

UMA REVISÃO INTEGRATIVA PARA SISTEMAS DE BUSCA POR PATENTES SIMILARES UTILIZANDO IA - AVANÇOS, DESAFIOS E APLICAÇÕES

Thiago Domingos Marques¹
Alexandre Leopoldo Gonçalves²

Resumo: O presente artigo é uma revisão integrativa para sistemas de busca por patentes similares. Com o crescente volume de dados de patentes disponíveis, que são armazenados em escritórios de patentes, tornou-se essencial desenvolver sistemas eficientes para identificação e verificação em documentos de patentes, analisando as suas semelhanças em termos de tecnologia, aplicação ou abordagem. Foram utilizadas diferentes bases para recuperação de artigos que tenham relevância com o tema. Neste sentido, este artigo objetiva examinar os avanços recentes na aplicação de técnicas de Processamento de Linguagem Natural, Aprendizado de Máquina e Mineração de Texto para melhorar a precisão e eficiência na busca por patentes. A partir dos resultados, vislumbra-se que este trabalho seja útil como fonte de consulta aos pesquisadores e profissionais envolvidos na busca e análise de patentes.

Palavras-chave: dados de patentes; busca de patentes similares; processamento de linguagem natural; aprendizado profundo; mineração de texto.

Abstract: This article is an integrative review of systems for searching similar patents. With the increasing volume of patent data available, stored in patent offices, it has become essential to develop efficient systems for the identification and verification of patent documents, analyzing their similarities in terms of technology, application, or approach. Different databases were used to retrieve articles relevant to the topic. In this context, this article aims to examine recent advances in the application of Natural Language Processing, Machine Learning, and Text Mining techniques to enhance the accuracy and efficiency of patent searching. From the results, it is envisioned that this work will serve as a valuable reference for researchers and professionals involved in patent search and analysis.

Keywords: patent data; search for similar patents; natural language processing; deep learning; text mining.

Resumen: Este artículo es una revisión integrativa de sistemas para buscar patentes similares. Con el creciente volumen de datos de patentes disponibles, almacenados en oficinas de patentes, se ha vuelto esencial desarrollar sistemas eficientes para la identificación y verificación de documentos de patentes, analizando sus similitudes en términos de tecnología, aplicación o enfoque. Se utilizaron diferentes bases de datos para recuperar artículos relevantes para el tema. En este contexto, este artículo tiene como objetivo examinar los avances recientes en la aplicación de técnicas de Procesamiento del Lenguaje Natural, Aprendizaje Automático y Minería de Texto para mejorar la precisión y eficiencia de la búsqueda de patentes. A partir de

¹ Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC.) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0001-9161-067X> e-mail: thiagomestradoufsc@gmail.com

² Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento – Universidade Federal de Santa Catarina (UFSC.) Florianópolis – Brasil. ORCID: <https://orcid.org/0000-0002-6583-2807> e-mail: alexandre.l.golcalves@gmail.com

los resultados, se vislumbra que este trabalho servirá como una referencia valiosa para investigadores y profesionales involucrados en la búsqueda y análisis de patentes.

Palabras clave: datos de patentes; búsqueda de patentes similares; procesamiento del lenguaje natural; aprendizaje profundo; minería de textos.

1 INTRODUÇÃO

A identificação e a correta proteção dos direitos de propriedade intelectual são de fundamental importância para o desenvolvimento econômico e social de uma nação desenvolvida ou em desenvolvimento, sendo um fomentador e caminho em direção à inovação, garantindo aos inventores e as empresas envolvidas as recompensas adequadas por suas criações. A busca e verificação de patentes idênticas ou semelhantes desempenha um papel crucial na garantia da originalidade e unicidade, sendo de suma importância por evitar a duplicação e o plágio, bem como por garantir maiores chances de êxito no pedido (Kim & Yoon, 2022; Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022; Yue et al., 2023; Roudsari et al., 2021).

A busca de patentes semelhantes na identificação de documentos de patentes relevantes, perante o Instituto Nacional de Propriedade Intelectual (INPI), ou outros escritórios de depósitos de patente, ajuda empresas a evitar violações de direitos de propriedade intelectual e a identificar oportunidades de inovação. Neste contexto, a Inteligência Artificial (do inglês *Artificial Intelligence - AI*) tem avançado como uma tecnologia poderosa, útil e eficaz, sendo capaz de analisar grandes volumes de dados, com maior agilidade e em menor tempo, visando identificar semelhanças em padrões complexos, constante em documentos de patentes (Ouyang et al., 2022; Krestel et al., 2021; Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022; Yue et al., 2023; Roudsari et al., 2021).

Para tal, este artigo apresenta uma revisão integrativa sobre o uso da *IA* na realização de buscas de patentes semelhantes em escritórios de depósitos de patentes. A busca por patentes semelhantes é um processo fundamental para empresas, pesquisadores e profissionais do campo da Propriedade Intelectual (do inglês *Intellectual Property - IP*). Desta forma, uma busca eficiente por patentes é essencial para o avanço da inovação e proteção da *IP* (Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022; Yue et al., 2023; Roudsari et al., 2021).

A utilização da *AI* na análise de patentes tem se mostrado uma ferramenta promissora nesse contexto, oferecendo métodos avançados de verificação em dados, possibilitando análises e comparações em grandes volumes de documentos de patentes depositados muitas vezes em diferentes locais (Krestel et al., 2021; Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022; Yue et al., 2023; Roudsari et al., 2021).

