

**AUTOMATIZAÇÃO DE REENGENHARIA DE TESAURUS:  
a utilização da ferramenta semantizar***AUTOMATION OF THESAURUS REENGINEERING:  
the use of the semantizar tool*

Vinícius Corrêa Braga<sup>1</sup>  
Universidade Federal de Minas Gerais

Gercina Ângela de Lima<sup>2</sup>  
Universidade Federal de Minas Gerais

**RESUMO**

Os sistemas de organização do conhecimento, instrumentos de representação que se configuram em diversas estruturas, tais como taxonomias, tesouros e ontologias, são estruturados com o propósito de facilitar os processos de organização e recuperação da informação. O foco desta pesquisa é, por meio do uso de uma ferramenta, verificar a possibilidade de automatizar o processo de enriquecimento semântico de tesouros. Classificada como uma pesquisa qualitativa, de objetivo exploratório e propósito aplicado iremos analisar, interpretar e comparar os relacionamentos semânticos gerados através da ferramenta com os relacionamentos gerados através do modelo de reengenharia de tesouros. Serão utilizados, como procedimentos, pesquisa bibliográfica e estudo de caso.

**Palavras-chave:** Reengenharia de tesouro; Relações Semânticas; Extração de Relações Semânticas; Semantizar;

**Keywords:** Thesaurus reengineering; Semantic Relations; Extraction of Semantic Relations; Semantizar;

**1 INTRODUÇÃO**

O avanço das tecnologias da informação mudou e continua a mudar, a cada dia, as relações entre os humanos e a forma como resolvemos problemas. Segundo

---

<sup>1</sup> Mestrando pelo Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Orcid: <https://orcid.org/0000-0002-0155-6786>. E-mail: 06correa@gmail.com.

<sup>2</sup> Docente do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Orcid: <https://orcid.org/0000-0003-0735-3856>. E-mail: limagercina@gmail.com.

Marcondes (2016), as tecnologias da informação e padrões sempre foram utilizadas para agregar diferentes acervos e têm sido conhecidos na literatura sob a denominação de “interoperabilidade”. A expansão da internet permitiu conectar informações a distâncias antes imagináveis e, em 1999 o emprego do termo “web semântica”, por Bernes-Lee e Fischetti, permitiu estudos que envolvam a incorporação de metadados aos recursos. Essa incorporação permite que acervos digitais sejam publicados diretamente da web e as informações possam ser reconhecidas por máquinas e interoperáveis entre distintos sistemas.

A Biblioteconomia e Ciência da Informação (BCI) sempre se preocuparam com a questão do controle semântico dos recursos informacionais, através das linguagens documentárias. Para Marcondes (2016) a BCI valoriza em especial as relações semânticas como fundamentais para agregar semântica aos recursos informacionais e tem se dedicado a estudá-las. Em ambientes digitais, as relações são a base para que computadores realizem “inferências”, descobrindo conhecimento não diretamente explicitado.

## **2 FUNDAMENTAÇÃO TEÓRICA**

No âmbito da BCI, conforme Maia (2018); Lima; Maculan (2017), os Sistemas de Organização do Conhecimento (SOC) organizam e representam domínios do conhecimento por meio de instrumentos, tais como taxonomia, sistemas de classificação, tesouros e ontologias para recuperar a informação formando uma base conceitual incorporando a característica de cada SOC. Estes são empregados na representação e estruturação semântica do conhecimento e no controle da terminologia de cada domínio. Assim é possível mapear o conhecimento reunindo assuntos correlatos e evitando a dispersão de termos associados.

A literatura apresenta a evolução e traz normalmente alguns modelos para o enriquecimento semântico de tesouros tradicionais, geralmente para convertê-los em ontologias (CAMPOS et AL, 2008; TEIXEIRA, 2013, SOERGEL et AL. 2004).

Entretanto alguns problemas podem ser percebidos. As conclusões do trabalho de Teixeira (2013) apontam dificuldades no momento da conversão e que necessita de uma supervisão de um especialista em ontologias. Soergel et al. (2004)

aponta limitações como a falta de abstração conceitual e a escassa representação semântica encontrada em tesouros.

Com base no exposto duas lacunas foram percebidas: (1) construir relações semanticamente ricas de maneira a não tornar inviável o trabalho devido ao tempo e complexidade; (2) automatizar a extração de relações semanticamente ricas com a mesma qualidade de um serviço manual ou modelos de conversão.

Alicerçado nas lacunas apresentadas, o seguinte problema foi levantado: é possível utilizar uma ferramenta para automatizar o processo de reengenharia de tesouro com o objetivo de enriquecer semanticamente a estrutura conceitual?

O objetivo geral visa utilizar o sistema de informação na web SEMANTIZAR para automatizar o processo de reengenharia de tesouro enriquecendo semanticamente as estruturas dos conceitos. Mais especificamente pretende: (1) **demonstrar** se o Sistema SEMANTIZAR é capaz de explicitar estruturas conceituais semanticamente ricas; (2) **aperfeiçoar** o sistema SEMANTIZAR e propor soluções que não tenham sido observadas durante o desenvolvimento do sistema e da tese de Maia (2018); (3) **verificar** se o sistema SEMANTIZAR é um meio de computação pragmaticamente eficiente não necessitando de um esforço heroico.

### 3 METODOLOGIA

De acordo com a concepção de Diehl e Tatim (2004) no que diz respeito às classificações de uma pesquisa científica, pode-se caracterizar este estudo da seguinte forma: quanto à abordagem do problema: trata-se de uma pesquisa qualitativa, pois descreve, interpreta e compara, os relacionamentos semânticos gerados através da ferramenta Semantizar com os relacionamentos gerados através do modelo de reengenharia de tesouros desenvolvido por Soergel et al. (2004) e Lauser et al. (2006); quanto ao objetivo geral: este trabalho tem a característica de ser uma pesquisa exploratória, que permite o conhecimento dos assuntos que intercederam a proposta da pesquisa; quanto ao propósito: esta pesquisa está classificada como aplicada com o objetivo de gerar conhecimentos para a aplicação prática.

