

**CRIAÇÃO DE UMA REDE DE CONHECIMENTO UTILIZANDO INTELIGÊNCIA
ARTIFICIAL A PARTIR DA ASSOCIAÇÃO DE BASES TÉCNICAS E
TECNOLÓGICAS BRASILEIRAS: PROPOSTA DE PESQUISA**

*CREATION OF A KNOWLEDGE NETWORK USING ARTIFICIAL INTELLIGENCE BASED
ON THE ASSOCIATION OF BRAZILIAN TECHNICAL AND TECHNOLOGICAL BASES:
RESEARCH PROPOSAL*

KATIA CINARA TREGNAGO CUNHA
UNINOVE – UNIVERSIDADE NOVE DE JULHO

CRISTINA DAI PRÁ MARTENS
UNINOVE – UNIVERSIDADE NOVE DE JULHO

CARLA BONATO MARCOLIN
FACULDADE DE GESTÃO E NEGÓCIOS - FAGEN/UFU

Agradecimento à órgão de fomento:

Os autores agradecem o apoio do Fundo de Amparo à Pesquisa - FAP-UNINOVE.

CRIAÇÃO DE UMA REDE DE CONHECIMENTO UTILIZANDO INTELIGÊNCIA ARTIFICIAL A PARTIR DA ASSOCIAÇÃO DE BASES TÉCNICAS E TECNOLÓGICAS BRASILEIRAS: PROPOSTA DE PESQUISA

Objetivo do estudo

Desenvolver uma rede de conhecimento para análise de inovação a partir da associação de bases técnicas e tecnológicas brasileiras, utilizando recursos de inteligência artificial.

Relevância/originalidade

Integração das bases de informação técnica e tecnológica brasileiras, com automatização da extração de informações de campos textuais, de forma a criar uma rede de conhecimento para fornecer conhecimento sistematizado para os processos de inovação.

Metodologia/abordagem

Estudo estruturado em três etapas interligadas, incluindo Revisão Sistemática da Literatura, construção de um banco de dados baseado nos Princípios inventivos da Teoria da Resolução Inventiva de Problemas (TRIZ) e integração de bases de informação técnica e tecnológica brasileiras.

Principais resultados

Identificação de métodos e ferramentas de extração de termos chave de documentos de patente e construção de um corpus terminológico para analisar dados textuais massivos para serem incorporados em modelos de aprendizado de máquina para a extração supervisionada de palavras-chave.

Contribuições teóricas/metodológicas

Oferecimento de uma base de efeitos técnicos traduzida para a língua portuguesa para aplicação em ferramentas de mineração de textos com o objetivo de extrair informações de documentos científicos.

Contribuições sociais/para a gestão

Interligar conhecimento acumulado em bases de informações técnicas e tecnológicas para subsidiar a gestão estratégica, gestão de projetos e ações de pesquisa e desenvolvimento.

Palavras-chave: Inovação, Inteligência Artificial, Conhecimento, Informação tecnológica, Informação Técnica

*CREATION OF A KNOWLEDGE NETWORK USING ARTIFICIAL INTELLIGENCE BASED
ON THE ASSOCIATION OF BRAZILIAN TECHNICAL AND TECHNOLOGICAL BASES:
RESEARCH PROPOSAL*

Study purpose

Develop a knowledge network for innovation analysis based on the association of Brazilian technical and technological bases, using artificial intelligence resources.

Relevance / originality

Integration of Brazilian technical and technological information bases, with automation of the extraction of information from textual fields, in order to create a knowledge network to provide systematized knowledge for innovation processes.

Methodology / approach

Study structured in three interconnected stages, including Systematic Literature Review, construction of a database based on the Inventive Principles of the Theory of Inventive Problem Solving (TRIZ) and integration of Brazilian technical and technological information bases.

Main results

Identification of methods and tools for extracting key terms from patent documents and building a terminological corpus to analyze massive textual data to be incorporated into machine learning models for supervised keyword extraction.

Theoretical / methodological contributions

Offering a base of technical effects translated into Portuguese for application in text mining tools in order to extract information from scientific documents.

Social / management contributions

Interconnect accumulated knowledge in technical and technological information bases to support strategic management, project management and research and development actions.

