

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT- 8 – Informação e Tecnologia

FERRAMENTAS COLABORATIVAS PARA A CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS: EVIDÊNCIAS NAS PUBLICAÇÕES CIENTÍFICAS NACIONAIS E INTERNACIONAIS

Everton Rodrigues Barbosa - Universidade Federal de Santa Catarina (UFSC)
Angel Freddy Godoy Viera - Universidade Federal de Santa Catarina (UFSC)

COLLABORATIVE TOOLS FOR CONTROLLED VOCABULARY CONSTRUCTION: EVIDENCES ON NATIONAL AND INTERNATIONAL PUBLICATIONS

Modalidade da Apresentação: Pôster

Resumo: Aborda as ferramentas colaborativas para a construção de vocabulários controlados nas publicações científicas nacionais e internacionais. Objetiva identificar técnicas e ferramentas utilizadas no processo de gestão terminológica. Apresenta a importância do uso de vocabulários controlados para a efetividade das buscas nos sistemas de recuperação da informação. A pesquisa é considerada bibliográfica, de caráter exploratório, com análise qualitativa baseado no método de revisão integrativa. Verifica que a grande maioria das ferramentas colaborativas utilizadas para a gestão dos vocabulários controlados são semiautomáticas, e se destinam especificamente à gestão de tesouros e ontologias. Constatou-se um forte direcionamento ao uso de ferramentas colaborativas *web*.

Palavras-Chave: Vocabulários controlados; Ferramentas colaborativas; Recuperação da informação.

Abstract: It presents an overview of collaborative tools for controlled vocabularies construction in national and international scientific publications. It aims to identify technics and tools for terminological maintenance process. It aims to discuss the importance of use indexing languages for the searches effectiveness in information retrieval systems. It is a bibliographic and exploratory research, with a qualitative analysis for an integrative review. It verifies that it is a large majority collaborative tools for controlled vocabularies construction has a semiautomatic approach, especially in thesaurus and ontology maintenance. It is verified the majority use of web collaborative as a controlled vocabulary construction tool.

Keywords: Controlled vocabulary; Collaborative tools; Information retrieval.

1 INTRODUÇÃO

Muitos são os esforços para otimizar a Recuperação da Informação (RI) em sistemas automatizados. Alguns sistemas utilizam linguagens controladas de indexação, que representam uma das categorias dos Sistemas de Organização do Conhecimento (SOC) para descrever conteúdos e facilitar o acesso à informação de maneira restrita.

Entre as SOC, destacaremos especificamente os vocabulários controlados por serem instrumentos de indexação amplamente conhecidos e mais utilizados nos Sistemas de Recuperação da Informação (SRI), especialmente os cabeçalhos de assunto, os tesauros, as ontologias e taxonomias. Segundo Harpring (2016, p. 37) vocabulário controlado é considerado “[...] um arranjo organizado de palavras e frases usados para indexar e/ou recuperar conteúdo por meio de navegação ou busca. Geralmente ele inclui termos preferidos e variantes, e possui um escopo definido ou descreve um domínio específico.”

De acordo com Hanandeh (2013) a atualização permanente de um vocabulário controlado é necessária e representa um benefício, pois é capaz de tornar os SRI mais efetivos diante das novas áreas e conceitos que surgem com a produção científica constante. Não obstante, Mari Váñez et al. (2015) afirmam que a construção e atualização desses vocabulários podem ser onerosas, devido a necessidade de envolvimento humano, implicando em custos e tempo suficientes para a efetividade de sua elaboração.

Objetivando sanar problemas relacionados a atualização de terminologias, alguns estudos vêm sendo desenvolvidos com o intuito de tornar automática a gestão desses instrumentos de indexação, pela utilização de sistemas de informação e algoritmos capazes de automatizar todo o procedimento. De outra forma, discute-se a possibilidade de o procedimento ser semiautomático, com a participação do especialista humano para fins de sua consolidação. (PETRAKI; KAPETIS; YANNOKOUDAKIS, 2015).

