REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

ANÁLISE DE DOCUMENTOS DE PATENTES UTILIZANDO UM SISTEMA MULTIMODAL: PROPOSTA DE INTEGRAÇÃO DA ANÁLISE DE IMAGEM (IMAGÉTICO) E DE REDAÇÃO (SEMÂNTICO) DOS DADOS PATENTÁRIOS

ANALYSIS OF PATENT DOCUMENTS USING A MULTIMODAL SYSTEM: PROPOSAL FOR INTEGRATING IMAGE (VISUAL) AND TEXTUAL ANALYSIS OF PATENT DATA

ANÁLISIS DE DOCUMENTOS DE PATENTES UTILIZANDO UN SISTEMA MULTIMODAL: PROPUESTA DE INTEGRACIÓN DEL ANÁLISIS DE IMÁGENES (VISUAL) Y DE REDACCIÓN (SEMÁNTICO) DE LOS DATOS DE PATE

Thiago Domingos Marques
Mestre em Propriedade Intelectual e Inovação
Universidade Federal de Santa Catarina - UFSC, Brasil
E-mail: thiagomestradoufsc@gmail.com

Alexandre Leopoldo Gonçalves
Doutor em Engenharia de Produção
Universidade Federal de Santa Catarina - UFSC, Brasil
E-mail: a.l.goncalves@ufsc.com

Resumo

Este artigo apresenta a proposta de um protótipo para análise de similaridade multimodal de patentes, integrando informações textuais e visuais com o objetivo de aprimorar as capacidades de busca e fomentar a análise de anterioridade. A metodologia emprega TF-IDF para a geração de embeddings textuais e uma abordagem simplificada para embeddings visuais, baseada no redimensionamento, conversão para escala de cinza e achatamento das imagens. A implementação foi realizada em ambiente Python, com ingestão direta de dados a partir de links de patentes provenientes do Google Patents, permitindo a execução automatizada dos processos de extração e análise. Os embeddings gerados são combinados por meio de uma função de pontuação multimodal ponderada, possibilitando a identificação de patentes semelhantes. Adicionalmente, é construído um grafo de conhecimento básico em memória, com a finalidade de representar as relações entre os documentos analisados. Os resultados demonstram a viabilidade e os benefícios da integração de múltiplas modalidades de dados, oferecendo um sistema inicial para o aprimoramento da recuperação de patentes. Embora fundamentado em técnicas simplificadas, o trabalho estabelece bases sólidas para pesquisas futuras, incluindo a adoção de modelos de embeddings mais avançados, soluções escaláveis de busca vetorial e a utilização de bancos de dados persistentes para grafos de conhecimento.
Palavras-chave: Análise multimodal; Patentes; Similaridade; Embeddings; Recuperação da informação.

Abstract

This article presents a prototype proposal for multimodal patent similarity analysis, integrating textual and visual information to enhance search capabilities and support prior art analysis. The

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

methodology employs TF-IDF for generating textual embeddings and a simplified approach for visual embeddings, based on image resizing, grayscale conversion, and flattening. The implementation was carried out in a Python environment, with direct data ingestion from patent links obtained from Google Patents, enabling automated execution of extraction and analysis processes. The generated embeddings are combined through a weighted multimodal scoring function, allowing the identification of similar patents. Additionally, a basic in-memory knowledge graph is constructed to represent relationships among the analyzed documents. The results demonstrate the feasibility and benefits of integrating multiple data modalities, providing an initial system for improving patent retrieval. Although based on simplified techniques, this work establishes a foundation for future advancements, including the adoption of more sophisticated embedding models, scalable vector search solutions, and persistent knowledge graph databases.
Keywords:
Multimodal analysis; Patents; Textual and visual embeddings; Document similarity; Information retrieval.

Resumen

Este artículo presenta una propuesta de prototipo para el análisis de similitud multimodal de patentes, integrando información textual y visual con el objetivo de mejorar las capacidades de búsqueda y fomentar el análisis de anterioridad. La metodología emplea TF-IDF para la generación de embeddings textuales y un enfoque simplificado para embeddings visuales, basado en el redimensionamiento, la conversión a escala de grises y el aplanamiento de las imágenes. La implementación se realizó en un entorno Python, con ingestión directa de datos a partir de enlaces de patentes provenientes de Google Patents, permitiendo la ejecución automatizada de los procesos de extracción y análisis. Los embeddings generados se combinan mediante una función de puntuación multimodal ponderada, lo que permite identificar patentes similares. Además, se construye un grafo de conocimiento básico en memoria con el fin de representar las relaciones entre los documentos analizados. Los resultados demuestran la viabilidad y los beneficios de integrar múltiples modalidades de datos, ofreciendo un sistema inicial para mejorar la recuperación de patentes. Aunque se basa en técnicas simplificadas, este trabajo establece bases sólidas para futuras investigaciones, incluyendo la adopción de modelos de embeddings más avanzados, soluciones escalables de búsqueda vectorial y el uso de bases de datos persistentes para grafos de conocimiento.
Palabras clave:
Análisis multimodal; Patentes; Embeddings textuales y visuales; Similitud de documentos; Recuperación de información.

1. INTRODUÇÃO

A análise de similaridade de patentes constitui atividade central na gestão da propriedade intelectual, sustentando processos de busca de anterioridade, mapeamento tecnológico, inteligência competitiva e tomada de decisão estratégica em pesquisa e desenvolvimento. Tradicionalmente, tais análises fundamentam-se predominantemente no conteúdo textual dos documentos patentários, em especial títulos, resumos, relatórios descritivos e reivindicações por meio de técnicas clássicas de recuperação da informação, mineração de texto e processamento de linguagem natural (PLN). No entanto, patentes são, por

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

natureza, artefatos comunicacionais multimodais, nos quais descrições verbais coexistem de forma indissociável com desenhos técnicos, diagramas e figuras que expressam aspectos estruturais e funcionais das invenções.

