



24º ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento

A EXTRAÇÃO DE RELAÇÕES SEMÂNTICAS COMO SUPORTE PARA A CONSTRUÇÃO DE ESTRUTURAS CONCEITUAIS

THE EXTRACTION OF SEMANTIC RELATIONSHIPS AS SUPPORT FOR THE CONSTRUCTION OF CONCEPTUAL STRUCTURES

Vinícius Corrêa Braga – Universidade Federal de Minas Gerais (UFMG)
Gercina Ângela de Lima – Universidade Federal de Minas Gerais (UFMG)
Lucinéia Souza Maia – Universidade Federal de Ouro Preto (UFOP)

Modalidade: Trabalho Completo

Resumo: A extração de relações pode ser utilizada na representação do conhecimento para a construção de estruturas conceituais. Nesse contexto, algoritmos são empregados para encontrar relações semânticas entre conceitos representativos de bases textuais. Assim, o Semantizar, uma ferramenta computacional de extração semiautomática de relações semânticas, indica possíveis relações semânticas entre conceitos em determinado contexto. Isso posto, o objetivo deste artigo é aplicar o Semantizar para apoiar a construção de uma estrutura conceitual a partir de uma fonte textual e um conjunto de termos validados por especialistas de um domínio. Na estrutura conceitual resultante, os conceitos são semanticamente relacionados entre si. A abordagem metodológica envolve um estudo de caso, cujos resultados apontam a aplicabilidade do Semantizar para indicar relações semânticas e seu potencial para apoiar os profissionais na elaboração de estruturas conceituais. Portanto, acredita-se que este artigo pode contribuir ao sugerir uma estratégia de apoio à organização e representação do conhecimento.

Palavras-chave: Extração de Relações; Relações Semânticas; Semantizar.

Abstract: The extraction of relations can support the representation of knowledge to build conceptual structures. In this context, algorithms find semantic relationships between concepts representing textual bases. Thus, Semantizar, a computational tool for semi-automatic extraction of semantic relations, indicates possible semantic relations between concepts in a given context. Thus, the objective of this article is to apply Semantizar to support the construction of a conceptual structure from a textual source and a set of terms validated by experts in a domain. In the resulting conceptual structure, concepts are semantically related to each other. The methodological approach involves a case study whose results point to the applicability of Semantizar to indicate semantic relationships and its potential to assist professionals in developing conceptual structures. Therefore, this article can contribute by suggesting a strategy to support knowledge organization and representation.

Keywords: Relation Extraction; Semantic Relations; Semantizar.

1 INTRODUÇÃO

Estruturas conceituais, semanticamente estruturadas e construídas a partir de linguagem natural como, por exemplo, tesouros, ontologias, redes semânticas e grafos do conhecimento, oferecem recursos essenciais para uma variedade de aplicações em uma ampla gama de domínios temáticos. As estruturas conceituais são vistas, de acordo com Bates (1986), como um caminho para superar o “problema de vocabulário”.

Uma das estratégias para apoiar a construção de estruturas conceituais é a Extração de Relações (ER). De acordo com Kumar (2013), a ER analisa se um determinado documento contém ou não uma relação para um par de conceitos. Segundo Pawar, Palshikar e Bhattacharyya (2017, p. 3, tradução nossa)¹ corroboram ao afirmar que a "tarefa de extração de relações consiste em identificar menções às relações de interesse em cada frase dos documentos fornecidos".

Nesse sentido, este artigo utiliza o sistema *Semantizar*, ferramenta computacional de extração semiautomática de relações semânticas em português, que pode auxiliar profissionais na construção de estruturas conceituais para representar o conhecimento não estruturado. O *Semantizar*, desenvolvido por Maia (2018), é uma aplicação para implementar o Modelo de Extração de Relações Semânticas. Nesse sentido, ele utiliza dois insumos como entrada de dados: um conjunto de conceitos e um documento. Este conjunto de conceitos é um artefato resultante da Representação do Conhecimento (RC) de um determinado documento, organizado como uma lista de conceitos. Já o documento pode ser um artigo, um livro, uma tese ou dissertação.

Ao extrair relações semânticas e elaborar estruturas conceituais a partir delas, representa-se um texto da linguagem natural em um arranjo semanticamente rico. Isso permite, segundo Uren *et al.* (2006), o acesso às informações de forma clara e otimiza os processos de recuperação de informações.

¹ "The task of relation extraction (RE) consists of identifying mentions of the relations of interest in each sentence of the given documents". (Pawar, Palshikar e Bhattacharyya, 2017, p. 3).

1.1 Problema

Para Mazzocchi e Tiberi (2009), estruturas conceituais são baseadas em linguagem natural que se transformam mediante certos tratamentos semânticos numa linguagem “artificial” e normalizada em que os conceitos são tornados basicamente monossêmicos, ou seja, apenas um único sentido, e as relações entre eles são explicitadas. Tradicionalmente, estruturas conceituais são desenvolvidas por especialistas de um determinado domínio do conhecimento e, devido à enorme quantidade de dados, os projetos necessitam, inevitavelmente, de muitos recursos como, por exemplo, financeiros e humanos.