Considerando aspectos de avaliação, controle e atualização permanente dos vocabulários controlados, Santos et al. (2010) reiteram que para manter a gestão terminológica é importante considerar a colaboração de todos os profissionais e atores sociais envolvidos no processo, incluindo as políticas institucionais de indexação e conformidade com as necessidades de informação dos usuários, pois isso pode influenciar de maneira significativa a qualidade da terminologia, e por consequência, a recuperação da informação.

Os pressupostos teóricos levam-nos a formulação do seguinte questionamento:
Quais são as evidências na literatura científica acerca da importância do uso de ferramentas colaborativas para a construção de vocabulários controlados?

Assim, o trabalho tem como objetivo realizar uma revisão integrativa para apresentar as evidências encontradas nas publicações científicas nacionais e internacionais sobre as contribuições das ferramentas eletrônicas colaborativas para a gestão de vocabulários controlados. Esse levantamento é aparte inicial da tese de doutorado que visa discutir sobre a integração e colaboração das equipes de gestão terminológica no processo de construção de linguagens de indexação.

2 FERRAMENTAS COLABORATIVAS

Os sistemas colaborativos são ferramentas de *software* capazes de dar suporte ao trabalho em grupo visando oferecer formas de interação, comunicação, coordenação e cooperação entre os membros de uma equipe de trabalho. As interações podem ocorrer de maneira síncrona ou assíncrona, e os membros podem executar suas tarefas no mesmo ambiente ou em locais distintos. (MURPHY, 2004; CAMARGO; KHOURI; GIAROLA, 2005).

Entre os sistemas colaborativos mais utilizados, especialmente na esfera empresarial estão os sistemas de *groupware* (trabalho em grupo) e *Workflow* (trabalho em fluxo). Esses integram a categoria de sistemas do tipo *Computer Supported Cooperative Work* (CSCW), cuja nomenclatura em português significa o Trabalho Cooperativo Apoiado por Computador. (NICOLACI-DA-COSTA; PIMENTEL, 2011).

Borges e Reis (2016) afirmam que com o desenvolvimento da *web 2.0* ocorreu uma prospecção de ferramentas eletrônicas colaborativas com intenção de promover a cooperação e o compartilhamento de informações entre os usuários da rede. Atualmente, essas ferramentas colaborativas contribuem significativamente para a efetividade dos SRI, como por exemplo, as marcações ou anotações sociais por meio das *tags*, folksonomias, *Wiki* semânticas e a filtragem colaborativa para os sistemas de recomendação.

De acordo com Sarmiento (2002) integrar trabalhadores do conhecimento é um dos requisitos fundamentais para o desenvolvimento das ferramentas colaborativas. O trabalho colaborativo permeia todas as esferas sociais, incluindo as atividades técnicas e científicas que enxergam no trabalho conjunto uma rica possibilidade de minimizar o tempo e reduzir a duplicidade de esforços.

Nessa perspectiva, partiu-se do pressuposto de que o uso dessas ferramentas colaborativas pode contribuir com a manutenção de vocabulários controlados, na medida em que tal atividade exige agilidade ao seu processo, bem como possui um forte direcionamento a integração da sua equipe de gestão.

3 METODOLOGIA

A pesquisa é considerada bibliográfica, de caráter exploratório, pois se trata de um levantamento das publicações nacionais e internacionais para a realização de uma revisão integrativa sobre a temática em questão. As pesquisas foram feitas nas seguintes bases de dados: *Web of Science* (WOS); *Library and Information Science Abstracts* (LISA); e a Base de Dados em Ciência da Informação (BRAPCI). Foram utilizados os termos de busca: *Thesaurus*, “*Controlled vocabulary*”, *Ontology*, *Taxonomy* e “*Collaborative tools*” em inglês. Para a busca em português utilizou-se os termos: *Tesauros*, “*Vocabulários controlados*”, *Ontologias*, *Taxonomias* e “*Ferramentas colaborativas*”.

Com o intuito de ampliar os resultados, foram formuladas várias estratégias de busca, incluindo a utilização de termos simples e compostos, os operadores booleanos AND e OR, e operadores NEAR para a aproximação de termos. O processo de busca seguiu o conjunto de critérios de inclusão e exclusão conforme o quadro 1.