A literatura especializada reconhece que uma parcela significativa da informação técnica relevante em documentos patentários é veiculada por representações visuais, frequentemente não capturadas de maneira satisfatória por abordagens exclusivamente textuais. Sob a perspectiva semiótica, a imagem técnica não desempenha papel meramente ilustrativo, mas atua como elemento estruturante do sentido, ancorando, complementando e, em muitos casos, desambiguando o conteúdo verbal (Barthes, 1990; Kress; Van Leeuwen, 2006).

No contexto jurídico-tecnológico da propriedade intelectual, essa articulação verbo-imagética assume especial relevância, uma vez que os desenhos integram o escopo descritivo da patente e influenciam diretamente a delimitação da proteção conferida.

O crescimento exponencial do volume de dados patentários, aliado à complexidade crescente das invenções contemporâneas, impõe desafios significativos aos sistemas tradicionais de busca e análise. Conforme apontado por Farhat e Gonçalves-Segundo (2022), o aumento expressivo dos repositórios tecnológicos demanda ferramentas analíticas mais sofisticadas, capazes de lidar com múltiplas camadas informacionais de forma integrada. Apesar disso, a maioria das soluções atualmente empregadas no exame de anterioridade e na prospecção tecnológica permanece limitada à exploração textual, desconsiderando o potencial informativo das imagens técnicas que acompanham os documentos.

Com o avanço das técnicas de aprendizado de máquina, PLN e visão computacional, emergem oportunidades para o desenvolvimento de sistemas multimodais, capazes de integrar dados textuais e visuais em um mesmo ambiente analítico. Estudos recentes demonstram que abordagens multimodais, ao projetarem diferentes modalidades de informação em espaços vetoriais compartilhados, apresentam desempenho superior em tarefas de classificação, recuperação da informação e análise de similaridade, quando comparadas a

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

métodos unimodais. Tais evidências sugerem que a análise exclusivamente textual pode ser insuficiente para capturar a complexidade informacional dos documentos patentários, sobretudo em domínios intensivos em engenharia e sistemas técnicos complexos.

Nesse contexto, o presente artigo propõe o desenvolvimento e a avaliação de um pipeline prototípico para análise multimodal de similaridade de patentes, combinando características textuais e visuais em um único modelo analítico. A abordagem integra representações textuais baseadas em técnicas de PLN com descritores visuais extraídos de imagens técnicas, fundamentando-se em uma matriz verbo-imagética de natureza semiótica. Ainda que o sistema utilize componentes deliberadamente introdutórios, o objetivo é demonstrar, de forma reprodutível e de ponta a ponta, os benefícios concretos da fusão multimodal na identificação de patentes semelhantes e na verificação de anterioridades tecnológicas.

A relevância do estudo justifica-se pela necessidade de soluções analíticas mais robustas e integradas, capazes de ampliar a capacidade interpretativa dos sistemas de busca e análise de patentes, em consonância com a Lei nº 9.279/1996 (Lei da Propriedade Industrial) e com os compromissos internacionais assumidos pelo Brasil no âmbito do Acordo TRIPS/ADPIC e das diretrizes da Organização Mundial da Propriedade Intelectual (WIPO). A adoção de uma abordagem multimodal representa, assim, um avanço metodológico e estratégico para a gestão da inovação, a mitigação de riscos de litígio e o fortalecimento da segurança jurídica em atividades de pesquisa e desenvolvimento.

A análise integrada de texto e imagem contribui, ainda, para o atendimento ao requisito de suficiência descritiva previsto na Lei nº 9.279/1996, reforçando a segurança jurídica no exame de anterioridade e na delimitação do escopo de proteção patentária.

O objetivo geral desse artigo é analisar documentos de patentes por meio de um sistema multimodal que integra dados textuais e imagéticos, fundamentado em uma matriz verbo-imagética semiótica, com vistas a aprimorar a identificação de similaridades tecnológicas e padrões discursivo-visuais.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Diante disso, este estudo propõe um pipeline multimodal para análise de similaridade de patentes, integrando *embeddings* textuais e visuais em um modelo unificado, com o objetivo de avaliar sua eficácia na identificação de anterioridade tecnológica. A principal contribuição do estudo reside na proposição de um pipeline reprodutível que integra, de forma estruturada, fundamentos semióticos à análise multimodal de documentos de patentes.

Para orientar o leitor quanto à estrutura deste artigo, e para atender ao objetivo proposto, o trabalho está organizado em seções. Inicialmente, apresenta-se esta introdução, na qual são contextualizados o tema, os objetivos e a relevância da pesquisa. Em seguida, a seção de referencial teórico discute os principais conceitos, na sequência, a metodologia detalha os procedimentos adotados. Posteriormente, são apresentados e discutidos os resultados obtidos, à luz da literatura pertinente. Por fim, as considerações finais.

2. REVISÃO DA LITERATURA

A análise de documentos de patentes exige abordagens metodológicas capazes de lidar com sua natureza técnico-jurídica e intrinsecamente multimodal. A integração entre texto e imagem atende diretamente ao requisito da suficiência descritiva, previsto no art. 24 da Lei nº 9.279/1996 (LPI), uma vez que a descrição clara, precisa e completa da invenção depende da articulação coerente entre elementos verbais e gráficos.

O escopo do artigo caracteriza-se como uma prova de conceito de natureza exploratória. Nesse contexto, a literatura recente tem enfatizado o papel dos sistemas multimodais e da inteligência artificial como instrumentos estratégicos para o exame de anterioridade, a gestão da informação tecnológica e a segurança jurídica da proteção patentária.