Com base no exposto e dado, o esforço para encontrar relações semânticas entre dois conceitos, duas lacunas foram percebidas: a complexidade para a construção de estruturas conceituais e, os recursos utilizados para o desenvolvimento manual das estruturas.

Alicerçado nas lacunas apresentadas, o seguinte problema foi levantado: como o sistema *Semantizar* contribui com os profissionais da informação na construção de estruturas conceituais em fontes textuais em linguagem natural?

1.2 Objetivo

Estudar o potencial do sistema *Semantizar* para a extração de relações semânticas de fontes em linguagem natural para a construção de estruturas conceituais.

1.3 Justificativa

A Extração de Relações (ER), segundo Guarino (1995), pode ser útil para diversas atividades de Recuperação de Informação (RI), resposta a perguntas, resumo de textos, construção de ontologias e geração de tesouros.

Para Green (2001), os relacionamentos envolvidos, sobretudo na Organização do Conhecimento (OC), são numerosos e geralmente complexos. Essa magnitude e complexidade atrapalham seu uso consistente pelos profissionais da informação, assim como pelos usuários finais. Segundo a autora, muitas vezes não existe consenso sobre como tratar certos relacionamentos e os usuários finais não entendem o que está sendo comunicado pelo relacionamento e pelo padrão de notações de relacionamento utilizado.

Hjørland (2007) também afirma que o número de relações semânticas pode ser infinito e que diferentes domínios desenvolvem novos tipos de relações. E mais, segundo o autor, as contribuições sobre relações semânticas são muito diferentes e difíceis de se apresentarem

de forma coerente, porque elas não estão relacionadas umas com as outras ou sistematicamente relacionadas com visões mais amplas.

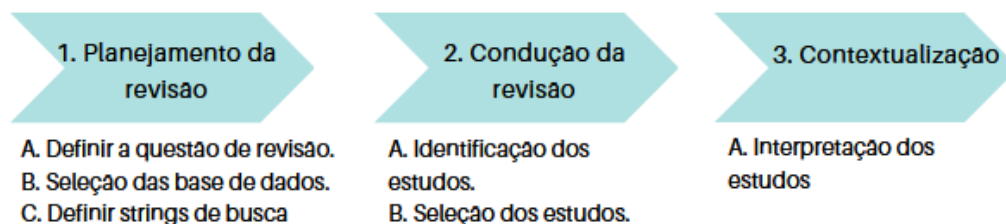
Nesse sentido, a extração de relações é crucial no campo da OC, implicando na eficiência e qualidade. Logo, este estudo desempenha um papel fundamental, entre eles: (1) aponta a extração de relações como um facilitador na organização de grandes volumes de dados. No mundo atual em que a quantidade de informações cresce exponencialmente, é essencial que dados não sejam apenas coletados, mas também organizados de maneira que sejam acessíveis e utilizáveis; (2) sugere que a extração de relações melhora a recuperação da informação, visto que ao estabelecer vínculos entre conceitos, facilita-se a construção de sistemas de recuperação da informação eficientes. Isso é importante, principalmente em ambientes acadêmicos e profissionais, onde a relevância e precisão são fundamentais; (3) demonstra que a extração de relações contribui significativamente para a construção de Sistemas de Organização do Conhecimento (SOC), técnicas importantes para a padronização e interoperabilidade entre sistemas e, por fim; (4) mostra que a extração de relações também é relevante ao suporte, à análise semântica e à compreensão do texto. Isso possibilita a identificação de contextos e significados mais profundos em textos, permitindo uma análise mais rica e detalhada.

Este artigo está organizado da seguinte forma: Nesta seção, realizou-se uma introdução com a intenção de contextualizar o leitor quanto ao teor da problemática da pesquisa. A seguir, expõe-se a metodologia de pesquisa, discorrendo sobre a caracterização da pesquisa com base no problema levantado, assim como no objetivo traçado. Na seção sobre fundamentação teórica, abordam-se os principais conceitos relatados para permitir o entendimento dos assuntos abordados. Em seguida, explana-se a revisão narrativa de literatura realizada para investigar estudos sobre extração das relações. Durante os resultados do estudo de caso, destacam-se as contribuições para os estudos sobre extração de relações semânticas e, por fim, na última seção, faz-se as considerações finais.

2 METODOLOGIA

Para a revisão de literatura, procedeu-se à elaboração de um esquema com o fluxo das três etapas da revisão de literatura, conforme a Figura 1.

Figura 1 - Esquema com as etapas da revisão de literatura



Fonte: Elaborado pelos autores (2024).

Na primeira etapa, definiu-se a questão que conduziu a revisão de literatura (1A): como as relações extraídas de fontes textuais mediante PLN, utilizando metodologia baseada em regras, são organizadas para a construção de EC?