Através desta revisão sistemática da literatura, foram identificados os estudos mais relevantes que abordam a aplicação de *AI* na busca de patentes semelhantes, considerando diferentes abordagens e técnicas utilizadas. Os resultados destacam a eficácia da *AI* na identificação de patentes relevantes, mas também evidenciam desafios a serem superados (Ji, Y, 2021; Kim, & Yoon, 2022; Yue et al., 2023).

Inicialmente, a verificação de patentes idênticas era realizada manualmente por examinadores de patentes, o que exigia uma análise minuciosa dos registros existentes. No entanto, com o crescente do volume de patentes e a necessidade de eficiência no processo, métodos computacionais começaram a ser desenvolvidos, avançando de acordo com as necessidades impostas (Ghaffari et al., 2023; Kim, & Yoon 2022; Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022).

Destaca-se ainda as aplicações práticas dos sistemas de busca por patentes similares em diversos setores e áreas de pesquisa. Esses sistemas desempenham um papel crucial na proteção de *IP*, verificação de patentes semelhantes e irregulares, identificação de oportunidades de pesquisa e desenvolvimento, avaliação da concorrência, apoio a estratégias de propriedade intelectual e facilitação de parcerias e colaborações, sendo um diferencial e indutor à inovação (Park, S et al., 2021; Krestel et al., 2021; Lee, 2022; Ouyang et al., 2022; Villa & Wirz, 2022; Zhu & Motohashi, 2022; Yue et al., 2023; Roudsari et al., 2021).

Espera-se que esta revisão integrativa forneça uma visão abrangente dos sistemas de busca por patentes similares, contribuindo para o avanço da pesquisa na área de *IP* e oferecendo *insights* para pesquisadores, profissionais da área e tomadores de decisão envolvidos na busca e análise de patentes.

Este trabalho está estruturado em seis seções, incluindo esta de introdução e a listagem das referências ao final. A segunda seção apresenta os procedimentos metodológicos utilizados. O trabalho continua na seção 3 explicitando a base teórica relativa ao tema. Na seção 4, são apresentados e discutidos os resultados do estudo. Finalmente, na seção 5, são expostas as considerações finais do trabalho.

2 METODOLOGIA

A definição das bases de dados levou em conta questões de relevância científica e multidisciplinares. As bases foram determinadas com base no tema de pesquisa e por serem consideradas multidisciplinares e na área de tecnologia. A partir da definição dos termos de busca, foram definidas as bases e realizada uma busca utilizando termos-chave relacionados ao tema. Os artigos foram selecionados levando-se em conta sua relevância para o objetivo do estudo.

Desta forma, foi efetuada a pesquisa utilizando a expressão a seguir, "*patent analysis*" AND "*litigation analysis*" AND ("*deep learning*" OR "*machine learning*" OR "*artificial intelligence*"). Como ocorreram poucos resultados a expressão foi redefinido para: "*patent analysis*" AND ("*deep learning*" OR "*machine learning*" OR "*artificial intelligence*"), tradução; "análise de patentes" E ("aprendizado profundo" OU "aprendizado de máquina" OU "inteligência artificial"). Adicionalmente, utilizou-se filtro temporal delimitando os anos de 2018 até 2023, para manter a pesquisa alicerçada em artigos recentes.

Inicialmente, encontrou-se com a busca uma grande quantidade de artigos distribuídos em diferentes bases de dados, *ACM Digital Library*[®], *BDTD*[®], *IEEEExplore*[®], *Scielo*[®], *Science Direct*[®], *Scopus*[®], *Springer Link*[®] e *Web of Science*[®]. Com a redefinição por data, com período compreendendo 2018 a 2023, chegou-se a um total de mais de 80 mil artigos. Conforme pode ser demonstrado no quadro 1 abaixo:

Quadro 1 - Resultados das bases de dados

<i>ACM Digital Library</i>	<i>BDTD</i>	<i>IEEE Xplore</i>	<i>Science Direct</i>	<i>Scopus</i>	<i>Springer Link</i>	<i>Web of Science</i>
204,830	0	225	487	141	144	122
<i>Aplicação de Filtro de 2018 até 2023</i>						
80,963	0	145	315	108	144	87

Fonte: autor (2023).

Sendo assim, foi efetuada uma nova redefinição da pesquisa, com novas formações das palavras-chave, utilizando as expressões de busca a seguir: ("*patent analysis*" OR "*patent analytics*" OR "*litigation analysis*") AND ("*deep learning*" OR "*machine learning*" OR "*artificial intelligence*").

Como resultado inicial obteve-se um total de 1681 artigos distribuídos nas diferentes bases. Com a redefinição por data, com período compreendendo os anos de 2018 a 2023, chegou-se a um total de mais de 1123 artigos. Visando a obtenção de artigos mais recentes e atualizados, efetuou-se nova redefinição por data, compreendendo o período de 2021 a 2023, totalizando ao final 31 artigos que foram efetivamente aproveitados. Após a identificação dos artigos, foi realizada uma análise prévia dos mesmos com o intuito de eliminar as duplicatas. Foram identificados durante a leitura dos títulos e resumos que alguns artigos não possuíam relação direta com os temas abordados na pesquisa, sendo então excluídos. Os artigos que necessitavam de pagamento para serem acessados foram descartados, sendo o resultado final demonstrado no quadro 2, abaixo:

Quadro 2 - Resultados das bases de dados

<i>ACM Digital Library</i>	<i>BDTD</i>	<i>IEEE Xplore</i>	<i>Science Direct</i>	<i>Scopus</i>	<i>Springer Link</i>	<i>Web of Science</i>
02	0	06	Repete	03	04	16

Fonte: autor (2023).