2.1 Sistemas multimodais e documentos patentários

Estudos apontam que os documentos de patentes constituem fontes primárias de conhecimento científico e tecnológico justamente por combinarem

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

descrições textuais detalhadas com representações visuais das invenções. Farhat e Gonçalves-Segundo (2022) destacam que a multimodalidade permite a construção de significados que não emergem da análise isolada de texto ou imagem, mas da interação entre diferentes sistemas semióticos. Assim, a análise integrada desses elementos possibilita uma compreensão mais precisa das soluções técnicas reivindicadas.

A literatura sugere que a complexidade dos documentos patentários demanda métodos analíticos capazes de considerar simultaneamente seus aspectos visuais e verbais. Sistemas multimodais têm se mostrado eficazes na extração de informações relevantes, ao facilitar a interpretação de desenhos técnicos e sua correlação com reivindicações e descrições, contribuindo tanto para a inovação tecnológica quanto para a gestão do conhecimento (Farhat; Gonçalves-Segundo, 2022).

No campo da mineração de patentes, Ji (2021) observa que a análise automatizada de grandes bases patentárias auxilia na proteção dos direitos de propriedade intelectual e no direcionamento estratégico da pesquisa científica. De forma complementar, Awale *et al.* (2025) demonstram que interfaces multimodais baseadas em texto e imagem promovem maior acessibilidade e compreensão dos documentos técnicos, superando limitações dos sistemas tradicionais centrados exclusivamente em texto.

2.2 Inteligência artificial aplicada à análise de patentes

A transformação digital pode ser compreendida como a aplicação intensiva de tecnologias digitais, como a inteligência artificial, para reconfigurar fluxos de produção de valor em diferentes setores econômicos (Vial, 2019). No âmbito da engenharia e da gestão do conhecimento, tais tecnologias têm sido incorporadas para enfrentar desafios relacionados à gestão do capital intelectual e à tomada de decisão estratégica (Kitsios; Kamariotou, 2021).

No contexto da propriedade intelectual, observa-se um crescimento expressivo de pesquisas que utilizam técnicas de aprendizado de máquina e processamento de linguagem natural para análise de patentes. Vasconcelos Lobo *et al*. (2025) destacam que essas abordagens reduzem o tempo de processamento

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

e aumentam a precisão das análises, impactando positivamente a eficiência dos sistemas patentários. Métodos como *TF-IDF*, *embeddings* semânticos, *Naive Bayes* e árvores de decisão têm sido amplamente empregados, embora apresentem limitações ao desconsiderar informações visuais relevantes (Krestel *et al.,* 2021; Haleem *et al.,* 2019).

Como já destacado, estudos recentes têm demonstrado o avanço das aplicações de inteligência artificial na análise de patentes. Técnicas de classificação de texto com base em *TF-IDF, word embeddings, Naive Bayes* e árvores de decisão vêm sendo amplamente utilizadas para extrair conhecimento de grandes volumes de documentos (Krestel *et al.*, 2021; Haleem *et al.,* 2019). Esses métodos permitem categorizar, agrupar e prever relações entre patentes de modo eficiente, mas apresentam limitações ao ignorar as informações visuais que frequentemente são essenciais à compreensão técnica das invenções.

No campo das patentes, desenvolveu-se o uso de informações textuais e informações de classificação de patentes. Primeiro, na pesquisa de texto, o texto usado na especificação do pedido de patente é usado para pesquisa. O usuário pesquisa a literatura do estado da técnica usando vários dicionários de sinônimos e termos técnicos. Embora seja possível realizar uma ampla gama de pesquisas, é essencial ter conhecimento especializado na criação de consultas de pesquisa (Higuchi; Yanai, 2023).

Em particular, as tecnologias de pesquisa baseadas em palavras-chave e classificações de patentes têm sido comumente usadas no campo de patentes e, com o recente desenvolvimento do processamento de linguagem natural, a pesquisa e o desenvolvimento no campo de patentes deram um grande salto à frente. No entanto, questões ainda precisam ser abordadas com relação aos desenhos de patentes (Higuchi; Yanai, 2023).

A análise de imagens em documentos de patentes, por sua vez, tem avançado com o uso de redes neurais convolucionais, capazes de identificar padrões estruturais nos desenhos técnicos (Xu *et al.*, 2023; Zhang *et al.*, 2022). Pesquisas em aprendizado multimodal indicam que a combinação de *embeddings* textuais e visuais resulta em maior robustez analítica e maior acurácia na

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

identificação de similaridades tecnológicas (Li *et al.*, 2024).

2.3 Inteligência artificial generativa em buscas de patentes na propriedade intelectual

A inteligência artificial generativa (*GenAI*) destaca-se por sua capacidade de criar conteúdos novos a partir do aprendizado de padrões estatísticos em grandes volumes de dados multimodais (Volpe, 2025). Modelos baseados em *deep learning* (DL) ampliam as possibilidades de automação e apoio à análise técnica e jurídica, mas também suscitam debates relevantes quanto ao uso de dados protegidos e à geração de conteúdos sintéticos.

A literatura aponta que a eficácia da GenAI depende diretamente da qualidade e representatividade dos dados utilizados em seu treinamento, podendo gerar vieses ou limitações caso esses dados sejam insuficientes ou pouco diversificados (Volpe, 2025). Nesse cenário, abordagens como a geração aumentada por recuperação (*Retrieval-Augmented Generation – RAG*) têm sido propostas para mitigar problemas como alucinações, ao integrar fontes externas de conhecimento e aumentar a confiabilidade das respostas (Naveed et al., 2023; Yager, 2023; Anjos, 2025).