Prosseguindo, selecionou-se as seguintes bases de dados (1B) para levantamento dos estudos a partir do Portal de Periódicos CAPES²: ACM *Digital Library* (ACM), ArXiv.org (ArXiv), Google Acadêmico (GA), Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE), *Scopus* e *Web of Science* (WoS). Elaborou-se as *strings* de busca (1C), conforme apresentado no Quadro 1, com objetivo de identificar e selecionar estudos relevantes.

Quadro 1 – Expressões de busca utilizadas nas bases de dados

Base de Dados	Expressões de busca
ACM	[Abstract: extract*] AND [[Abstract: "unstructured text"] OR [Abstract: "unstructured document"]] AND [[Abstract: taxonom*] OR [Abstract: ontolog*] OR [Abstract: thesaur*]]
ArXiv	abstract=extract*; AND abstract="unstructured text" OR "unstructured document"; AND abstract=Taxonom* OR Ontolog* OR Thesaur*
GA	allintitle: (Unstructured) (Relation OR Relationship OR Extraction OR Extracting OR Entity) (Taxonomy OR Taxonomies OR Ontology OR Ontologies OR Thesaurus OR Thesauri)
IEEE	("Abstract":extract*) AND ("Abstract": "unstructured text" OR "Abstract": "unstructured document") AND ("Abstract":Taxonom* OR "Abstract":Ontolog* OR "Abstract":Thesaur*)
Scopus	(TITLE-ABS-KEY (extract*) AND TITLE-ABS-KEY (entity) AND TITLE-ABS-KEY (relation*) AND TITLE-ABS-KEY ("unstructured text" OR "unstructured document") AND TITLE-ABS-KEY (taxonom* OR ontolog* OR thesaur*))

² Portal de Periódicos CAPES. Disponível em: <https://www-periodicos-capes-gov-br/>. Acesso em: 11 abr. 2024.

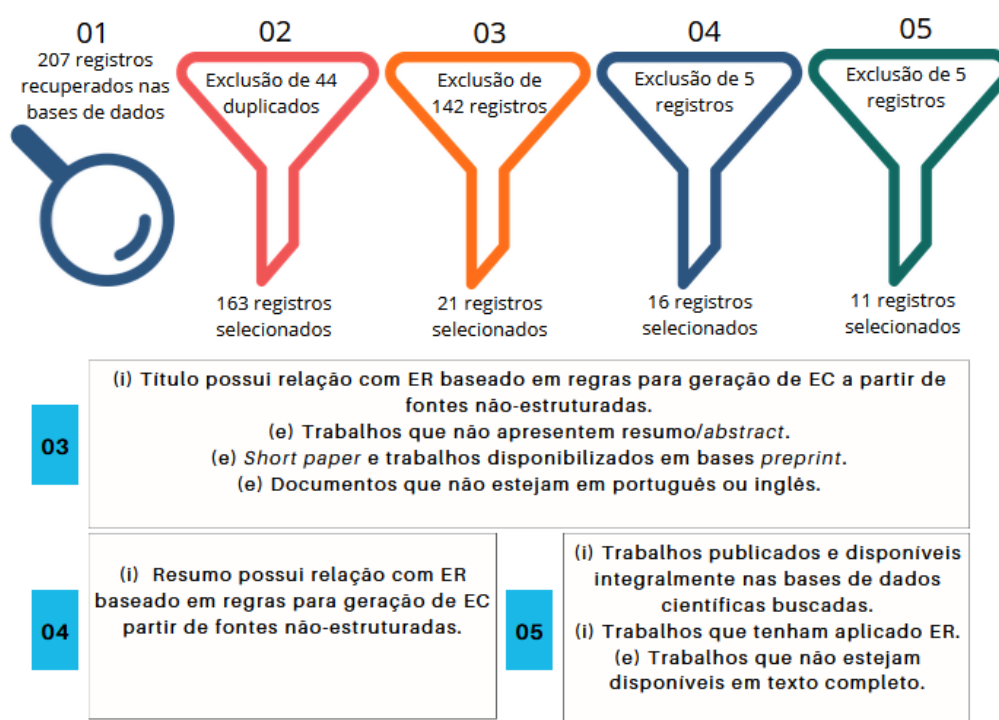
XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

WoS	(((AB=(extract*)) AND AB=(entity)) AND AB=(relation*)) AND AB=("unstructured text" OR "unstructured document")) AND AB=(Taxonom* OR Ontolog* OR Thesaur*)
-----	---

Fonte: Dados da pesquisa (2024).

Na segunda etapa, condução da revisão, é iniciada pela identificação dos estudos (2A) seguida pela seleção dos estudos (2B) relevantes, seguindo o processo de triagem dos estudos e os critérios de inclusão (i) e exclusão (e), conforme demonstrado na Figura 2. Como resultado, foram selecionados 11 (onze) estudos que contemplaram os critérios determinados.

Figura 2 – Esquema do processo de seleção de estudos



Fonte: Elaborado pelos autores (2024).