Os artigos selecionados foram lidos, analisados e sintetizados para extrair as principais informações relacionadas aos sistemas de busca de patentes similares. Foram identificadas as abordagens, técnicas, metodologias e resultados apresentados em cada artigo

3 RESULTADOS

Nos últimos anos, técnicas avançadas de Processamento de Linguagem Natural (do inglês *Natural Language Processing - NLP*), Mineração de Texto (do inglês *Text Mining - TM*), Aprendizado de Máquina (do inglês *Machine Learning - ML*), incluindo Redes Neurais Artificiais (do inglês *Artificial Neural Network - ANN*) e algoritmos de similaridade de texto, têm sido aplicadas no desenvolvimento de sistemas automatizados para verificação de patentes idênticas.

De acordo com Choi et al. (2021), com as rápidas mudanças na tecnologia e nas cadeias de valor da indústria, tornou-se essencial para as empresas identificar tecnologias emergentes promissoras que possam responder melhor às forças disruptivas externas e serem utilizadas para lançar novos negócios ou melhorar os negócios existentes. Neste sentido, o estudo em bancos de dados de patentes é essencial, visto que é possível a identificação de tecnologias emergentes

promissoras durante a análise de patentes. As patentes há muito tempo são consideradas uma fonte de dados útil sobre tecnologias; consequentemente, vários estudos anteriores se utilizaram de patentes para definir tecnologias em ascensão.

Segundo Jeon et al. (2022), a análise de patentes utilizando técnicas de mineração de texto é uma forma eficaz de identificar tecnologias inovadoras. No entanto, os resultados de estudos anteriores têm sido de uso limitado na prática, uma vez que requerem conhecimento específico do domínio e refletem as características tecnológicas limitadas das patentes.

Seguindo Roudsari et al. (2022), a classificação de patentes é uma tarefa cara e demorada que tradicionalmente tem sido realizada por especialistas na área. Portanto, automatizar essa tarefa cara e trabalhosa é essencial para auxiliar especialistas no gerenciamento de documentos de patentes, facilitando a busca confiável, a recuperação e outras tarefas de análise de patentes. A transferência de aprendizado e os modelos de linguagem pré-treinados recentemente alcançaram resultados de ponta em muitas tarefas de Processamento de Linguagem Natural.

Algumas das dificuldades que podem surgir durante a pesquisa em banco de dados de patentes, é: a heterogeneidade dos dados de patentes, a falta de padronização nos critérios de similaridade, a necessidade de atualização constante dos sistemas e a superação de barreiras linguísticas e culturais em um contexto internacional.

Com base nesse entendimento referente a importância de sistemas de busca de patentes, foi efetuada a organização e categorização das informações extraídas dos artigos relevantes de acordo com as definições estabelecidas, sendo elas: aplicações práticas, desafios e técnicas de pesquisa utilizados, facilitando a identificação de artigos relevantes.

Por meio de leitura e avaliação dos artigos sob as lentes do objeto de pesquisa, podemos avaliar e examinar os avanços recentes na aplicação de técnicas de *NLP*, *ML* e *AI* em análise de patentes.

Abaixo é apresentado o Quadro 3 com os 31 artigos encontrados, com o título e tema de estudo, aplicações práticas, desafios e técnicas de pesquisa utilizadas.

Quadro 3 – Artigos selecionados

No.	Título	Aplicações Práticas	Desafios	Técnicas de Pesquisa Usadas
01	Inteligência Artificial na Prospectiva Estratégica - Práticas Atuais e Potenciais Futuros de Aplicação	Uso da Inteligência Artificial na prospecção estratégica por parte de organizações e empresas	Garantir que a <i>AI</i> seja capaz de lidar com conjuntos de dados cada vez maiores e mais complexos, de forma a realizar uma análise precisa e abrangente das tendências e padrões emergentes, garantir a interpretação correta dos resultados da Inteligência Artificial e abordar questões éticas e sociais associadas à sua aplicação	Algoritmos avançados de Aprendizado de Máquina e Processamento de Linguagem Natural como técnicas utilizadas na aplicação da Inteligência Artificial na prospecção estratégica
02	<i>TechPat</i> : Extração de Frases Técnicas para Mineração de Patentes	Identificação de tendências tecnológicas, Análise de concorrência e propriedade intelectual	Extração precisa de frases técnicas, Mineração eficiente de patentes.	Mineração e extração de textos em frases técnicas de patentes
03	<i>IA</i> para Patentes: uma estrutura nova, mas eficaz e eficiente para análise de patentes	Análise de patentes utilizando Inteligência Artificial	Efetividade e eficiência da estrutura de <i>IA</i> na análise de patentes, ligada a complexidade e alto volume de dados de patentes	Inteligência Artificial, Análise de patentes
04	Um novo método de extração de conhecimento de patentes para design inovador	Extração de conhecimento de patentes para o desenvolvimento de projetos de design inovadores	Extração precisa de conhecimento de patentes, Design inovador adaptado às necessidades específicas	Extração de conhecimento, Design inovador
05	Método de análise de sumarização inteligente de texto de patente	Desenvolvimento de um método de análise de sumarização inteligente de texto de patente	Sumarização precisa de texto de patente, com a linguagem técnica e complexidade dos documentos de patente	Análise de sumarização
06	Reconhecimento de Entidade Nomeada de Aprendizado Pouco Supervisionado do Texto de Patente de Sensor de Pressão Baseado em <i>MLM</i>	Reconhecimento de entidade nomeada em textos de patentes pode facilitar a busca e a análise de patentes, baseado em <i>MLM</i> - "Masked Language Modeling" (Modelagem de Linguagem Enmascarada). <i>MLM</i> é uma técnica de processamento de linguagem natural usada em aprendizado de máquina, onde palavras ou tokens em uma sentença são mascarados (ocultados) e o modelo é treinado para prever as palavras ou tokens mascarados com base no contexto das palavras circundantes	Aprendizado semi supervisionado para reconhecimento de entidade nomeada, acessar e utilizar informações relevantes nos textos	Reconhecimento de Entidade Nomeada, Aprendizado Semi Supervisionado, Texto de patente, Sensor de Pressão, <i>MLM</i> (Masked Language Model)
07	Aprendizado de Representação em Grafos para Análise de Redes de Patentes	Melhorar à análise e compreensão de redes de patentes, por meio de conexões entre patentes	Representação eficaz de redes de patentes, Utilização de técnicas de aprendizado em grafos para análise de patentes	Aprendizado de Representação, Grafos, Análise de Redes
08	Classificação de Patentes com Extração Inteligente de Palavras-chave	Melhorar a identificação de patente	Extração precisa de palavras-chave, Classificação eficiente e precisa de patentes	Classificação de Patentes, Extração Inteligente de Palavras-chave
09	Uma pesquisa sobre aprendizado profundo para análise de patentes	Identificação de patente por meio da Aprendizagem Profunda	Utilização efetiva de Aprendizado Profundo na análise de patentes	Aprendizado Profundo, Análise de patentes
10	Mapeando a dinâmica da inovação tecnológica em domínios de inteligência	Mapeamento de Inovações de Mercado	Análise eficiente e abrangente de patentes, Identificação precisa de	Mapeamento de dinâmica de inovação tecnológica,