Após o surgimento da *GenAI*, seu uso tornou-se mais amplo, o Escritório de Patentes e Marcas dos Estados Unidos (*United States Patent and Trademark Office – USPTO*) anunciou o lançamento do *Artificial Intelligence Search Automated Pilot (ASAP*), uma iniciativa voltada à aplicação de técnicas de inteligência artificial (IA) na condução de buscas de anterioridade (*prior art*) em pedidos de patentes utilitárias, antes do início do exame substantivo. O programa tem como objetivo avaliar o potencial da IA para ampliar a eficiência, a abrangência e a precisão da fase pré-exame, oferecendo aos depositantes uma visão antecipada e estruturada das referências tecnológicas mais relevantes ao pedido (Abapi, 2025).

A ferramenta interna de IA desenvolvida pelo *USPTO* realiza a análise integral do conteúdo do pedido, incluindo a especificação descritiva, as reivindicações e o resumo, bem como as classificações atribuídas segundo o sistema *Cooperative Patent Classification (CPC).* Com base nessa análise

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

contextual e semântica, o sistema executa de forma automatizada buscas em extensos repositórios públicos de informação patentária, abrangendo patentes concedidas nos Estados Unidos, publicações de pedidos ainda não concedidos e documentos de origem estrangeira. Como resultado, o sistema identifica e hierarquiza até dez documentos considerados mais relevantes, ordenados de acordo com o grau de pertinência técnica em relação ao objeto reivindicado (Abapi, 2025).

A *GenAI* está em constante evolução, com novos modelos e técnicas sendo desenvolvidos regularmente. Ela tem o potencial não apenas de automatizar tarefas criativas, mas também de abrir novas possibilidades para a criatividade humana, oferecendo ferramentas que podem inspirar e ampliar a expressão criativa. A favor da Inteligência Artificial Generativa, argumenta-se que ela pode gerar novas ideias em forma de texto e em forma de arte, além de proporcionar o aprendizado sobre diferentes culturas e suas perspectivas. (Hessel; Lemes, 2023)

No entanto, o desempenho da recuperação de imagens de patentes, mesmo usando aprendizado métrico profundo, permaneceu insatisfatório e nenhum método ou sistema de fato apareceu no campo de patentes. A razão é que os desenhos de patentes são descritos como desenhos abstratos em preto e branco, que diferem significativamente das imagens naturais em termos de modalidade, e não é fácil alcançar o desempenho desejado por uma simples aplicação de métodos convencionais. Além disso, o campo de patentes geralmente requer o conhecimento de especialistas em propriedade intelectual (advogados de patentes, pesquisadores de patentes) para reconhecer e entender desenhos. A oferta (número de especialistas) é pequena em relação à demanda, e construir um novo conjunto de dados para aprendizado profundo é caro e requer uma enorme quantidade de tempo e esforço (Higuchi; Yanai, 2023).

Nessa linha, Pustu-Iren, Bruns e Ewerth (2021) propõem uma abordagem multimodal para a recuperação semântica de imagens de patentes, combinando características visuais e textuais. Imagens de patentes, como desenhos técnicos, contêm informações valiosas e são frequentemente utilizadas por especialistas para comparar patentes. No entanto, as abordagens atuais de recuperação de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

informações de patentes concentram-se amplamente em dados textuais.

Por conseguinte, Pustu-Iren, Bruns e Ewerth (2021) destacam que a integração de características textuais e visuais em um sistema multimodal de recuperação de patentes possibilita uma busca semântica mais precisa, explorando simultaneamente as informações contidas nas imagens e nos textos dos documentos. Segundo os autores, a combinação de modalidades — como a análise de figuras por redes neurais profundas e o reconhecimento de texto — amplia significativamente a capacidade de identificar similaridades semânticas entre patentes, demonstrando que abordagens multimodais superam métodos unidimensionais (Pustu-Iren; Bruns; Ewerth, 2021).

Dessa forma, a literatura evidencia que a convergência entre inteligência artificial, aprendizado multimodal e fundamentos semiótico-jurídicos constitui uma base teórica sólida para o desenvolvimento de sistemas avançados de análise de patentes, capazes de refletir a natureza híbrida dos documentos patentários e atender às exigências normativas do regime de propriedade intelectual.

2.4 Recuperação multimodal e busca por anterioridade

A maioria dos sistemas de busca de patentes ainda se apoia predominantemente em texto e classificações, como IPC e CPC, o que subutiliza a natureza multimodal desses documentos (Higuchi; Yanai, 2023). Embora eficazes, tais métodos exigem elevado conhecimento especializado e apresentam limitações na interpretação direta de desenhos técnicos.

Avanços recentes buscam suprir essas lacunas por meio da recuperação multimodal. Awale *et al.* (2025) apresentam o sistema *iPatent*, que integra análise textual e visual por meio de modelos de aprendizado profundo, permitindo buscas unimodais, cruzadas e multimodais, além de oferecer visualizações interativas. Resultados semelhantes são observados por Pustu-Iren, Bruns e Ewerth (2021), que demonstram que a combinação de modalidades amplia a precisão da recuperação semântica. Já Shukla *et al.* (2025) reforçam que modelos multimodais especializados no domínio patentário incorporam elementos estruturais próprios das figuras técnicas, permitindo interpretações mais fidedignas das invenções.

Por fim, Consoloni, Giordano e Fantoni (2024) acrescentam que métricas de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

coerência entre texto e imagem são fundamentais para avaliar e calibrar bases de dados multimodais, contribuindo para sistemas mais confiáveis.