Para orientar a aplicação da ferramenta *Semantizar*, elaborou-se um protocolo conforme Figura 3. Esse protocolo levou em consideração a interpretação dos estudos da revisão de literatura, fase 3A do esquema de revisão, a abordagem de aplicação e a adaptabilidade ao *Semantizar*.

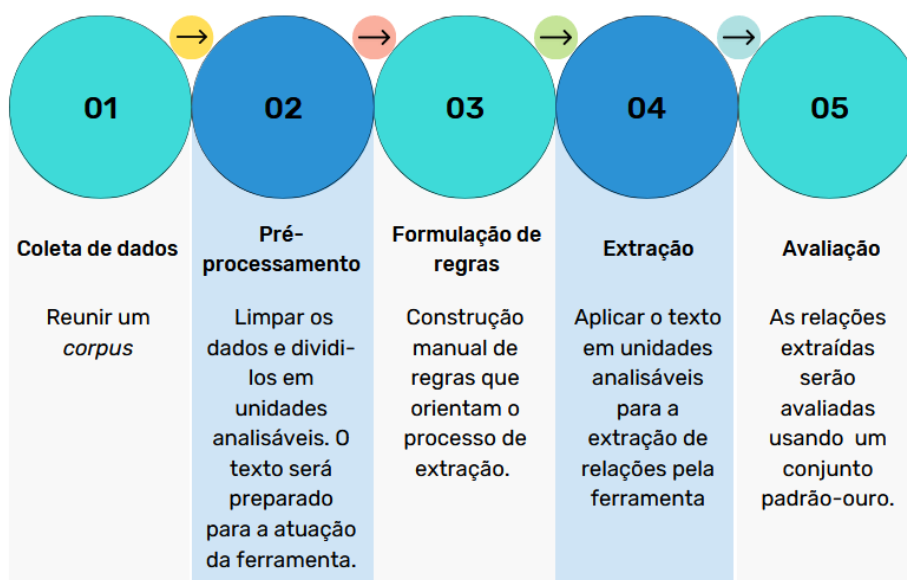
Na coleta de dados, é selecionado documentos para a extração que, para este estudo de caso, foi selecionado o livro “Evolução agrária e pressão demográfica”.

No pré-processamento, é analisada a estrutura textual do conteúdo do documento sendo esta composta das seguintes partes: (1) pré-textuais: inclui a folha de rosto e o sumário

com a descrição dos capítulos; (2) textual: inclui, além da introdução e conclusão, 14 (quatorze) capítulos e; (3) um elemento pós-textual composto pelas referências utilizadas pela autora. Todo o conteúdo está distribuído em 141 (cento e quarenta e uma) páginas, em formato PDF, convertido para um editor de texto simples para o processamento.

A formulação de regras, etapa 3 do protocolo, foi desenvolvida por Oliveira (2018), e envolve cinco etapas: (1) determinação do domínio do conhecimento; (2) seleção de 26 conceitos conforme Quadro 2. Em seguida, (3) definição da terminologia selecionada; (4) determinação do método para a representação das unidades de conhecimento do domínio; e (5) estabelecimento e refinamento das relações semânticas.

Figura 3 – Protocolo de aplicação do Semantizar



Fonte: Elaborado pelos autores (2024).

Quadro 2 - Lista de conceitos para a amostra

Agricultura extensiva	Modernização	Pousio em floresta
Agricultura intensiva	Mudança tecnológica	Produção de alimentos
Biodiversidade	Período de cultivo	Produção de alimentos para humanos
Crescimento populacional	Período de pousio	Recurso fundiário
Cultura anual	Policultura	Sistema de cultivo
Escassez de terra	Posse da terra	Solo
Fertilidade do solo	Pousio	Território e paisagem
Inovação agropecuária	Pousio com vegetação arbustiva	Uso da terra
Intensificação agropecuária	Pousio curto	

Fonte: Oliveira (2018, p. 111)

A extração envolve a aplicação da ferramenta no documento selecionado e, por fim, a avaliação qualitativa dos relacionamentos semânticos gerados com o padrão ouro. Para Xu *et al.* (2022), o termo "padrão ouro" (*gold standard*), refere-se a um conjunto de dados anotados e validados por especialistas humanos, usado como referência para avaliar a precisão e a eficácia de sistemas automatizados.

3 FUNDAMENTAÇÃO TEÓRICA

Com o advento da Internet, a produção diária de textos digitais na forma de mídias sociais, *e-mails*, *blogs*, notícias, livros, artigos de pesquisa e fóruns de perguntas e respostas aumentou significativamente. Esses textos, não estruturados ou semiestruturados, contêm uma enorme quantidade de informações, algumas vezes "desordenadas". Jurafsky e Martin (2023) afirmam que a Extração de Informação (EI) visa obter de um texto em linguagem natural, informações para representá-los em um formato estruturado.