No.	Título	Aplicações Práticas	Desafios	Técnicas de Pesquisa Usadas
	artificial: Evidências de uma análise global de patentes		padrões e tendências de inovação tecnológica	Inteligência Artificial, Análise de patentes
11	Apresentando Patentes com Conexão Indireta (PCI) para Estabelecer Estratégias de Patentes	Melhora na precisão da coleta e identificação de patentes	Estabelecer conexões indiretas entre patentes, Utilização efetiva de conexões indiretas para estratégias de patentes	Patentes com Conexão Indireta, Estratégias de Patentes
12	Estabelecendo um Modelo de Previsão de Risco de Patente para Tecnologias Emergentes usando Aprendizado Profundo e Augmentação de Dados	Previsão de riscos	Previsão precisa de risco de patente para tecnologias emergentes, Utilização efetiva de técnicas de Aprendizado Profundo	Modelo de Previsão de Risco de Patente, Tecnologias Emergentes, Aprendizado Profundo
13	Aprendizado de máquina para mapeamento de patentes usando transformadores e incorporação de grafo	Maior precisão na identificação de patente	Mapeamento preciso de patentes, Utilização efetiva de técnicas de Aprendizado de Máquina, Incorporação de grafo em patentes	Aprendizado de Máquina, Mapeamento de Patentes, Arquitetura de Redes Neurais do tipo <i>transformers</i> , Incorporação de Grafo
14	Um método baseado em <i>deep learning</i> que se beneficia das características das patentes para a classificação de relações semânticas	Melhoramento na classificação e extração de patente	Classificação precisa de relações semânticas	Aprendizado Profundo, Classificação de relações semânticas
15	Análise de Patentes Utilizando Análise de Dados Bayesiana e Modelagem de Redes	Verificação e identificação de patente com Análise de Dados Bayesiana	Utilização efetiva de análise de dados bayesiana e modelagem de redes	Análise de Dados Bayesiana, Modelagem de Redes, Patentes
16	Uma revisão sobre as tendências de digitalização em bancos de dados de informações de patentes e ferramentas de interrogatório	Fornecer uma visão geral das tendências e avanços na digitalização de informações de patentes e nas ferramentas de interrogatório.	Análise abrangente e precisa de tendências de digitalização, Desenvolvimento de ferramentas eficientes de interrogatório	Digitalização, Bancos de dados de informações de patentes, Ferramentas de interrogatório
17	Identificando a convergência tecnológica usando informações de texto de patentes: uma abordagem baseada em redes convolucionais de grafos (GCN)	Identificar padrões de convergência tecnológica por meio da análise de informações de texto de patentes.	Utilização efetiva de informações de texto de patentes, Utilização de redes convolucionais de grafos para identificar convergência tecnológica	Convergência tecnológica, Informações de texto de patentes, Redes convolucionais de grafos (do inglês <i>Graph Convolutional Network - GCN</i>)
18	<i>PatentNet</i> : classificação de documentos de patentes com várias etiquetas usando compreensão de linguagem baseada em <i>deep learning</i>	Melhorias na capacidade de automatizar e aprimorar o processo de classificação de documentos de patentes, Busca e recuperação de informações: monitoramento de tendências, Avaliação de portfólio de patentes	Classificação eficiente e precisa de documentos de patentes, Compreensão avançada de linguagem com DL	Classificação de documentos de patentes, Compreensão de linguagem, DL
19	Uma abordagem utilizando <i>doc2vec</i> e <i>local outlier factor</i> para medir a novidade de patentes	A aplicação prática deste artigo está relacionada à medição da novidade de patentes e pode ter as seguintes aplicações: Avaliação de patentes e Identificação de tendências emergentes	Medição precisa da novidade de patentes, Utilização efetiva de técnicas <i>doc2vec</i> e <i>local outlier factor</i>	Medição de novidade de patentes