Keywords: Innovation, Artificial Intelligence, Knowledge, Technological information, Technical information

CRIAÇÃO DE UMA REDE DE CONHECIMENTO UTILIZANDO INTELIGÊNCIA ARTIFICIAL A PARTIR DA ASSOCIAÇÃO DE BASES TÉCNICAS E TECNOLÓGICAS BRASILEIRAS: PROPOSTA DE PESQUISA

1 Introdução

O conhecimento é um recurso essencial para a sociedade. Nos ambientes de inovação, o conhecimento é um insumo crítico (Takeuchi, 2013; Zheng et al., 2011). A capacidade de inovar e de obter vantagem competitiva sustentável é alcançada por meio da gestão dos recursos intelectuais (Cabrilo & Dahms, 2018). As modernas tecnologias digitais têm contribuído na disseminação e preservação dos resultados das atividades de Pesquisa e Desenvolvimento (P&D) (Antopolskii & Efremenko, 2019), fundamentais para aumentar as oportunidades de inovações científicas e tecnológicas e o progresso da sociedade (Smirnova et al., 2021).

Neste contexto em que o conhecimento e a inovação desempenham um papel cada vez mais importante na sobrevivência e no desenvolvimento das empresas, a integração das bases de informação técnica e tecnológica é um desafio vivenciado por diversas nações (Antopolskii & Efremenko, 2019; Namdarian et al., 2021). Dentre as dificuldades, está a complexidade de automatizar a extração de informações de campos textuais não homogêneos e com construção semântica distinta. Ainda, devem ser estabelecidas conexões entre as diferentes bases, de modo que os sistemas inteligentes de máquinas possam se basear neles para recuperar e analisar informações (Krestel et al., 2021).

Assim, a questão central de pesquisa do presente estudo é: Como integrar as bases técnicas e tecnológicas brasileiras, utilizando recursos de inteligência artificial, para oferecer uma rede de conhecimento para os ambientes de inovação? O objetivo é desenvolver uma rede de conhecimento para análise de inovação a partir da associação de bases técnicas e tecnológicas brasileiras. Este documento apresenta esta proposta de pesquisa e os passos para seu desenvolvimento.

2 Referencial teórico

Nos últimos anos, a necessidade de mesclar e integrar diferentes tecnologias tem se tornado cada vez mais importante nos esforços para desenvolver novos produtos e serviços no campo da inovação. Assim, o conhecimento de um campo do domínio pode ser apropriado por outro campo do conhecimento, permitindo desenvolver produtos e processos. Para tanto, é necessária uma grande quantidade de dados que, geralmente, não estão unificados, exigindo do pesquisador diferentes métodos de acesso, recursos e tempo (Su et al., 2022).

Na era de rápida inovação tecnológica, cada vez mais as empresas têm recorrido à inteligência técnica para, a partir de informações de Ciência, Tecnologia e Inovação (C&T&I), monitorar o mercado, a concorrência, acompanhar tendências e antecipar mudanças (Ashton e Stacey, 2014). Constitui inteligência técnica o conhecimento disponibilizado em bases cognitivas (Yoon & Park, 2004), incluindo informações técnicas e tecnológicas que se dedicam à resolução de um problema técnico.

Dentre as bases de informação de C&T&I, tem-se os repositórios de patentes, que são a principal fonte de inovação tecnológica (Xu et al., 2022), onde são descritos novos produtos e/ou novos processos, desenvolvidos a partir da intervenção humana; a produção científica, acumulada em diversas bases, compartilhando a produção técnica e tecnológica por meio de relatos técnicos e artigos tecnológicos, oferecendo importante contribuição prática para a sociedade; por fim, as bases enciclopédicas que reúnem conhecimento especializado em diversas áreas do conhecimento.

Para a criação de uma rede interconectada para fundir conhecimento de diferentes documentos técnicos e tecnológicos, deve ser construído um *corpus* terminológico para analisar

dados textuais massivos codificado em uma ontologia, que sintetiza e carrega o conhecimento central de um determinado domínio tecnológico (Liwei, 2022). Este corpus terminológico, parte do estágio de pré-processamento de um modelo de aprendizado de máquina, especifica o contexto aos modelos de aprendizado de máquina para permitir que realizem a extração supervisionada de palavras-chave e a análise (Singh, 2018). Atualmente, não há corpus anotado direcionado, completo e em larga escala na literatura de patentes (Liwei, 2022; Xu et al., 2019).