A Figura 4 mostra um exemplo de EI. Nela, é possível observar a organização das informações obtidas de uma fonte textual (que podem ser consideradas como dados), nomeado no exemplo como texto livre. Essa estruturação, conhecida como *template*, possibilita a interpretação computacional dos dados coletados para abastecer um banco de dados ou realizar inferências lógicas, por exemplo.

Figura 4 – O exemplo de uma entrada e saída de extração de informação



Fonte: Elaborado pelos autores (2024).

Segundo Jurafsky e Martin (2023), a EI é normalmente dividida em tarefas de interesse dependendo do tipo de informação que se quer extrair. Na literatura, as tarefas mais citadas são Reconhecimento de Entidades Nomeadas (REN), Extração de Eventos (EE), e Extração de Relações (ER). Este artigo concentra-se na ER.

A ER é um método de Processamento de Linguagem Natural (PLN) que visa identificar e extrair relações semânticas em um texto. Em vez de utilizar algoritmos de aprendizado de máquina que requerem grandes conjuntos de dados para treino, a ER depende do conhecimento especializado e da compreensão do domínio para criar padrões específicos que possam capturar as relações desejadas. Esse conhecimento especializado, é particularmente útil em domínios com uma linguagem altamente estruturada ou técnica, em que os padrões são relativamente estáveis e previsíveis, e pode ter a vantagem de ser mais transparente, fácil de interpretar e mais preciso para conjuntos de dados menores ou mais específicos, os quais, os modelos de aprendizado de máquina podem não ter desempenho eficiente.

De acordo com Sarawagi (2008), uma das formas de resolução de ER é: dado um par de entidades nomeadas (tratadas neste artigo como conceitos), precisa-se descobrir qual a relação existente entre elas. Seguindo essa lógica, o Semantizar foi elaborado.

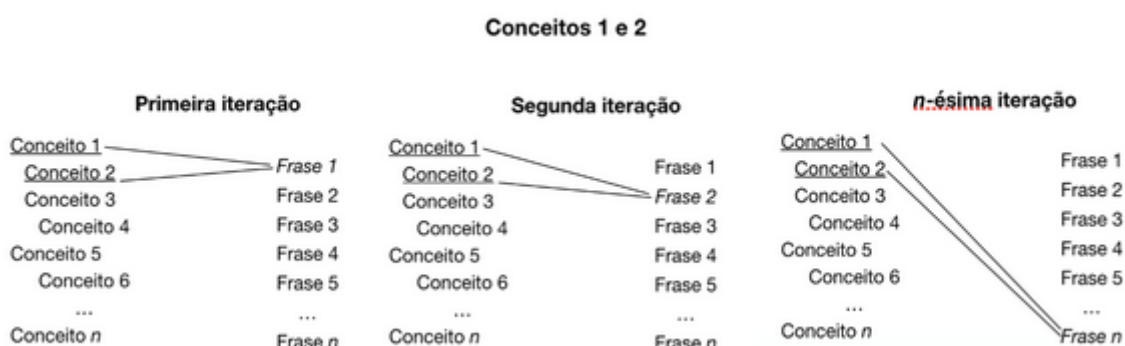
3.1 Semantizar

O Semantizar é um sistema de informação na web que foi desenvolvido para dar suporte à extração de relações semânticas a partir de estruturas classificatórias, que

representam documentos acadêmicos específicos. Desse modo, o Semantizar realiza buscas, nesses documentos, de pares de conceitos em frases, indicando a existência de relação semântica.

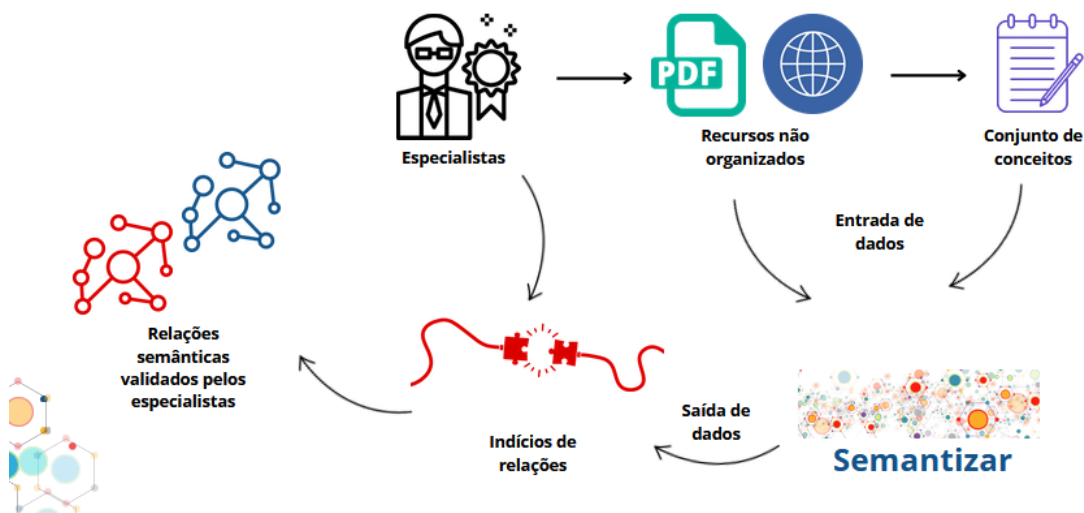
De acordo com Maia (2018), dos conceitos contidos na lista de conceitos, para cada par formado por eles, uma busca é realizada em cada frase do documento para verificar se esses dois conceitos existem na frase. A Figura 5 mostra uma representação das iterações das buscas de pares de conceitos nas frases do documento. A Figura 6 ilustra o esquema de funcionamento descrito.