Quadro 1: Critérios de inclusão/exclusão

Critérios de inclusão/Exclusão	
a) Artigos de periódicos revisados por pares	(X) Incluir () Excluir
b) Artigos de revisão	() Incluir (X) Excluir
c) Nos idiomas português, inglês ou espanhol	(X) Incluir () Excluir
d) Publicados entre 2000 a 2017	(X) Incluir () Excluir

Fonte: Dos autores (2017)

Aplicado os critérios supracitados ao processo de busca, foi realizada uma análise preliminar dos trabalhos a partir das leituras dos resumos. Após verificar a adequação dos artigos, procedeu-se a leitura integral dos mesmos.

Foram analisados os seguintes aspectos: a) ano de publicação; b) autores; c) principais referências; d) tipos de linguagens; e) técnicas e ferramentas de apoio a análise dos dados; f) principais ferramentas colaborativas; g) abordagens de automação; h) principais conclusões.

3 PUBLICAÇÕES SOBRE FERRAMENTAS COLABORATIVAS PARA A CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS

As publicações selecionadas – disponíveis em ordem cronológica decrescente no quadro 2 –, revelaram um forte direcionamento ao uso de ferramentas colaborativas *web* na construção de vocabulários controlados. Dessa forma, infere-se que a garantia do usuário se torna um requisito fundamental para a manutenção das terminologias, especialmente na *web* onde ocorre um grande fluxo de dados produzidos livremente pelos usuários da rede.

Quadro 2: Trabalhos analisados

Autor(es)	Título	Periódico	Data de Publicação
Ravagli, Carlo; Pognan, Francois; Marc, Philippe.	<i>OntoBrowser: a collaborative tool for curation of ontologies by subject matter experts</i>	<i>Bioinformatics</i> , v. 33, n. 1, p. 148–149.	2017
Gabriel Júnior, R. F.; Laipelt, R. C. F.	THESA: ferramenta para construção de tesauro semântico aplicado interoperável	<i>Revista P2P e inovação</i> , v. 3, n. 2, p. 124-145.	2017
Mari Vállez et al.	<i>Updating controlled vocabularies by analysing query logs</i>	<i>Online Information Review</i> , v. 39, n. 7, p. 870-888.	2015
Tudorache, Tania et al.	<i>WebProtégé: a collaborative ontology editor and knowledge acquisition tool for the web</i>	<i>Semant Web</i> , v. 4, n. 1, p. 89-99.	2013
Marie-Angélique Laporte, M. A.; Mougenot, I.; Garnier, E.	ThesauForm-Traits: A web based collaborative tool to develop a thesaurus for plant functional diversity research	<i>Ecological Informatics</i> , v. 11, p. 34-44.	2012
Zhitomirsky-Geffet, M. et al.	<i>A generic framework for collaborative multi-perspective ontology acquisition</i>	<i>Online Information Review</i> , v. 34, n. 1, p. 145-159.	2010
Moura, M. A.	Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias	Informação & Sociedade: Estudos , v.19, n.1, p. 59-73, jan./abr.	2009
Sini, M. et al.	<i>The AGROVOC concept server: rationale, goals and usage.</i>	<i>Library Review</i> , v. 57, n. 3, p. 200-212.	2008
Lima, V. M. A. Et al.	Estudos para implantação de ferramenta de apoio à gestão de linguagens documentárias: vocabulário controlado da USP	TransInformação , Campinas, v. 18, n. 1, p. 17-25, jan./abr.	2006
Lykke Nielsen, M. L.	<i>A framework for work task based thesaurus design</i>	<i>Journal of Documentation</i> , v. 57, n. 6, p. 774-797.	2001

Fontes: Dados da pesquisa (2017)

Percebeu-se que as ferramentas colaborativas são destinadas basicamente a construção e manutenção de dois tipos de linguagens específicas: as ontologias e tesouros, onde a grande maioria apresenta uma abordagem semiautomática de construção. Entre as ferramentas colaborativas *web* que mais se destacam no suporte a construção das linguagens

controladas estão as anotações ou marcações sociais (*tagging*). Além dessas, observou-se a utilização de sistemas colaborativos baseados no trabalho em grupo (*groupware*) e trabalho em fluxo (*workflow*).