3 Metodologia

O presente estudo é estruturado em três etapas interligadas (Da Costa et al., 2019). Uma primeira etapa se dedicou a analisar o conhecimento acumulado a respeito de técnicas de Inteligência Artificial para a mineração de campos textuais de documentos de patente. Nesta primeira etapa foi realizada uma Revisão Sistemática da Literatura (RSL) para analisar o conhecimento acumulado a respeito de técnicas de mineração de textos dedicadas à exploração e aquisição de dados de campos textuais de documentos de patente, identificando as estratégias, os métodos, abordagens e campos patentários explorados.

Para a seleção das publicações foram utilizadas as Bases *Scopus* e *Web of Science*. Para a recuperação dos documentos, a estratégia de busca utilizou termos adquiridos em *handbooks* de linguística a fim de serem obtidos sinônimos. Para a pesquisa, em ambas as bases de consulta foram privilegiados os campos de título, resumo e palavras-chave. A consulta selecionou artigos e documentos de conferência, este último dada a atualidade do tema, sendo consideradas as publicações a partir de 2018 até 29/05/2023, em inglês, espanhol e português. Na primeira etapa da RSL, foram obtidos 1217 artigos, sendo excluídas 243 duplicatas. Para a triagem, foi realizada a leitura do título e do resumo dos 974 artigos, sendo selecionadas 163 publicações. Após a leitura na íntegra, foram incluídos 69 artigos na RSL.

Uma segunda etapa, em desenvolvimento, objetiva construir um banco de dados de efeitos na forma de uma lista de fenômenos físicos ou aplicações de fenômenos físicos, baseada nos Princípios inventivos da Teoria da Resolução Inventiva de Problemas (TRIZ), para ser utilizada como uma ligação semântica para integrar as bases técnicas e tecnológicas. Nesta etapa, os dados obtidos da base de efeitos TRIZ *Oxford Creativity*, contendo 23.861 termos de entrada em inglês, foram armazenados em planilha Excel, subdivididos em subcampos relacionados ao método, função, objeto, conceito e aplicação. Os termos de entrada foram relacionados com o respectivo link da *Wikipédia*. Os termos foram traduzidos para o idioma português a fim de constituir o corpus terminológico para a etapa de associação das bases e extração de informações por meio de ferramentas de mineração de texto. Esta base de efeitos será o suporte para a aplicação do método de mineração e para a testagem das possíveis ligações semânticas que viabilizarão a integração das bases técnicas e tecnológicas.

Por fim, a terceira etapa objetiva integrar as bases de informação técnica e tecnológica brasileiras para criar uma rede de conhecimento para fornecer conhecimento sistematizado para os processos de inovação. Nesta terceira etapa, ainda não iniciada, serão associadas as bases de informação técnica ou tecnológica, utilizando as ligações semânticas testadas na segunda etapa.

4 Análise dos resultados preliminares e proposição de rede de conhecimento

A revisão sistemática da literatura evidenciou que a maior parte dos estudos selecionados descreve métodos que utilizam abordagem de PLN (77% dos artigos), seguido de aprendizado profundo (15%) e aprendizado de máquina (8%). No entanto, vem crescendo, ao longo do tempo, a utilização de técnicas de aprendizado profundo, principalmente em virtude das dificuldades de processamento da grande quantidade de dados (Shalaby & Zadrozny, 2018). Para a extração de termos chave, 77% dos estudos selecionados utiliza técnicas de aprendizado não supervisionado. Nesta categoria, não é requerida a rotulagem de nenhum dado de

treinamento. Por sua vez, o aprendizado supervisionado requer a rotulagem de dados, o que torna a tarefa de extração de palavras-chave mais trabalhosa.

No que diz respeito aos campos textuais explorados pelas ferramentas de mineração de texto, dos artigos que revelam explicitamente na metodologia, 56% (n=40) utilizam de um a três campos de dados textuais de patentes, dentre título, resumo, reivindicações, descrição detalhada ou campo técnico. Nove artigos (13%) revelam utilizar a totalidade dos textos do documento de patente nas atividades de mineração de texto e dois artigos descrevem ferramentas de mineração de texto que utilizam campos textuais e classificação de patente (campo estruturado) como suporte para a extração de informações.

A rede de conhecimento pretendida de ser obtida ao final do estudo é apresentada na Figura 1.



Figura 1 - Diagrama da estrutura da Rede de Conhecimento
Fonte: Elaborado pelos Autores (2023).