Figura 5 – Exemplo das iterações das buscas do par de conceitos 1 e 2 nas frases



Fonte: Adaptado de Maia (2018).

Figura 6 – Esquema de funcionamento do Semantizar



Fonte: Elaborado pelos autores (2024).

4 REVISÃO DE LITERATURA

Os onze estudos selecionados para revisão foram: Xiao *et al.* (2013), Shin; Jung; Yi; (2015), Vlachidis; Tudhope (2016), Husin; Saad; Noah. (2017), Tissera; Weerasinghe; (2019), Srinivasa; Thilagam (2019), Thomas, A; Sangeetha, S. (2019), Rossanez, *et al.* (2020), Saber; Abdel-Galil; Belal (2022), Sun; Ma; Liu (2022) e Xu *et al.* (2022).

No que diz respeito às áreas de aplicação, os 11 estudos selecionados, ciências biomédicas/saúde e tecnologia da informação/ontologias representam 18,18%, cada. As demais áreas analisadas, ou seja, gestão e negócios, cultura e esportes, engenharia/construção civil, arqueologia, segurança pública/criminalidade, estudos religiosos e linguística/estudos árabes correspondem a 9,09%, cada. Essa distribuição reflete a diversidade de aplicações das técnicas de ER e geração de conhecimento em diferentes áreas.

Nos trabalhos analisados, identificou-se algumas etapas no processo de extração de relações baseada em regras, sendo elas: análise textual, coleta de dados, pré-processamento, reconhecimento de entidades nomeadas, aprendizado de máquina, extração de relações e avaliação.

Prosseguindo com a análise dos artigos, constatou-se as seguintes estruturas semânticas resultantes: base do conhecimento que representa 27,27% dos estudos, consiste em sistemas que organizam e consolidam informações de forma acessível e utilitária; grafos do conhecimento que compõem 18,18% dos estudos, são estruturas que conectam entidades e relações de maneira que os dados possam ser navegados e consultados semanticamente; representação estruturada, abrangendo 18,18% dos estudos, refere-se a várias formas de organização de dados, como estruturas hierárquicas; e finalmente, ontologias, representando 36,36% dos estudos, envolvem a criação de esquemas formais que descrevem as relações entre conceitos dentro de um domínio.

5 RESULTADOS

Prosseguindo, na etapa de extração, os dois primeiros capítulos da amostra, (1) A dinâmica e o uso da terra e (2) A interdependência entre uso da terra e mudança técnica, juntamente com os 26 termos, foram submetidos ao *Semantizar*. No primeiro capítulo, a ferramenta apresentou 18 (dezoito) indícios de relações e, no segundo capítulo, 30 (trinta).

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

Em seguida, os 48 (quarenta e oito) indícios de relações semânticas foram analisados. Após a análise, 4 relações semânticas foram validadas, as quais foram comparadas com a estrutura semântica de Oliveira (2018), considerada como padrão ouro para esta proposta.

As relações semânticas validadas são apresentadas no Quadro 3, composta de: frase na qual os conceitos foram extraídos, os indícios encontrados pela ferramenta *Semantizar* e o estabelecimento da relação semântica.

Quadro 3 – Relações semânticas validadas

Frase 1	Em vastas regiões da África e da América Latina, e em algumas partes da Ásia, o sistema de uso da terra é extensivo, seguindo-se a cada um ou dois anos de cultivo um período de pousio de pelo menos vinte anos	
	Indícios	Período de pousio. Uso da terra.
	Relação semântica	Uso da terra precisa de período de pousio.
Frase 2	Não se pode negar que o potencial de produção de alimentos tenha sido comprometido por populações que desconheciam a maneira de compatibilizar seus números crescentes com o uso mais intensivo de terra sem prejudicar o solo temporária ou definitivamente	
	Indícios	Produção de alimentos - Solo.
	Relação semântica	Produção de alimentos afeta o uso do solo.
Frase 3	Quando o pousio é encurtado, ou mesmo eliminado, em um dado território, outros métodos de preservação ou recuperação da fertilidade do solo têm de ser introduzidos	
	Indícios	Pousio - Fertilidade do solo
	Relação semântica	Pousio preserva fertilidade do solo Pousio recupera fertilidade do solo

Fonte: Elaborado pelos autores (2024).