Quadro 3: Resumo da revisão Integrativa

Itens	Observações
Ano	2001-2017
Autores	Instituição: Universidades (7), Institutos de Pesquisa (5)
	Titulação: Profissionais/graduados (7), mestres (1), doutores (26) e pós-doutores (4)
	Área de formação: Ciência da Informação (17), Ciência da Computação (16), Bioinformática (3), Comunicação social (2)
Principais referências	Fisseha e Liang (2003), Lancaster (1986), ISO 5964 (1985), ISO 2788 (1986); ISO 25.964.1 (2011); ANSI/NISO Z39.19 (2005)
Tipos de linguagens	Tesouros, Ontologias, Cabeçalhos de assuntos
Técnicas e ferramentas de apoio a análise dos dados	Grupo focal, Análise do discurso, Análise de <i>logs</i> , <i>Google Analytics</i> , <i>Google Trends</i>
Principais Ferramentas colaborativas	Marcação social (<i>tagging</i>), Formulário eletrônico, <i>e-mail</i> , <i>Groupware</i> , <i>Workflow</i>
Abordagens de automação	Automática e semiautomática
Principais conclusões	a) agilizar e dar visibilidade ao processo de manutenção dos vocabulários controlados; b) acompanhar o processo de análise e validação de termos sugeridos; c) considerar a perspectiva dos usuários através da mineração dos termos utilizados nas consultas; d) Promover a integração entre os membros da equipe de trabalho.

Fonte: Dos autores (2017)

No campo das metodologias de apoio a coleta e análise dos dados, merece destaque a utilização de análise de *logs* de consultas com o objetivo de verificar se os vocabulários controlados refletem as necessidades informacionais dos usuários. Dessa forma, as palavras-chave formuladas nas interfaces de busca são colocadas como uma alternativa para a atualização das terminologias.

As principais citações encontradas no *corpus* analisado, excluindo as autocitações, são dos autores: Fisseha e Liang (2003), referente a temática de ontologias; Lancaster (1986), utilizado para a fundamentação teórica sobre os tesouros e vocabulários controlados com destaque para a área de recuperação da informação; e as normas ISO 5964:1985, ISO 2788:1986, ISO 25.964.1:2011¹ e ANSI/NISO Z39.19:2005². Essas normas servem de guia para a construção e manutenção de tesouros, incluindo os monolíngues e multilíngues, e a orientação para interoperabilidade com outros tipos de vocabulários controlados.

¹ <https://www.iso.org/home.html>

² <http://www.niso.org/home/>

As técnicas de apoio a coleta e análise de dados denominadas grupo focal e análise do discurso, foram empregadas na verificação das solicitações de pesquisas via *e-mail* feitas pelos usuários. Os conteúdos dos *e-mails* auxiliaram na incorporação de novos termos ao tesouro.

Apesar de não diretamente expressas nas publicações, verificou-se que a maioria das ferramentas utilizadas são assíncronas e integram os colaboradores geograficamente dispersos, especialmente os especialistas de domínio e usuários da informação, sendo que esses últimos quase sempre colaboram com a manutenção das linguagens de maneira implícita, ou seja, sem saber que estão contribuindo.

4 CONSIDERAÇÕES FINAIS

A revisão integrativa possibilitou verificar as evidências do uso de ferramentas colaborativas para a construção de vocabulários controlados, mostrando que o trabalho em equipe é fundamental para a consolidação das linguagens abordadas, em especial os tesouros e as ontologias.

Foi possível identificar as principais técnicas, ferramentas, as referências mais citadas nas publicações, bem como as contribuições dos estudos para a efetividade das linguagens controladas por meio da análise das principais conclusões.

Dessa forma, o levantamento possibilitou identificar os trabalhos com potencial para integrar a revisão narrativa da tese. Como etapa seguinte, pretende-se ampliar esse levantamento aplicando os mesmos critérios de busca em outras bases de dados, a fim de expandir o número de trabalhos referente a temática proposta.