2.5 Lacuna de Pesquisa e Contribuição do Estudo

Apesar dos avanços recentes na aplicação de técnicas de processamento de linguagem natural, aprendizado de máquina e inteligência artificial à análise de documentos patentários, a literatura ainda apresenta lacunas significativas no que se refere à integração efetiva e sistemática das modalidades textual e imagética em modelos unificados de análise de similaridade. Nesse diapasão, destaca-se que grande parte dos estudos concentra-se em abordagens unimodais, predominantemente textuais, ou em aplicações pontuais de visão computacional dissociadas de uma fundamentação semiótica e jurídico-normativa consistente, o que limita a capacidade interpretativa dos sistemas e sua aderência às exigências da suficiência descritiva no contexto da propriedade intelectual.

Ademais, observa-se a escassez de propostas que apresentem *pipelines* multimodais completos, reprodutíveis e transparentes, capazes de articular, de forma integrada, extração de características, indexação, fusão de modalidades e mecanismos de explicabilidade.

Diante desse cenário, o presente estudo contribui ao propor e demonstrar um *pipeline* prototípico de análise multimodal de similaridade de patentes, fundamentado em uma matriz verbo-imagética de natureza semiótica, que integra informações textuais e visuais em um modelo analítico unificado.

Ao fazê-lo, o trabalho avança metodologicamente ao evidenciar os ganhos interpretativos e analíticos da fusão multimodal, oferecendo subsídios teóricos e práticos para o aprimoramento de sistemas de busca de anterioridade, gestão da informação tecnológica e fortalecimento da segurança jurídica no âmbito da propriedade intelectual.

2.6 Síntese da Revisão Metodológica

Em síntese, a revisão da literatura evidencia uma evolução metodológica que parte de abordagens textuais tradicionais, avança para métodos baseados em

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

IA e culmina em modelos multimodais, os quais representam uma nova fronteira na análise de patentes. A integração entre texto e imagem mostra-se essencial para capturar a complexidade técnica dos documentos patentários, corroborando as conclusões de Li *et al.* (2024) e consolidando o papel da inteligência artificial multimodal na gestão da propriedade intelectual.

Logo, por conseguinte, a abordagem proposta confirma as conclusões ao evidenciar que modelos multimodais representam uma nova fronteira na inteligência aplicada à propriedade intelectual, unindo precisão analítica à contextualização técnica e semântica.

A Figura 01 apresenta visualmente esses eixos teóricos, evidenciando a convergência entre propriedade intelectual, engenharia, gestão do conhecimento e aprendizado multimodal apoiado por inteligência artificial

Figura 01 - Convergência entre fundamentos jurídicos e avançadas técnicas de análise integrada de documentos patentários.

Fonte: Autores (2026).

Ao representar graficamente a articulação entre o arcabouço jurídico-normativo — especialmente a Lei de Propriedade Industrial e o exame de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

anterioridade — e as abordagens computacionais contemporâneas de fusão textual e imagética, conforme Figura 01, reforça-se a compreensão de que a análise de patentes demanda métodos capazes de refletir sua natureza híbrida e complexa.

Logo, nessa linha, a literatura especializada indica que a combinação de análises semânticas e visuais proporciona ganhos expressivos de abrangência e precisão na avaliação de similaridade entre patentes, evidenciando que modelos contrastivos multimodais são particularmente eficazes na redução de vieses e na priorização de casos com elevado grau de similaridade técnica (Morais; Dias, 2025).

Dessa forma, a literatura recente destaca que abordagens unimodais apresentam limitações na análise de patentes, sobretudo por não capturarem adequadamente a complementaridade entre texto e imagem. Técnicas como TF-IDF, embeddings semânticos e classificadores tradicionais têm sido amplamente utilizadas na análise textual, enquanto avanços em visão computacional, como redes neurais convolucionais e modelos contrastivos (ex.: CLIP), permitem extrair características visuais de maior nível semântico.

3. METODOLOGIA

Esta seção apresenta e discute os principais métodos de análise de patentes identificados na literatura, abrangendo abordagens tradicionais baseadas em texto, métodos apoiados em inteligência artificial, técnicas de análise de imagens e, mais recentemente, abordagens multimodais que integram informações textuais e visuais.

Ressalta-se que este tipo de estudo enquadra-se em uma pesquisa aplicada com experimentação computacional voltada à modelagem e prototipagem de um sistema multimodal na área de recuperação da informação e mineração de patentes, com ênfase em aprendizado multimodal e propriedade intelectual. Ferramentas de apoio à escrita foram utilizadas de forma instrumental, sem interferência na autoria científica.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

3.1 Abordagem metodológica

A pesquisa caracteriza-se como aplicada, de natureza quantitativa e qualitativa, com objetivos exploratórios e explicativos. Adota-se um desenho metodológico baseado em experimentação computacional e análise interpretativa semiótica.

3.2 Materiais

O corpus da pesquisa é composto por um conjunto de documentos de patentes extraídos do *Google Patentes*, contemplando um *Dataset* com relatório descritivo, reivindicações, resumo e desenhos técnicos. Os dados textuais são obtidos em formato estruturado, enquanto as imagens correspondem aos desenhos anexos aos pedidos de patente (Figura 02).

No que se refere aos recursos computacionais, utilizam-se linguagens e bibliotecas voltadas à ciência de dados, aprendizado de máquina, visão computacional e processamento de linguagem natural. A Figura 02 destaca a fonte de dados e sua respectiva proveniência.

