

## EXTRAÇÃO DE RELAÇÕES SEMÂNTICAS NA CONSTRUÇÃO DE ESTRUTURAS CONCEITUAIS AGROPECUÁRIAS

*Semantic relationship extraction in the construction of  
agricultural conceptual structures*

### 1 INTRODUÇÃO

A Representação do Conhecimento (RC) tem avançado significativamente nos últimos anos, impulsionada pelas Tecnologias da Informação e Comunicação (TIC). Nesse contexto, destaca-se a importância do controle semântico dos recursos informacionais, que pode ser alcançado por meio da construção de Estruturas Conceituais (EC).

Essas estruturas não apenas facilitam a Organização do Conhecimento (OC), como também contribuem para a superar a diversidade terminológica existente, estabelecendo pontes semânticas que otimizam a Recuperação da Informação (RI).

Com o objetivo de explorar os limites e possibilidades do uso de ferramentas computacionais na RC, esta pesquisa propõe responder à seguinte questão: como a extração semiautomática de Relações Semânticas (RS) pode auxiliar na construção de Estruturas Conceituais (EC) aplicadas ao domínio da Intensificação Agropecuária (IA)?

Uma das estratégias fundamentais nesse processo é a Extração de Relações (ER). Ao extrair relações a partir de textos em Linguagem Natural (LN), torna-se possível melhorar o acesso às informações e otimizar os processos de RI.

BELO HORIZONTE - MG  
04 A 06 JUNHO DE 2025

**VI FORPED PPGOC  
UFMG**

**Fórum de Pesquisas  
Discentes do Programa  
de Pós-Graduação em  
Gestão e Organização  
do Conhecimento**

**ISSN: 2965-4068**

**MODALIDADE:  
RESUMO EXPANDIDO**



**Vinicius Braga**

Mestre do Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Universidade Federal de Minas Gerais, Brasil.

<https://orcid.org/0000-0002-0155-6786>

@6correa@gmail.com



**Gercina Ângela de Lima**

Docente do Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Universidade Federal de Minas Gerais, Brasil.

<https://orcid.org/0000-0003-0735-3856>

limagercina@gmail.com



**Lucinéia Sousa Maia**

Docente do Departamento de Computação e Sistemas de Informação, Universidade Federal de Ouro Preto, Minas Gerais, Brasil.

<https://orcid.org/0000-0002-2343-2944>

lucineia@ufop.edu.br



## 2 FUNDAMENTAÇÃO TEÓRICA

Para Mazzocchi e Tiberi (2009), as EC baseiam-se em LN e, por meio de tratamentos semânticos específicos, convertem-se em uma linguagem “artificial” e normalizada. No entanto, Lima (2020) ressalta que, apesar da clareza e precisão oferecidas por essas EC, sua construção ainda enfrenta diversos desafios.

Buscando mitigar esses desafios, foram desenvolvidas técnicas e ferramentas que permitem automatizar a ER, reduzindo a necessidade de intervenção manual, conforme apontam Claro *et al.* (2023). Entre essas ferramentas, destaca-se o Semantizar, um sistema computacional criado por Maia (2018), voltado à extração semiautomática de relações semânticas na língua portuguesa.

## 3 METODOLOGIA

Esta pesquisa caracteriza-se como aplicada, pois visa soluções práticas para problemas específicos. Quanto aos objetivos, adota-se uma abordagem exploratória, com o intuito de aprofundar o conhecimento sobre o tema em questão. Utiliza-se uma metodologia mista, combinando a análise qualitativa das RS identificadas com a quantificação dos resultados para avaliar a precisão e a eficácia do sistema. No que se refere aos procedimentos técnicos, trata-se de um estudo de caso, em que a ferramenta Semantizar foi aplicada empiricamente para avaliar sua capacidade de estabelecer RS no domínio temático da IA.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Para avaliar quantitativamente a tarefa de ER, as *Message Understanding Conferences* (Conferências de Compreensão de Mensagem), estabeleceram, conforme Claro *et al.* (2023), as métricas de precisão e de revocação (*recall*) como indicadores principais do desempenho.