Nenhuma das relações semânticas encontradas até o momento, foram contempladas na estrutura conceitual de Oliveira (2018). Dos pares de conceitos sugeridos pelo *Semantizar*, cujas relações semânticas foram validadas, somente o par "uso da terra" e "período de pousio" também estão presentes em Oliveira (2018). Nesse caso, na estrutura conceitual da autora, a relação semântica entre esses conceitos é: uso da terra **influencia** período de pousio. No caso da relação validada sugerida pelo *Semantizar*, a relação semântica é: uso da terra **precisa de** período de pousio.

6 CONSIDERAÇÕES FINAIS

Os resultados apresentados são incipientes. Até então, dois capítulos do livro utilizado como insumo foram utilizados. Outros dez serão analisados no *Semantizar* para explicitarem relações semânticas.

Em Oliveira (2018), para os 26 (vinte e seis) termos da amostra, foram explicitadas 313 (trezentos e treze) relações semânticas. Dessas, 70 (setenta) eram estritamente entre os conceitos da amostra. Como trabalhos futuros, as relações de Oliveira serão comparadas com as relações validadas a partir das sugestões feitas pelo *Semantizar*. Desse modo, acredita-se que o *Semantizar* pode sugerir relações que não estão contempladas na estrutura conceitual de Oliveira (2018), como apontado no recorte realizado, em que os pares de conceitos: "produção de alimentos" e "solo" e "pousio" e "fertilidade do solo" não estão explicitados na estrutura conceitual de Oliveira. As quatro relações semânticas explicitadas a partir do *Semantizar*, não existem na estrutura conceitual. Assim, o *Semantizar* ampliou a estrutura conceitual. Portanto, o uso de uma ferramenta de extração de relações, pode contribuir com profissionais da informação para a construção ou o enriquecimento de estruturas conceituais.

Ressalta-se que, a estrutura conceitual apresentada por Oliveira (2018) seguiu uma metodologia que considera termos que excedem os 26 (vinte e seis) da amostra. Por sua vez, o *Semantizar* tem como característica analisar exatamente os termos da amostra. Logo, o quantitativo de relações semânticas explicitadas a partir do *Semantizar*, pode ser menor que os explicitados na estrutura conceitual de Oliveira (2018) devido ao algoritmo que limita as relações entre os conceitos da amostra. Contudo, isso não invalida o trabalho, pelo contrário, mostra outro ponto de vista de relações entre os conceitos.

Para os resultados definitivos, os tipos das relações semânticas serão explorados, conforme sugere o *Semantizar*, pois eles desempenham um papel crucial na compreensão de sistemas complexos em diversos campos de estudo. Do mesmo modo, as propriedades das relações semânticas serão determinadas. Assim, será possível compreender, em sua completude, a relação semântica explicitada, garantindo assim a validade da relação semântica. Por fim, as relações inversas também serão reveladas. Destarte, acredita-se que outras relações semânticas poderão surgir a partir da amostra.

REFERÊNCIAS

- BATES, Marcia J. Subject access in online catalogs: A design model. **Journal of the American Society for Information Science**, Syracuse, v. 37, n. 6, p. 357-376, 1986. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(198611\)37:6<357::AID-ASI1>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-4571(198611)37:6<357::AID-ASI1>3.0.CO;2-H). Acesso em: 12 jan. 2025.
- GREEN, R. Relationships in the Organization of Knowledge: an overview. *In*: BEAN, C. A.; GREEN, R. (Ed.). **Relationships in the Organization of Knowledge**. Boston: Kluwer Academic Publishers, 2001. p. 3-18. (Information Science and Knowledge Management, 1. ed., v. 2).
- GUARINO, N. Formal ontology, conceptual analysis, and knowledge representation. **International Journal of Human-Computer Studies**, Londres, v. 43, n. 5-6, p. 625-640, 1995. Disponível em: <https://www.loa.istc.cnr.it/old/Papers/FormOntKR.pdf>. Acesso em: 17 maio 2024.
- HJØRLAND, B. Semantics and knowledge organization. **Annual Review of Information Science and Technology**, Syracuse, v. 41, n. 1, p. 1-30, 2007. Disponível em: https://eclass.aueb.gr/modules/document/file.php/INF180/bibliography/Semantics_and_KO_K.pdf. Acesso em: 12 set. 2024.
- HUSIN, M. Z.; SAAD, S.; NOAH, S. A. M. Syntactic rule-based approach for extracting concepts from quranic translation text. *In*: INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING AND INFORMATICS, 6., 2017, Langkawi. **Proceedings** [...]. Langkawi: IEEE, 2017. p. 1-6. Disponível em: https://www.researchgate.net/publication/323821531_Syntactic_rule-based_approach_for_extracting_concepts_from_quranic_translation_text. Acesso em: 22 maio 2024.
- JURAFSKY, D.; MARTIN, J. H. **Speech and language processing: an introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3. ed. New Jersey: Prentice Hall PTR, 2023.
- KUMAR, Ela. **Natural language processing**. New Delhi: IK International Publishing House Pvt. Ltd., 2013.
- MAIA, L. S. **Extração e explicitação de relações semânticas para representação de documentos acadêmicos: um estudo de caso a partir de uma estrutura classificatória**. 2018. 248 f. Tese (Doutorado em Gestão e Organização do Conhecimento) – 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 jun. 2023.
- MAZZOCCHI, F.; TIBERI, M. Knowledge organization in the philosophical domain: dealing with polysemy in thesaurus building. **Knowledge Organization**, [S.l.], v. 36, n. 2-3, p. 103-112, 2009. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2009-2-3-103/knowledge-organization-in-the-philosophical-domain-dealing-with-polysemy-in-thesaurus-building-volume-36-2009-issue-2-3?page=1>. Acesso em: 16 maio 2024.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