No.	Título	Aplicações Práticas	Desafios	Técnicas de Pesquisa Usadas
20	Identificando tecnologias emergentes para visualizar um ecossistema de inovação futuro: uma abordagem de aprendizado de máquina para dados de patentes	Vigilância tecnológica e Planejamento estratégico, Formulação de políticas de inovação	Identificação precisa de tecnologias emergentes, Utilização efetiva de aprendizado de máquina em dados de patentes	Identificação de tecnologias emergentes, Aprendizado de Máquina, Dados de patentes
21	Modelo Integrado de Sobrevivência para Prever o Risco de Litígio de Patentes.	Previsão do risco de litígio de patentes. A aplicação prática do artigo está relacionada ao gerenciamento de risco, avaliação de portfólio de patentes e planejamento de pesquisa e desenvolvimento com base nas previsões de litígio de patentes	O desafio abordado neste artigo é o desenvolvimento de um modelo de sobrevivência integrado para prever o risco de litígio de patentes	Aprendizado de Máquina
22	Sumarização de múltiplos documentos para documentos de patentes com base em redes generativas adversárias	Análise e classificação de patentes, pesquisa e desenvolvimento de produtos	Sumarização eficiente de documentos de patentes, Utilização de redes generativas adversárias para sumarização	Sumarização de múltiplos documentos, Redes generativas adversárias
23	Avaliação de patenteabilidade por meio de análise semântica de patentes - Uma abordagem matemático-lógica	Avaliação de risco de litígios, pesquisa de estado da técnica, análise semântica de patentes, permitindo uma avaliação mais objetiva e precisa da relevância e originalidade das invenções, bem como sua conformidade com os requisitos de patenteabilidade	Avaliação precisa de patenteabilidade, Utilização de abordagem matemático-lógica para análise semântica de patente	Análise semântica de patentes, Avaliação de patenteabilidade, Abordagem matemático-lógica
24	Estimulando P&D por meio da Descoberta de Patentes Econômicas: Um Modelo de Processo e uma Comparação entre Métodos de Avaliação Diferentes	Seleção de patentes estratégicas, análise de competidores e tendências de inovação, identificação de patentes "frugais"	Descoberta eficiente de patentes econômicas, Comparação abrangente e precisa de métodos de avaliação. Estímulo eficiente de P&D, Comparação abrangente e precisa de métodos de avaliação	Descoberta de Patentes Econômicas, Estimulação de P&D, Comparação de Métodos de Avaliação
25	Mapeamento de tecnologia baseado em tópicos usando análise de dados de patentes: um estudo de caso de pneus para veículos.	Mapeamento de tecnologias com base em dados de patentes, estudo de tendências	Identificação precisa de tópicos de tecnologia, Análise de dados de patentes para mapeamento de tecnologia	Mapeamento de tecnologia, Análise de dados de patentes, Estudo de caso.
26	Explorando a inovação em IA e cuidados de saúde: análise de patentes baseada em processamento de linguagem natural para mapeamento tecnológico impulsionado.	Identificação de oportunidades de pesquisa e desenvolvimento, por meio de <i>NLP</i>	Explorar a inovação via inteligência artificial, por meio da análise de patentes baseada em processamento de linguagem natural para realizar um mapeamento impulsionado por tecnologia.	Inteligência artificial no estudo de <i>roadmapping</i> baseado em processamento de linguagem natural e análise de patentes
27	Uma abordagem sequencial de busca de patentes que combina semântica e inteligência artificial para identificar documentos iniciais de Estado da Técnica.	Suporte à inovação e desenvolvimento de novas tecnologias. Melhoria do processo de patenteamento: A abordagem proposta pode ter implicações significativas no processo de patenteamento	Combinação eficiente de semântica e inteligência artificial na busca de patentes, Identificação precisa de documentos de Estado da Técnica	Busca de patentes, Semântica, Inteligência Artificial, Documentos de Estado da Técnica
28	Descoberta de Oportunidades Tecnológicas usando Mineração de Texto baseada em Aprendizado	Uso do conhecimento gráfico e da mineração de texto baseada em aprendizado profundo para apoiar a tomada de decisões em projetos de pesquisa e desenvolvimento. A	Descoberta eficiente de oportunidades tecnológicas, Mineração de texto baseada em aprendizado profundo e grafo de conhecimento	Descoberta de Oportunidades Tecnológicas, Mineração de Texto, Aprendizado

No.	Título	Aplicações Práticas	Desafios	Técnicas de Pesquisa Usadas
	Profundo e um Grafo de Conhecimento.	aplicação do aprendizado profundo na mineração de texto pode auxiliar na análise e previsão de tendências tecnológicas		Profundo, Grafo de Conhecimento
29	Inteligência artificial e produtividade: evidências globais a partir de dados de patentes de IA e bibliométricos	Análise dos dados de patentes e bibliometria relacionados à IA para descobrir tendências tecnológicas emergentes e áreas de pesquisa promissoras. Isso pode orientar a alocação de recursos para P&D e inovação	Análise abrangente de dados de patentes de IA e bibliométricos, Relaciona inteligência artificial com produtividade	Inteligência Artificial, Produtividade, Dados de patentes, Bibliométricos
30	Um pipeline baseado em aprendizado profundo com entradas de múltiplos canais para classificação de patentes.	A utilização do aprendizado profundo na classificação de patentes pode aumentar significativamente a eficiência do processo, reduzindo o tempo e o esforço necessários para a categorização manual. Isso pode liberar recursos humanos para outras tarefas mais complexas e estratégicas	Desenvolvimento de pipeline eficiente para classificação de patentes, Utilização de múltiplos canais de entrada	Aprendizado profundo, Classificação de patentes
31	Um método baseado em aprendizado profundo para extrair informações semânticas de documentos de patentes.	Utilização de aprendizado profundo na extração de informações semânticas para aumentar a eficiência da análise de patentes, reduzindo o tempo e o esforço necessários para revisar manualmente grandes volumes de documentos. Isso é particularmente útil para agências de patentes, escritórios de propriedade intelectual e empresas que precisam analisar um grande número de patentes	Extração eficiente de informações semânticas de documentos de patentes, Utilização de aprendizado profundo	Aprendizado profundo, busca de patentes

Fonte: autor (2023).