A delimitação do universo de pesquisa pode ser realizada a partir de diferentes aspectos, conforme apontam Lakatos e Marconi (2010), a partir dos quais foram demarcados os seguintes recortes: quanto ao campo de investigação: o campo de estudo compreende os aportes teóricos de relacionamentos semântico e reengenharia de tesouros; quanto ao recorte da pesquisa: o recorte será na temática da Intensificação Agropecuária; quanto ao universo de pesquisa: o universo de pesquisa utilizado na experimentação são estruturas classificatórias e seus respectivos documentos acadêmicos; quanto a amostra: livro intitulado “Evolução agrária e pressão demográfica<sup>3</sup>”, da autora Ester Boserup (1965), que explica e elucida o processo de Intensificação Agropecuária e a estrutura classificatória da Classificação Decimal Universal (classe 631 - Agricultura em geral).

Quanto aos procedimentos, definiu-se a realização da pesquisa bibliográfica e do estudo de caso. A pesquisa bibliográfica será realizada em dois momentos: no levantamento bibliográfico da fundamentação teórica-metodológica e no levantamento de trabalhos publicados sobre automatização de reengenharia de tesouros para a revisão de literatura. Como forma de aprofundamento da sua abordagem qualitativa, esta pesquisa se caracteriza como um estudo de caso, uma vez que haverá análise da aplicação empírica da ferramenta semantizar. Será verificado se o sistema supracitado contribuirá para automatizar o processo de reengenharia de tesouro enriquecendo semanticamente as estruturas conceituais.

#### **4 RESULTADOS PARCIAIS**

A pesquisa encontra-se em fase embrionária. Os procedimentos para o levantamento bibliográfico serão estruturados, para dar prosseguimento na primeira etapa da metodologia, a revisão de literatura.

---

<sup>3</sup> Versão traduzida do original em inglês “The conditions of agricultural growth: the economics of agrarian change under population pressure” e publicada em 1987 pelos autores Oriowaldo Queda e João Carlos Duarte.

## 5 CONSIDERAÇÕES FINAIS

Os SOCs se propõem a organizar o conhecimento e auxiliar no entendimento de dado domínio. A presente pesquisa utilizará o Sistema SEMANTIZAR para explicitar relações entre conceitos, tendo como insumos o livro *Evolução Agrária e Pressão Demográfica* de Ester Boserup e a estrutura classificatória da CDU. Com isso, espera-se que seja possível observar relações entre conceitos semanticamente ricos. Estes relacionamentos explicitados serão qualitativamente comparados com a estrutura semântica de um tesauro que foi enriquecido semanticamente pelo modelo de reengenharia desenvolvido por Soergel et al. (2004) e Lauser et al. (2006). Com isso, espera-se que a estrutura conceitual seja compreendida não somente por especialistas e avaliar se a ferramenta é pragmaticamente eficiente.

## REFERÊNCIAS

BOSERUP, Ester. **Evolução agrária e pressão demográfica**. São Paulo: HUCITEC: Polis, 1987. 141 p.

CAMPOS, M. L. M. et al. O uso de tesauro como base terminológica para a elaboração de ontologias de domínio: uma experiência com o domínio do Folclore e Cultura Popular. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: USP, 2008. Disponível em: <https://brapci.inf.br/index.php/res/download/176526>. Acesso em: 15 de jun. de 2023.

DIEHL, A. A.; TATIM, D. C. **Pesquisa em ciências sociais aplicadas: métodos e técnicas**. São Paulo: Pearson Prentice Hall, 2004. 168 p.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica: técnicas de pesquisa**. 7. ed. São Paulo: Atlas, 2010

LAUSER, B. et al. 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. Disponível em: <http://eprints.rclis.org/21109/1/Paper11.pdf>. Acesso em: 15 de jun. de 2023.

LIMA, G.A.B.; MACULAN, B. C. M. S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ciência da Informação**, Brasília, DF, v.46 n.1, p.60-72, jan./abr. 2017. Disponível em: <https://brapci.inf.br/index.php/res/download/55121>. Acesso em: 15 de jun. de 2023.

MAIA, Lucinéia Souza. **Extração e explicitação de relações semânticas para representação de documentos acadêmicos**: um estudo de caso a partir de uma estrutura classificatória. 248f. 2018. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/VAFA-BB5JHW>. Acesso em: 15 de jun. de 2023.

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 2, p. 61-83, abr./jun. 2016. Disponível em: <https://www.scielo.br/j/pci/a/8svGtzqw5HZCrfrPJbRypsb/?format=pdf&lang=pt>. Acesso em: 15 de jun. de 2023.

SOERGEL, D. et al. Reengineering thesauri for new applications: the AGROVOC example. **Journal of Digital Information**, v.4, n.4, 2004. Disponível em: <https://www.fao.org/3/af234e/af234e.pdf>. Acesso em: 15 de jun. de 2023.

TEIXEIRA, J. R. **Conversão de tesouros em ontologias [manuscrito]**: um estudo exploratório. 2013. 122f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013. Disponível em: [https://repositorio.ufmg.br/bitstream/1843/BUBD-9R7JYV/1/disserta\\_\\_o\\_joice\\_teixeira\\_2013.pdf](https://repositorio.ufmg.br/bitstream/1843/BUBD-9R7JYV/1/disserta__o_joice_teixeira_2013.pdf). Acesso em: 15 de jun. de 2023.