Figura 02 - Fonte Publica dos Documentos das Patentes.

```
--- Table of Extracted Abstracts ---
```

	URL	Abstract
0	https://patents.google.com/patent/BRPI0722238B...	method and apparatus for identifying wheel mod...
1	https://patents.google.com/patent/EP2818340B1/...	Abstract not found
2	https://patents.google.com/patent/US10181010B2...	A genetic surveillance system comprises a comm...
3	https://patents.google.com/patent/KR102451328B...	The present invention relates to a plant-based...
4	https://patents.google.com/patent/KR102729750B...	Provided are skin-protecting agents and extern...
5	https://patents.google.com/patent/BR2020170206...	colar protetor para pescoÃ§o contra linha de c...
6	https://patents.google.com/patent/BR2020230038...	Trata-se a presente criaÃ§Ã£o de um aperfeiÃ§o...
7	https://patents.google.com/patent/BRMU8900159U...	CAPACETE COM DISPOSITIVO PROTETOR CONTRA LINHA...

Fonte: Autores (2026).

Ressalta-se que o número total de patentes analisadas é reduzido, o que, no entanto, não compromete os objetivos do estudo, uma vez que o foco do trabalho reside na operacionalização e execução do sistema para comparação entre patentes similares. Nesse contexto, priorizou-se a diversidade tecnológica do

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

conjunto analisado, em detrimento da quantidade absoluta de documentos, adotando-se critérios de seleção mais apropriados à heterogeneidade do corpus.

Ademais, destaca-se que a inclusão de um volume elevado de patentes excessivamente semelhantes poderia comprometer a análise de similaridade, tanto textual quanto imagética, ao introduzir vieses e inflar artificialmente os resultados. Assim, a estratégia adotada privilegia a diversidade de patentes como forma de garantir maior robustez interpretativa e validade analítica dos experimentos realizados.

3.3 Métodos

Inicialmente, os textos das patentes são submetidos a pré-processamento, incluindo normalização, remoção de ruído, *tokenização* e vetorização por meio de técnicas como *TF-IDF* e modelos baseados em embeddings semânticos.

Na etapa seguinte, os vetores textuais e imagéticos são integrados em um espaço multimodal unificado. Essa integração é orientada por uma matriz verbo-imagética semiótica, que considera categorias como complementaridade, redundância, ancoragem e relevo semântico entre texto e imagem. Conforme Figura 03 é gerada uma matriz de similaridade.

Figura 03 - Similaridade entre as Patentes baseados em *embeddings*.

Textual Similarity Matrix (Cosine Similarity):

URL	BRPI0722238B1	EP2818340B1	US10181010B2	KR102451328B1	KR102729750B1	BR202017020609U2	BR202023003802U2	BRMU8900159U2
BRPI0722238B1	1.00	0.20	0.20	0.17	0.00	0.00	0.00	0.00
EP2818340B1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
US10181010B2	1.00	0.20	0.15	0.11	0.00	0.00	0.00	0.00
KR102451328B1	1.00	0.17	0.17	0.11	0.02	0.01	0.00	0.00
KR102729750B1	1.00	0.20	0.17	0.15	0.00	0.00	0.00	0.00
BR202017020609U2	1.00	0.48	0.48	0.00	0.00	0.00	0.00	0.00
BR202023003802U2	1.00	0.52	0.48	0.01	0.00	0.00	0.00	0.00
BRMU8900159U2	1.00	0.52	0.48	0.02	0.00	0.00	0.00	0.00

Fonte: Autores (2026).

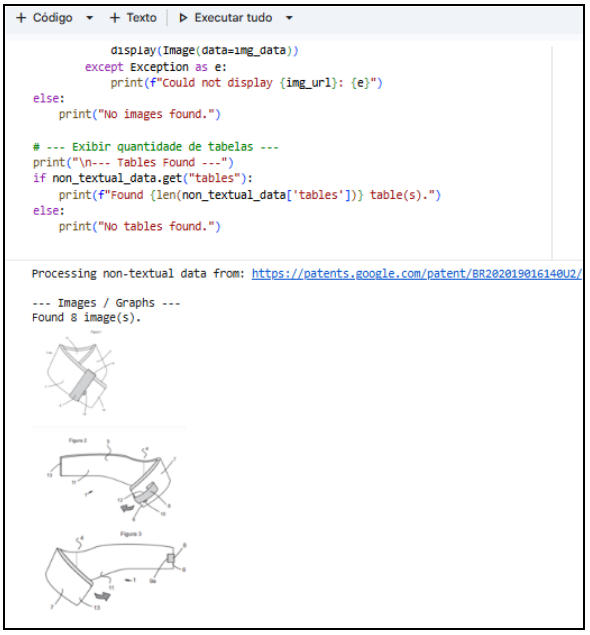
Por fim, aplicam-se algoritmos de agrupamento e similaridade para avaliar a proximidade entre documentos, bem como análises comparativas entre o desempenho do sistema multimodal e abordagens unimodais. Na Figura 04,

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

destaca-se o processo de extração de dados e a geração de imagens pelo sistema em execução, o que assegura a veracidade e a autenticidade das informações, permitindo sua verificação por meio de observação visual (Figura 04).

Figura 04 - Parte do Código de Similaridade entre com Imagens de Patentes.



Fonte: Autores (2026).

Complementarmente, ferramentas computacionais baseadas em IA foram utilizadas como apoio instrumental para aprimoramentos e revisão dos códigos.

3.4 Metodologia da Pesquisa - Literatura

Nesta pesquisa, adota-se a revisão integrativa da literatura, compreendida como um subtipo de revisão sistemática, com o objetivo de mapear o estado da arte e identificar as soluções já existentes relacionadas ao problema investigado. Tal abordagem permite a consolidação do conhecimento científico acumulado, fornecendo subsídios teóricos e metodológicos para o desenvolvimento do artefato proposto.