REFERÊNCIAS

BORGES, C. E. L.; REIS, Adriana N. dos. Sistemas colaborativos baseados em *wikimaps*: uma análise sob o ponto de vista do Modelo 3C. IN: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, 22, Porto Alegre, RS. **Anais eletrônicos....** Porto Alegre: Sociedade Brasileira de Computação, 2016. v. 2: Workshops e Sessão de Pôsteres. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wticwebmedia/2016/005.pdf>>. Acesso em: 30 jul. 2017.

CAMARGO, Álvaro A. B. de. KHOURI, L. H. El; GIAROLA, P. C. **O Uso de Sistemas Colaborativos na Gestão de Projetos: Fatores Relevantes para o Sucesso.** Trabalho de Conclusão de Curso. Fundação Instituto de Administração – FIA. 2005.

- FISSEHA, F.; LIANG, A.C. Reengineering AGROVOC to ontologies: steps towards better semantic structure. IN: NKOS WORKSHOP, 31 May 2003, Houston. **Anais eletrônicos...** Houston: Rice University, 2003. Disponível em: <ftp://ftp.fao.org/gi/gil/gilws/aims/publications/presentations/2003_7.pdf>. Acesso em: 01 ago. 2017.
- HANANDEH, E. Building an automatic thesaurus to enhance information retrieval. **International Journal of Computer Science Issues**, v. 10, Issue 1, n. 3, jan. 2013. Disponível: <<https://ijcsi.org/papers/IJCSI-10-1-3-676-686.pdf>>. Acesso em: 17 mar. 2017.
- HARPRING, P. **Introdução aos vocabulários controlados**: terminologia para arte, arquitetura e outras obras culturais. São Paulo: ACAM Portinari, 2016.
- LANCASTER, F.W. **Vocabulary control for information retrieval**. Arlington, Va: Information Resources Press, 1986.
- MARI VÁLLEZ et al. Updating controlled vocabularies by analysing query logs. **Online Information Review**, v. 39, n. 7, p. 870-884, 2015. Disponível: <<http://www.emeraldinsight.com/doi/abs/10.1108/OIR-06-2015-0180>>. Acesso em: 12 mar. 2017.
- MURPHY, E. Recognising and promoting collaboration in a online asynchronous discussion. **British Journal of Educational Technology**, v. 35, n. 4, p. 421-431. 2004. Disponível em: <http://www.ucs.mun.ca/~emurphy/bjet_401.pdf>. Acesso em: 04 maio. 2017.
- NICOLACI-DA-COSTA, A. M.; PIMENTEL, M. Sistemas colaborativos para uma nova sociedade e um novo ser humano. In: FUKS, H., PIMENTEL, M. **Sistemas colaborativos**. 1. ed. Elsevier, 2011. p. 3-15. Disponível em: <<http://www.sciencedirect.com/science/book/9788535246698>>. Acesso em: 31 jul. 2017.
- PETRAKI, E.; KAPETIS, C.; YANNAKOUDAKIS, E. J. Automated thesaurus population and management. **Qualitative and quantitative methods in libraries (QQML)**, v. 4, p.181-189, 2015. Disponível em: <http://www.qqml.net/papers/March_2015_Issue/4119QQML_Journal_2015_Petrakietal_1_81-189.pdf>. Acesso em: 12 mar. 2017.
- SANTOS, C. A. C. M. dos et al. Sistema de gestão para linguagem documentária: metadados e rede colaborativa no vocabulário controlado do SIBI/USP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, XVI, Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: SNBU, 2010. Disponível em: <<http://www.producao.usp.br/handle/BDPI/35732>> . Acesso em: 27 abr. 2017.
- SARMENTO, A. M. T. **Impacto dos sistemas colaborativos nas organizações**: estudo de casos de adoção e utilização de sistemas *workflow*. Dissertação (doutorado). Universidade do Minho. 2002. Disponível em: <[http://recipp.ipp.pt/bitstream/10400.22/3930/1/TD_AnabelaSarmiento_2002%20\(2\).pdf](http://recipp.ipp.pt/bitstream/10400.22/3930/1/TD_AnabelaSarmiento_2002%20(2).pdf)>. Acesso em: 31 jul. 2017.