OLIVEIRA, E. D. **Proposta de um modelo de hipertexto com abordagem semântica para a representação do conhecimento no domínio temático da intensificação agropecuária**. 2018. 233 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/ECIP-B2EJSW>. Acesso em: 04 abr. 2024.

PAWAR, S.; PALSHIKAR, G. K.; BHATTACHARYYA, P. Relation extraction: a survey, 2017. [Preprint submitted to Elsevier]. Disponível em: <https://arxiv.org/pdf/1712.05191v1>. Acesso em: 12 set. 2024.

ROSSANEZ, A.; REIS, J.; TORRES, R.; RIBAUPIERRE, H. KGen: a knowledge graph generator from biomedical scientific literature. **BMC medical informatics and decision making**, [S.l.], v. 20, p. 1-24, 2020. Disponível em: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01341-5>. Acesso em: 21 maio 2024.

SABER, Y. M.; ABDEL-GALIL, H.; BELAL, M. A. Arabic ontology extraction model from unstructured text. **Journal of King Saud University-Computer and Information Sciences**, Riyadh, v. 34, n. 8, p. 6066-6076, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157822000404>. Acesso em: 21 maio 2024.

SARAWAGI, S. Information extraction. *In*: SARAWAGI, S. **Foundations and Trends® in Databases**. Boston: Now Publishers, 2008. v. 1, n. 3, p. 261-377.

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**, Seul, v. 9, n. 1, p. 407-420, jan. 2015. Disponível em: http://kirc.kaist.ac.kr/papers/journal/2015_Building_a_Business_Knowledge_Base_by.pdf. Acesso em: 18 maio 2024.

SRINIVASA, K.; THILAGAM, P. Santhi. Crime base: Towards building a knowledge base for crime entities and their relationships from online newspapers. **Information Processing & Management**, [S.l.], v. 56, n. 6, p. 1-19, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306457318306885?via%3Dihub>. Acesso em: 12 jan. 2025.

SUN, H.; MA, Y.; LIU, J. Construction of mass cultural and sports activity knowledge graph based on the fusion of multisource data. *In*: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND COMMUNICATION SYSTEMS, 2022, Chickballapur. **Proceedings [...]**. Chickballapur: IEEE, 2022. p. 1-6.

THOMAS, A.; SANGEETHA, S. An innovative hybrid approach for extracting named entities from unstructured text data. **Computational Intelligence**, [S.l.], v. 35, n. 4, p. 799-826, 2019.

TISSERA, M.; WEERASINGHE, R. Auto generation of gold standard, class labeled data set and ontology extension tool [QuadW]. *In*: INTERNATIONAL CONFERENCE ON ADVANCED

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

COMPUTATIONAL AND COMMUNICATION PARADIGMS, 2., 2019, Gangtok. **Proceedings** [...]. Gangtok: IEEE, 2019. p. 1-6.

UREN, V.; CIMIANO, P.; IRIA, J.; HANDSCHUH, S.; VARGAS-VERA, M.; MOTTA, E.; CIRAVEGNA, F. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. **Journal of Web Semantics**, [S.l.], v. 4, n. 1, p. 14-28, 2006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1570826805000338>. Acesso em: 12 jan. 2025.

VLACHIDIS, A.; TUDHOPE, D. A knowledge-based approach to Information Extraction for semantic interoperability in the archaeology domain. **Journal of the Association for Information Science and Technology**, [S.l.], v. 67, n. 5, p. 1138-1152, 2016. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23485>. Acesso em: 12 jan. 2025.

XIAO, L.; TANG, K.; LIU, X.; YANG, H.; CHEN, Z.; XU, R. Information extraction from nanotoxicity related publications. *In*: INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE, 2013, Shanghai. **Proceedings** [...]. Shanghai: IEEE, 2013. p. 25-30.

XU, N.; CHANG, H.; XIAO, B.; ZHANG, B.; LI, J.; GU, T. Relation extraction of domain knowledge entities for safety risk management in metro construction projects. **Buildings**, [S.l.], v. 12, n. 10, p. 1-16, 2022. Disponível em: <https://www.mdpi.com/2075-5309/12/10/1633>. Acesso em: 18 maio 2024.