Para a obtenção do conhecimento necessário ao projeto do artefato, a revisão integrativa foi estruturada com base na abordagem metodológica de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Botelho, Cunha e Macedo (2011), a qual compreende as seguintes etapas sequenciais: *(a) identificação do tema e formulação da questão de pesquisa; (b) definição dos critérios de inclusão e exclusão dos estudos; (c) identificação dos estudos pré-selecionados e selecionados; (d) categorização dos estudos incluídos; (e) análise e interpretação dos resultados; e (f) apresentação da revisão, por meio da síntese do conhecimento produzido.* O seguimento dessas etapas assegura a transparência, a rastreabilidade e a reprodutibilidade da revisão conduzida.

A revisão da literatura constitui um instrumento fundamental para a compreensão aprofundada de um determinado campo de estudo, permitindo ao pesquisador construir uma base conceitual sólida que sustenta o desenvolvimento da pesquisa. Trata-se, portanto, de uma etapa inicial e indispensável no processo de construção do conhecimento científico. Conforme destacam Botelho, Cunha e Macedo (2011), a revisão da literatura pode ser classificada em revisão narrativa e revisão bibliográfica sistemática. A revisão narrativa caracteriza-se pela descrição e discussão do desenvolvimento ou do estado da arte de um tema sob uma perspectiva teórica ou contextual; contudo, por não adotar procedimentos rigorosos de sistematização, apresenta limitações quanto à replicabilidade e à resposta quantitativa aos questionamentos da pesquisa (Figura 05).

Em contraste, a revisão bibliográfica sistemática distingue-se pela adoção de métodos explícitos, planejados, sistematizados e passíveis de reprodução, com o objetivo de sintetizar evidências provenientes da literatura primária de forma estruturada e confiável (Botelho; Cunha; Macedo, 2011). Nesse escopo, Whittemore e Knafl (2005) categorizam as revisões sistemáticas em quatro tipos principais: *meta-análise, revisão qualitativa, revisão sistemática e revisão integrativa*, conforme ilustrado na Figura 05.

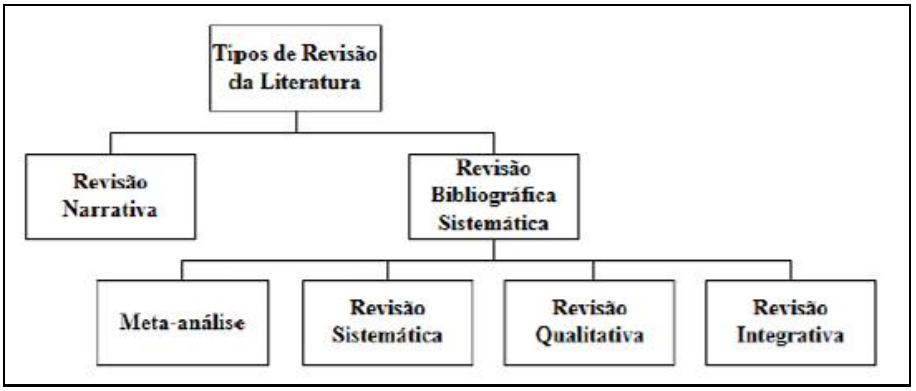
Segundo Whittemore e Knafl (2005), a revisão integrativa configura-se como a modalidade mais abrangente entre os tipos de revisão sistemática da literatura, uma vez que permite a incorporação simultânea de estudos empíricos e teóricos. Tal característica possibilita não apenas a apresentação do estado da arte sobre determinado tema, mas também a compreensão ampliada do fenômeno investigado, o apoio ao desenvolvimento de novas teorias e a orientação de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

aplicações práticas do conhecimento, justificando sua adoção como método nesta pesquisa (Figura 05).

Figura 05 – Tipos de Revisão da Literatura.



Fonte: Botelho, Cunha e Macedo (2011).

Os desafios identificados e categorizados ao longo desta pesquisa foram mapeados por meio de uma revisão integrativa da literatura, a qual possibilitou uma compreensão aprofundada dos problemas existentes no domínio investigado. Essa etapa configurou-se como fundamental para o delineamento da pesquisa e para o desenvolvimento de artefatos, especialmente em estudos de natureza tecnológica, nos quais a fundamentação teórica sólida orienta a concepção de soluções inovadoras.

A condução da revisão integrativa contemplou as etapas de análise e interpretação dos resultados, bem como a apresentação da síntese do conhecimento, conforme preconizado na literatura metodológica. Segundo Whittemore e Knafl (2005), a revisão integrativa da literatura exige a formulação de uma questão de pesquisa claramente definida, a adoção de métodos explícitos e sistematizados e o emprego de uma estratégia de busca abrangente, capaz de contemplar fontes relevantes e representativas do campo de estudo.

A etapa de implementação foi realizada utilizando a linguagem de programação *Python*, escolhida por sua ampla adoção em contextos acadêmicos e industriais, bem como pela disponibilidade de *frameworks*, bibliotecas e ferramentas especializadas que facilitam o desenvolvimento de soluções voltadas à análise de dados, aprendizado de máquina e processamento multimodal (Anjos,

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

2025).

4. RESULTADOS E DISCUSSÃO

Em um cenário de sobrecarga de análise e limitação de recursos humanos especializados, a adoção dessa tecnologia representa um avanço na eficiência e na padronização dos processos, contribuindo para reduzir prazos, otimizar decisões e aumentar a assertividade na proteção de ativos intangíveis. Apesar dessas restrições, os resultados obtidos indicam que o sistema tem potencial significativo para apoiar profissionais de propriedade intelectual na análise técnica de anterioridade, contribuindo para maior eficiência e assertividade nos processos de proteção de ativos intangíveis (Aguiar *et al.,* 2025).