O objetivo inicial do artigo foi alcançado ao examinarmos os avanços recentes na aplicação de técnicas de *NLP*, *ML* e *TM* para melhorar a precisão e eficiência da busca por patentes similares, explorando os fundamentos da aprendizagem profunda; analisando as suas principais aplicações, limitações e desafios na aplicação da busca de patentes.

Algumas das técnicas de pesquisa utilizadas, não estavam detalhadas em todos os artigos descritos. Bem como, algumas informações sobre as técnicas estavam ausentes ou não especificadas nos dados fornecidos. Outra aplicação é na classificação automática, onde as patentes são geralmente classificadas em categorias específicas com base na tecnologia a que se referem. Observamos que há uma constante de trabalhos recentes na área em diferentes países.

A *AI* pode ser treinada para classificar automaticamente as patentes em categorias apropriadas, com base em recursos extraídos dos documentos de patentes. Essa classificação

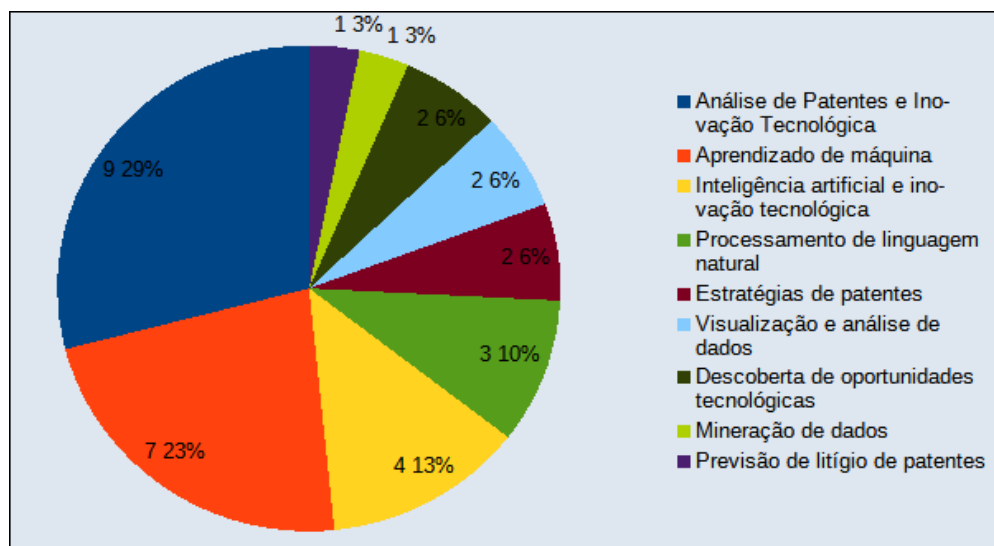
automatizada facilita a pesquisa e a recuperação de patentes relevantes em um determinado campo, melhorando a eficiência da busca.

Ao examinar os artigos, foi possível identificar alguns tópicos recorrentes, tais como: Análise de texto e processamento de linguagem natural em patentes, aplicação de técnicas de aprendizado profundo em análise de patentes; extração de informações e conhecimento a partir de bases de dados de patentes; análise de tendências e padrões de inovação em áreas específicas; previsão de risco de litígio com base em análise de patentes; análise de redes de patentes para identificação de convergência tecnológica; desenvolvimento de modelos de classificação de patentes utilizando *AI*.

Na Figura 1, observa-se que a análise de patentes e inovação tecnológica, aprendizado de máquina e inteligência artificial e inovação tecnológica, são os temas mais abordados nos artigos com 20,65% definidos, com base no histórico de artigos pesquisados e analisadas, de patentes relevantes.

Já os artigos relacionados à descoberta de oportunidades tecnológicas, mineração de dados e previsão de litígio de patentes representam 5,20%. Essa abordagem utiliza algoritmos de aprendizado de máquina para identificar padrões e similaridades entre as patentes e oferecer sugestões com base nos interesses e nas necessidades do usuário operadores.

Figura 1 – Gráfico com as principais categorias definidas nos artigos



Fonte: autores (2023).

Esses tópicos refletem a diversidade de abordagens e aplicações da *AI* na análise de patentes. Através da revisão desses artigos, é possível obter uma visão abrangente das pesquisas

e avanços nessa área, contribuindo para o desenvolvimento de estratégias de inovação, monitoramento tecnológico e tomada de decisões relacionadas à patentes.

Pensando em futuras pesquisas, essa análise pode subsidiar direcionamentos de pesquisas e desenvolvimentos baseadas em métodos das áreas discutidas neste artigo, *NLP*, *MT* e *DL*, onde patentes representam o objeto principal de estudo.

4 CONSIDERAÇÕES FINAIS

A busca por patentes similares desempenha um papel fundamental na pesquisa, inovação e estratégias de propriedade intelectual, sendo de suma importância para as empresas, governos, pesquisadores das mais diversas áreas e população, que podem se beneficiar das facilidades promovidas por sistemas automatizados e eficazes de buscas de patentes. Identificou-se, por meio da revisão integrativa, os principais trabalhos envolvendo sistemas de busca por patentes similares, explorando os avanços, desafios e aplicações dessas tecnologias.