5 Considerações finais

O estudo pretende contribuir com a infraestrutura de CT&I brasileira, oferecendo uma base de conhecimento sistematizado que integra informações técnicas e tecnológicas úteis para o setor produtivo e para os ambientes de inovação. Para a área da Administração, esta base de conhecimento se propõe a subsidiar informações relevantes para a gestão estratégica, gestão de projetos e ações de pesquisa e desenvolvimento. No contexto da propriedade intelectual e dos sistemas de informação, o oferecimento de uma base de efeitos traduzida para a língua portuguesa permite aplicações em ferramentas de mineração de textos para extração de informações de documentos científicos, dentre os quais patentes.

Referências

- Antopolskii, A. B., & Efremenko, D. V. (2019). Setting Up the Modern Digital Infrastructure for the Storage and Analysis of Scientific and Technological Information. *Scientific and Technical Information Processing*, 46(2), 123–132.
<https://doi.org/10.3103/S0147688219020138>
- Ashton, W. B., & Stacey, G. S. (2014). Technical intelligence in business: Understanding technology threats and opportunities. *International Journal of Technology Management*.
<https://www.inderscienceonline.com/doi/10.1504/IJTM.1995.025615>
- Cabrilo, S., & Dahms, S. (2018). How strategic knowledge management drives intellectual capital to superior innovation and market performance. *Journal of Knowledge*

- Management*, 22(3), 621–648. <https://doi.org/10.1108/JKM-07-2017-0309>
- Da Costa, P. R., Ramos, H. R., & Pedron, C. D. (2019). Proposição de Estrutura Alternativa para Tese de Doutorado a Partir de Estudos Múltiplos. *Revista Ibero-Americana de Estratégia*, 18(2), 155–170. <https://doi.org/10.5585/riae.v18i2.2783>
- Jiang, M., & Shang, J. (2020). *Scientific Text Mining and Knowledge Graphs*. 3537–3538. <https://doi.org/10.1145/3394486.3406465>
- Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. *World Patent Information*, 65, 102035. <https://doi.org/10.1016/j.wpi.2021.102035>
- Liwei, Z. (2022). Chinese technical terminology extraction based on DC-value and information entropy. *Scientific Reports*, 12(1), 20044. <https://doi.org/10.1038/s41598-022-23209-6>
- Namdarian, L., Alidousti, S., & Rasuli, B. (2021). Developing a comprehensive framework for analyzing national scientific and technical information policy: Application of HeLICAM in Iran. *Online Information Review*, 45(7), 1381–1403. <https://doi.org/10.1108/OIR-11-2020-0493>
- Shalaby, W., & Zadrozny, W. (2018). Toward an Interactive Patent Retrieval Framework based on Distributed Representations. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 957–960. <https://doi.org/10.1145/3209978.3210106>
- Singh, S. (2018). Natural Language Processing for Information Extraction. arXiv pré-impressão arXiv:1807.02383. <https://doi.org/10.48550/arXiv.1807.02383>
- Smirnova, O. V., Denisova, G. V., Alevizaki, O. R., Ilyichenko, D. S., & Antipova, A. S. (2021). Space-Related Scientific and Technical Information in Current Media Discourse: Research Results. *Scientific and Technical Information Processing*, 48(2), 133–138. <https://doi.org/10.3103/S0147688221020118>
- Takeuchi, H. (2013). La Visión de la Estrategia basada en el Conocimiento. *UCJC Business and Society Review (formerly known as Universia Business Review)*, (40).
- Xia, Y. Q., Wei, X. C., Hou, Y. Y., & Lv, H. Q. (2015). *Effect of Integration of Science and Technology Information on Achievements Transformation*. 764–766. <https://doi.org/10.2991/itms-15.2015.181>
- Xu, H., Yue, Z., Pang, H., Elahi, E., Li, J., & Wang, L. (2022). Integrative model for discovering linked topics in science and technology. *Journal of Informetrics*, 16(2), 101265. <https://doi.org/10.1016/j.joi.2022.101265>
- Xu, J., Guo, L., Jiang, J., Ge, B., & Li, M. (2019). A deep learning methodology for automatic extraction and discovery of technical intelligence. *Technological Forecasting and Social Change*, 146, 339–351. <https://doi.org/10.1016/j.techfore.2019.06.004>
- Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, 15(1), 37–50. <https://doi.org/10.1016/j.hitech.2003.09.003>
- Yoon, B., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change*, 72(2), 145–160. <https://doi.org/10.1016/j.techfore.2004.08.011>
- Zheng, S., Zhang, W., & Du, J. (2011). Knowledge-based dynamic capabilities and innovation in networked environments. *Journal of Knowledge Management*, 15(6), 1035–1051. <https://doi.org/10.1108/13673271111179352>