A precisão, que mede a proporção de resultados corretamente classificados como positivos em relação ao total de resultados positivos identificados, é calculada por: Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Positivos). Aplicando-se essa fórmula, o Semantizar identificou corretamente 31 relações dentre os 144 indícios retornados, obtendo uma precisão de 21%.



Já a revocação, que avalia a capacidade do sistema em recuperar corretamente todos os exemplos positivos, é calculada pela fórmula: Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Negativos). O Semantizar identificou 31 relacionamentos corretamente (verdadeiros positivos) e deixou de identificar 52 (falsos negativos), resultando em uma revocação de 37%. A Tabela 1 apresenta uma comparação das métricas obtidas neste estudo com os resultados encontrados na literatura para outras ferramentas.

Tabela 1: Comparativo das métricas da literatura com o Semantizar

Estudos	Precisão	Revocação
Shin; Jung; Yi (2015)	79,21%	70,51%
Tissera; Weerasinghe (2019)	95%	92%
Semantizar - estudo de caso (2024)	21%	37%

Fonte: Elaborada pelo autor, 2024.

Shin, Jung e Yi (2015) utilizaram algoritmos de Aprendizado de Máquina (AM) para classificar palavras, reconhecer entidades e extrair relações, complementados por regras especializadas para refinar os resultados. Já o alto desempenho alcançado por Tissera e Weerasinghe (2019) é atribuído ao uso combinado de AM e ontologia.

Sob uma perspectiva qualitativa, o Semantizar mostrou capacidade para enriquecer e expandir a estrutura de referência inicial, propondo 28 (vinte e oito) novas RS. A Tabela 2 evidencia exemplos da expansão da estrutura.

Tabela 2: Exemplo de relações propostas pelo Semantizar além da estrutura inicial

Conceito 1	Conceito 2	Notação da RS	RS Explicitada	Tipo de RS
Pousio	Solo	afeta	Pousio afeta o solo.	Associativa
Pousio	Pousio curto	influência	O pousio influencia o pousio curto.	Associativa

Fonte: Elaborada pelo autor, 2024.

## 5 CONSIDERAÇÕES FINAIS

Considerando a complexidade inerente na definição de conceitos e relações, bem como o aumento expressivo no volume de informações disponíveis, esta



pesquisa investigou a contribuição da ferramenta Semantizar na construção de EC a partir de fontes textuais em LN, com foco no domínio da IA.

A aplicação da ferramenta possibilitou a ampliação da estrutura semântica inicial com 24 novas RS, enriquecendo as conexões conceituais nesse domínio. A EC resultante mostra-se promissora para subsidiar a criação de vocabulários controlados voltados à organização e recuperação eficiente da informação.

## REFERÊNCIAS

CLARO, D. *et al.* Extração de Informação. *In*: CASELI, E. M.; NUNES, M. G. V. (org.). **Processamento de Linguagem Natural**: Conceitos, técnicas e aplicações em português. São Carlos: BPLN, 2023.

LIMA, G. A. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em ciência da informação**, Belo Horizonte, v. 25, número especial, p. 57-97, 2020.

MAIA, L. S. **Extração e explicitação de relações semânticas para representação de documentos acadêmicos**: um estudo de caso a partir de uma estrutura classificatória. 248 f. 2018. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018.

MAZZOCCHI, F.; TIBERI, M. Knowledge organization in the philosophical domain: dealing with polysemy in thesaurus building. **Knowledge Organization**, Germany, v. 36, n. 2-3, p. 103-112, 2009.

SHIN, S.; JUNG, H.; YI, M. Y. Building a business knowledge base by a supervised learning and rule-based method. **KSII Transactions on Internet and Information Systems (TIIS)**, Republic of Korea, v. 9, n. 1, p. 407-420, 2015.

TISSERA, M.; WEERASINGHE, R. Auto Generation of Gold Standard, Class Labeled Data Set and Ontology Extension Tool [QuadW]. *In*: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTATIONAL AND COMMUNICATION PARADIGMS (ICACCP), 2., 2019, Gangtok, India. **Proceedings** [...] Gangtok, India: IEEE, 2019. p. 1-6.