Foram observados que os métodos tradicionais de busca por patentes similares estão sendo aprimorados, todavia, ainda apresentam limitações em termos de precisão e eficiência, especialmente diante do grande volume de dados disponíveis. No entanto, os avanços recentes nas técnicas de processamento de linguagem natural, aprendizado de máquina e aprendizado profundo têm proporcionado novas abordagens para identificar patentes similares com base em semelhanças semânticas, estruturais ou de conteúdo, representando um vasto campo a ser explorado e com grande potencial futuro.

Através da revisão, verificou-se que tais tecnologias oferecem funcionalidades avançadas, cobertura de dados abrangente e relativa eficácia na entrega de resultados relevantes para os interessados. Verificou-se ainda a importância e possibilidade de utilização em diversos setores e áreas de pesquisa, fornecendo *insights* valiosos para mapear oportunidades de pesquisa e desenvolvimento, monitorar a concorrência, apoiar estratégias de propriedade intelectual e promover parcerias e colaborações.

No entanto, também foram identificados desafios a serem superados. A heterogeneidade dos dados de patentes, a falta de padronização nos critérios de similaridade e a necessidade de atualização constante dos sistemas são alguns dos desafios enfrentados pelos sistemas de busca por patentes similares. Além disso, a superação de barreiras linguísticas e culturais em um contexto internacional é outro aspecto importante a ser considerado.

Diante desses avanços e desafios, é necessário um esforço contínuo para aprimorar a precisão, a abrangência e a eficiência dos sistemas de busca por patentes similares. Além disso,

são necessárias pesquisas futuras para explorar ainda mais as aplicações dessas tecnologias em diferentes setores e para abordar as limitações existentes.

Os sistemas de busca por patentes similares desempenham um papel fundamental na aceleração da inovação e no avanço da pesquisa científica. Espera-se que esta revisão integrativa forneça uma perspectiva adequada, ainda que inicial, do estado atual dessas tecnologias, estimule discussões e inspire pesquisas adicionais no campo da propriedade intelectual e inovação.

Conclui-se que a utilização de Inteligência Artificial aplicada à busca de patentes tem um grande potencial para melhorar e agilizar esse processo crucial para a inovação e proteção da propriedade intelectual. Os estudos selecionados demonstraram que a utilização de áreas e técnicas de *AI* podem oferecer benefícios significativos.

Entre os benefícios identificados, destacam-se a capacidade da Inteligência Artificial em lidar com grandes volumes de dados, identificar padrões e similaridades que seriam difíceis de serem percebidos por métodos tradicionais de busca de patentes. Além disso, a *AI* permite a otimização dos resultados de busca, aprimorando a eficiência nos diversos processos envolvidos na localização de patentes similares e economizando tempo para os pesquisadores e demais interessados nesse campo de atuação.

REFERÊNCIAS

- Brandtner, P., & Mates, M. (2021). Artificial Intelligence in Strategic Foresight – Current Practices and Future Application Potentials. *The 2021 12th International Conference on E-business, Management and Economics*, July 2021, Pages 75–81. DOI: 10.1145/3481127.3481177.
- Chen, L., Xu, S., Zhu, L., Zhang, J., Lei, X., & Yang, G. (2020). A deep learning based method for extracting semantic information from patent documents. *Scientometrics*, 125(1), 289-312. DOI: 10.1007/s11192-020-03634-y.
- Choi, Y., Park, S., & Lee, S. (2021). Identifying emerging technologies to envision a future innovation ecosystem: a machine learning approach to patent data. *Scientometrics*, 126(7), 5431-5476. DOI: 10.1007/s11192-021-04001-1.
- Chi, Y., & Wang, H. (2022). Establish a patent risk prediction model for emerging technologies using deep learning and data augmentation. *Advanced Engineering Informatics*, 52, 101509. DOI: 10.1016/j.aei.2021.101509.
- Ghaffari, M., Aliahmadi, A., Khalkhali, A., Zakery, A., Daim, T. U., & Yalcin, H. (2023). Topic-based technology mapping using patent data analysis: a case study of vehicle tires. *Technological Forecasting And Social Change*, 193, 122576. DOI: 10.1016/j.techfore.2023.122576.
- Heydari, M., & Teimourpour, B. (2021). Graph Representation Learning Towards Patents Network Analysis. DOI: 10.1109/ACCESS.2022.3176877.

