos para implantação de ferramenta de apoio à gestão de linguagens documentárias: vocabulário controlado da USP. **Transinformação**, Campinhas, v. 18, n. 1, p. 17-25, jan./abr. 2006.

MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesauro**: remodelagem do THESAGRO. 2015. 339f. Tese (doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação. Disponível em: http://hdl.handle.net/1843/BUBD-9ZKMUV. Acesso em: 23 maio 2018.

MACULAN, C. M. S.; LIMA, G. Â. B. O. Buscando uma definição para o conceito de conceito. **Perspectivas em Ciência da Informação**, v. 22, n. 2, p. 54-87, abr./jun. 2017.

MOCHÓN BEZARES, G.; MÉNDEZ RODRÍGUEZ, E.; SORLI ROJO, A. Etiquetado social y blog-scraping como alternativa para la actualización de vocabularios controlados: aplicación práctica a un tesauro de Biblioteconomía y Documentación. **Información, cultura y sociedade**, La Rioja [España], n. 37, p. 13-26, dez. 2017. Disponível em: https://dialnet.unirioja.es/servlet/articulo?codigo=6560341. Acesso em: 13 out. 2018.

NARUKAWA, C. M.; PALÁCIO, F. G. L.; MORAES, M. I. de; SILVA, F. M. da; SILVA, M. dos R. da; SAMPAIO, E. A.; MATOS, R. L. de. Metodologia de trabalho para atualização do Vocabulário Controlado da Universidade de São Paulo (USP) da área jurídica. *In:* SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado (RS). **Anais** [...]. Gramado: UFRGS, 2012. Disponível em: https://www.researchgate.net/publication/281118219_Metodologia_de_trabalho_para_atualizacao_do_vocabulario_controlado_da_Universidade_de_Sao_Paulo_USP_da_area_juridica. Acesso em: 19 nov. 2018.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**: ANSI/NISO Z39.19-2005 (R2010). Baltimore, Maryland, U.S.A., 2010.

RIBAS, R. R. de B.; PARRA, L. S. A construção e atualização da linguagem Unesp. *In:* FUJITA, M. S. L. (Org.). **Política de indexação para bibliotecas**: elaboração, avaliação e implantação. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. p. 67-87.

SIMÕES, M. da G. **Da abstração à complexidade formal**: relações conceptuais num tesauro. Coimbra: Almedina, 2008.

SOERGEL, D.; LAUSER, B.; LIANG, A.; FISSEHA, F.; KEIZER, J.; KATZ, S. Reengineering thesauri for new applications: the AGROVOC example. **Journal of Digital Information**, v. 4, n. 4, 2004. Disponível em: https://journals.tdl.org/jodi/index.php/jodi/article/view/112. Acesso em: 15 ago. 2018.

SOERGEL, D. **Indexing languages and thesauri**: construction and maintenance. Los Angeles, CA: Wiley, 1974.

SOUZA, S. T. de. **Modelagem de domínios em Sistemas de Organização do Conhecimento (SOC)**: uma investigação em tesauros e ontologias para a informação legislativa. 2017. 326 f. Tese (doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2017. Disponível em: http://hdl.handle.net/1843/ECIP-ANJHH3. Acesso em: 24 maio 2018.

SU, Y.; ANDREWS, J.; HUANG, H.; WANG, Y.; KONG, L.; CANNON, P.; XU, P. Reengineering of MeSH thesauri for term selection to optimize literature retrieval and knowledge reconstruction in support of stem cell research. **BMC Medical Informatics and Decision Making**, v. 16, n. 54, p. 1-10, 2016. Disponível em: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-016-0298-z. Acesso em: 14 nov. 2018.

VÁLLEZ, M.; PEDRAZA-JIMENEZ, R.; CODINA, L.; BLANCO, S.; ROVIRA, C. Updating controlled vocabularies by analysing query logs. **Online Information Review**, v. 39, n. 7, p. 870-884, 2015.

VOCABULÁRIO Controlado Básico (VCB) na Rede Virtual de Bibliotecas – Congresso Nacional – RVBI. Brasília: Senado Federal, 16 abril 2009. 45 slides. Palestra na 2ª Reunião Ordinária do Comitê Gestor do LeXML. Palestrante Stelina Maria Pinha e Maria Fátima Jaegger. Disponível em: http://www2.senado.leg.br/bdsf/bitstream/handle/id/158959/VCB%20LexML%202009.pdf?sequence=4. Acesso em: 18 set. 2018.

WU, Y. Enriching a thesaurus as a better question-answering tool and information retrieval aid. **Journal of Information Science**, v. 44, n. 4, p. 512-525, 2018.

METHODOLOGIES FOR REVIEWING AND UPDATING THESAURUS: LITERATURE MAPPING

ABSTRACT
Introduction: Thesauri play an important role in terminological standardization to facilitate the communication of information. Given this importance, instruments are not reviewed and updated using consistent methodological approaches. **Objective:** The aim of the study was to identify the main methodologies used for thesauri reviewing and updating, using literature mapping, as support for the development of more consistent conceptual structures. **Method**: To carry out the mapping process, the systematic review protocol proposed by COMO… (2015) was used. The searches were conducted in main national and international databases in the field of Information Science. **Results**: The results indicate different approaches used for reviewing and updating thesauri, such as the procedures for intellectual conceptual modeling, study of users' search logs, adoption of folksonomies, and semantic web technologies. Based on the reports identified, we observed that three basic aspects are considered when reviewing and updating thesauri: 1) general terminological and conceptual aspects; 2) managing aspects; and 3) empirical studies using different updating techniques. **Conclusion**: Although a multiplicity of approaches and techniques for thesaurus updating were identified, the assessment of user context, the information needs of the user, and review purposes for effective procedures should be given more attention.

Descriptors/Keywords: Thesaurus. Reviewing. Updating. Methodology.

METODOLOGÍAS PARA REVISIÓN Y ACTUALIZACIÓN DE TESAUROS: MAPEO DE LA LITERATURA

RESUMEN
Introducción: Los tesauros siguen desempeñando un importante papel en la estandarización de la terminología para facilitar la comunicación de la información. Dada esta reconocida importancia, tales instrumentos no son revisados y actualizados según

fundamentos metodológicos claros. **Objetivo:** En vista de este problema, se objetivó diagnosticar, a través del mapeo de la literatura, las principales metodologías empleadas para la revisión y la actualización de tesauros como un subsidio al desarrollo de estructuras conceptuales más consistentes. **Metodología:** Para la realización de este mapeo, se utilizó el protocolo de revisión sistemática propuesto por COMO... (2015). Se realizaron las búsquedas en las principales bases de datos del área de Ciencia de la información en el contexto nacional e internacional. **Resultados:** Los resultados indican diferentes enfoques empleados en la revisión y actualización de tesauros, a ejemplo de procedimientos para el modelado conceptual intelectual, estudios de *logs* de investigación de los usuarios, la adopción de *folsksonomias* y el empleo de tecnologías de la web semántica. Desde los informes identificados, se observó que la revisión y la actualización de los tesauros implican tres aspectos fundamentales: 1) aspectos terminológicos y conceptuales generales; 2) aspectos administrativos y 3) estudios empíricos con el empleo de diferentes técnicas de actualización. **Consideraciones finales:** Aunque se identifique una multiplicidad de enfoques y técnicas para la actualización de tesauros, es necesario considerar la evaluación del contexto de uso, las necesidades de información de los usuarios y los propósitos de la revisión para que los procedimientos adoptados sean efectivos.

Descriptores: Tesauro. Revisión. Actualización. Metodología.

Recebido em: 28/03/2019

Aceito em: 11/02/2020