- Jeon, D., Ahn, J. M., Kim, J., & Lee, C. (2022). A doc2vec and local outlier factor approach to measuring the novelty of patents. *Technological Forecasting And Social Change*, 174, 121294. <http://dx.doi.org/10.1016/j.techfore.2021.121294>.
- Ji, Y. (2021). Intelligent Patent Text Summarization Analysis Method. *7th International Conference on Systems and Informatics (ICSAI)*. DOI: 10.1109/ICSAI53574.2021.9664064.
- Joshi, Y. U., Bansal, M., Hedaoo, M., More, V., & Fatnani, P. (2022). Patent Classification with Intelligent Keyword Extraction. In *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)* (pp. 1-5). Pimpri Chinchwad College of Engineering (PCCOE), Pune, India. Aug 26-27, 2022. IEEE. <http://dx.doi.org/10.1109/iccubea54992.2022.10010888>.
- Junyoung Son, Hyeonseok Moon, Jeongwoo Lee, Seolhwa Lee, Hanjun Park, Wonkyung Jung, & Heuseok Lim. (2022). AI for Patents: A Novel Yet Effective and Efficient Framework for Patent Analysis. *IEEE Access*. DOI: 10.1109/ACCESS.2022.3176877.
- Kim, S., & Yoon, B. (2022). Multi-document summarization for patent documents based on generative adversarial network. *Expert Systems With Applications*, 207, 117983. DOI: 10.1016/j.eswa.2022.117983.
- Kim, Y., Park, S., Lee, J., Jang, D., & Kang, J. (2021). Integrated Survival Model for Predicting Patent Litigation Hazard. *Sustainability*, 13(4), 1763. DOI: 10.3390/su13041763.
- Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. *World Patent Information*, 65, 102035. <http://dx.doi.org/10.1016/j.wpi.2021.102035>
- Kronemeyer, L. L., Draeger, R., & Moehrle, M. G. (2023). Stimulating R&D by Finding Frugal Patents: a process model and a comparison between different evaluation methods. *Ieee Transactions On Engineering Management*, 70(2), 615-630. DOI: 10.1109/tem.2021.3058003.
- Lee, J., Park, S., & Kang, J. (2021). Introducing Patents with Indirect Connection (PIC) for Establishing Patent Strategies. *Sustainability*, 13(2), 820. <http://dx.doi.org/10.3390/su13020820>.
- Lee, M., Kim, S., Kim, H., & Lee, J. (2022). Technology Opportunity Discovery using Deep Learning-based Text Mining and a Knowledge Graph. *Technological Forecasting And Social Change*, 180, 121718. DOI: 10.1016/j.techfore.2022.121718.
- Lee, Myoungsoon; Kim, Suhyeon; Kim, Hangyeol; Lee, Junghye. (2022). Exploring AI-healthcare innovation: natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes*, 52(4), 1173-1189. DOI: 10.1108/k-03-2021-0170.
- Liu, N., Shapira, P., Yue, X., & Guan, J. (2021). Mapping technological innovation dynamics in artificial intelligence domains: evidence from a global patent analysis. *Plos One*, 16(12), 0262050. DOI: 10.1371/journal.pone.0262050.
- Liu, Y., Wu, H., Huang, Z., Wang, H., Ning, Y., Ma, J., Liu, Q., & Chen, E. (2023). TechPat: technical phrase extraction for patent mining. *ACM Transactions on Knowledge Discovery from Data*, 17(9), 1-31. DOI: 10.1145/3596603.
- Ouyang, X., Sun, Z., & Xu, X. (2022). Patent system in the digital era - Opportunities and new challenges. *Journal Of Digital Economy*, 1(3), 166-179. <http://dx.doi.org/10.1016/j.jdec.2022.12.003>.

- Park, S., Choi, S., & Jun, S. (2021). Bayesian Structure Learning and Visualization for Technology Analysis. *Sustainability*, 13(14), 7917. DOI: 10.3390/su13147917.
- Parteka, A., & Kordalska, A. (2023). Artificial intelligence and productivity: global evidence from ai patent and bibliometric data. *Technovation*, 125, 102764. DOI: 10.1016/j.technovation.2023.102764.
- Roudsari, A. H., Afshar, J., Lee, W., & Lee, S. (2021). PatentNet: multi-label classification of patent documents using deep learning based language understanding. *Scientometrics*, 127(1), 207-231. DOI: 10.1007/s11192-021-04179-4.
- Schmitt, V. J., Walter, L., & Schnittker, F. C. (2023). Assessment of patentability by means of semantic patent analysis – A mathematical-logical approach. *World Patent Information*, 73, 102182. DOI: 10.1016/j.wpi.2023.102182.
- Sofean, M. (2021). Deep learning based pipeline with multichannel inputs for patent classification. *World Patent Information*, 66, 102060. DOI: 10.1016/j.wpi.2021.102060.
- Teza, P., Miguez, V. B., Fernandes, R. F., & Dandolin, G. A. (2013). Direcionadores do processo de inovação: o papel da estratégia, liderança e cultura. *Navus – Revista de Gestão e Tecnologia*, 3(2), 77-88. DOI: 10.18815/navus.v3i2.156. ISSN 2237-4558.
- Uhm, D., & Jun, S. (2022). Zero-Inflated Patent Data Analysis Using Generating Synthetic Samples. *Future Internet*, 14(7), 211. DOI: 10.3390/fi14070211.
- Villa, A. M., & Wirz, M. (2022). A sequential patent search approach combining semantics and artificial intelligence to identify initial State-of-the-Art documents. *World Patent Information*, 68, 102096. DOI: 10.1016/j.wpi.2022.102096.
- Walter, L., Denter, N. M., & Keibel, J. (2022). A review on digitalization trends in patent information databases and interrogation tools. *World Patent Information*, 69, 102107. <http://dx.doi.org/10.1016/j.wpi.2022.102107>.
- Wang, Y. H., & Lin, G. Y. (2022). Exploring AI-healthcare innovation: natural language processing-based patents analysis for technology-driven roadmapping. *Kybernetes*, 52(4), 1173-1189. DOI: 10.1108/k-03-2021-0170.
- Yue, D., Honghui, L., & Xueliang, F. (2021). Few-shot Learning Named Entity Recognition of Pressure Sensor Patent Text Based on MLM. *In IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)* (pp. 1-5). IEEE. DOI: 10.1109/TOCS53301.2021.9688929.
- Yue, G., Liu, J., Hou, Y., & Zhang, Q. (2023). A Novel Patent Knowledge Extraction Method for Innovative Design. *Ieee Access*, 11, 2182-2198. DOI: 10.1109/access.2022.3229490.
- Zhu, C., & Motohashi, K. (2022). Identifying the technology convergence using patent text information: a graph convolutional networks (gcn)-based approach. *Technological Forecasting And Social Change*, 176, 121477. DOI: 10.1016/j.techfore.2022.121477.