4.1 Arquitetura Geral - Sistema e Processamento

O *pipeline* de similaridade multimodal é composto por quatro etapas principais: *(i) pré-processamento e embedding textual, (ii) embedding simplificado de imagens, (iii) indexação por similaridade em cada modalidade e (iv) fusão multimodal por meio de uma função de escore ponderada.* A arquitetura foi concebida de forma modular, permitindo a substituição futura de componentes por modelos mais sofisticados.

Dessa forma, é percorrida várias etapas-chave, como o pré-processamento e incorporação de dados textuais usando TF-IDF, a incorporação simples de dados de imagem, a construção de índices de similaridade para cada modalidade usando *NearestNeighbors*, e por fim uma função de pontuação multimodal que combina as pontuações de similaridade de texto, imagens e metadados (códigos IPC) em uma única pontuação de relevância. A arquitetura é projetada para ser extensível, permitindo a futura integração de modelos de incorporação e mecanismos de indexação mais sofisticados.

4.1.1 Pré-processamento e *Embedding* Textual

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Os textos das patentes — títulos, resumos e reivindicações — são concatenados e submetidos à vetorização por *TFIDF*, utilizando o *TfidfVectorizer* da *biblioteca scikit-learn.* Adota-se um vocabulário limitado (máx. 256 termos) e n-grams de ordem 1 e 2, visando capturar termos simples e expressões frequentes. Os vetores resultantes são normalizados pela norma L2, assegurando que a similaridade do cosseno reflita exclusivamente a orientação vetorial.

Nessa linha, os textos, passam por pré-processamento para convertê-los em uma representação numérica adequada para comparação de similaridade (TF-IDF/Term Frequency-Inverse Document Frequency), uma medida estatística que avalia o quão relevante uma palavra é para um documento em uma coleção de documentos. Especificamente, o TfidfVectorizer de sklearn.feature_extraction.text é usado.

Essa normalização garante que o comprimento dos vetores não influencie os cálculos de similaridade de cosseno, permitindo que a similaridade seja baseada puramente no ângulo entre os vetores.

4.1.2 *Embedding* de Imagens

Para fins de demonstração, emprega-se uma técnica simples e computacionalmente leve de representação visual. As imagens associadas às patentes são redimensionadas para 32×32 *pixels*, convertidas para tons de cinza e achatadas em vetores unidimensionais. Em seguida, aplica-se normalização L2. Embora limitada do ponto de vista semântico, essa abordagem ilustra o conceito de incorporação de informação visual ao pipeline.3.

Ocorre, assim, Pré-processamento e Incorporação de Imagem, o *array* de *pixels* em escala de cinza é então achatado em um vetor unidimensional. Finalmente, de forma semelhante às incorporações de texto, este vetor de imagem passa por normalização L2 para garantir o dimensionamento consistente. Este método fornece uma representação básica, porém eficaz, para avaliações iniciais de similaridade de imagem, demonstrando o conceito multimodal sem depender de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

modelos de visão pré-treinados.

4.1.3 Função de Geração de Embeddings Visuais

O núcleo do processo de representação visual é a função *image_embedding_simple*, responsável pela geração dos *embeddings* de imagem. Essa função recebe como entrada um caminho de arquivo ou um objeto do tipo PIL *Image* e executa uma sequência de etapas padronizadas: conversão da imagem para tons de cinza, redimensionamento para um tamanho fixo de 32 × 32 *pixels* e transformação da matriz bidimensional resultante em um vetor unidimensional.

Em seguida, o vetor é normalizado pela norma L2, assegurando que todos os embeddings apresentem o mesmo comprimento e sejam comparáveis por meio da similaridade do cosseno. Trata-se de uma abordagem deliberadamente simples, adequada para fins demonstrativos e para cenários em que a distinção visual geral é suficiente.

Contudo, reconhece-se que esse método não captura padrões visuais complexos, como aqueles extraídos por arquiteturas de *deep learning* (DL), a exemplo de CLIP ou *Vision Transformers*, o que abre espaço para aprimoramentos futuros.

4.1.4 Extração e Organização das Imagens de Patentes

Na etapa de configuração, foi criado um diretório temporário destinado ao armazenamento das imagens extraídas dos PDFs. Em seguida, definiu-se uma lista contendo as URLs dos documentos de patente analisados. O código percorre cada URL, realizando o download do PDF, sua abertura por meio da biblioteca *PyMuPDF* e a varredura de todas as páginas do documento.

Para cada página, todas as imagens incorporadas são extraídas e salvas localmente, com nomes de arquivo que incluem o identificador do PDF de origem, o número da página e o índice da imagem. Paralelamente, são armazenados metadados associados a cada imagem, como o documento de origem e sua

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

localização, viabilizando rastreabilidade e análises posteriores.

4.1.5 Construção dos *Embeddings* e do Índice de Similaridade

Após a extração completa das imagens, o código percorre o conjunto de arquivos gerados, aplicando a função de *embedding* visual a cada imagem.

Os vetores resultantes são então empilhados, formando uma matriz de *embeddings* que constitui a base para a análise de similaridade. Sobre essa matriz, é construído um índice de vizinhos mais próximos utilizando o algoritmo *NearestNeighbors*, configurado para identificar até cinco vizinhos por consulta e adotando a métrica de distância do cosseno, adequada para vetores normalizados.

O modelo é ajustado à matriz de *embeddings*, criando uma estrutura eficiente para buscas de similaridade.

4.1.6 Indexação por Similaridade

A busca por vizinhos mais próximos é realizada por meio do algoritmo *NearestNeighbors*, com métrica de distância do cosseno, implementado na *biblioteca scikit-learn*. Índices independentes são construídos para os *embeddings* textuais e visuais, possibilitando a recuperação eficiente dos documentos mais semelhantes em cada modalidade.

Índices separados são construídos para as incorporações de texto e as incorporações de imagem. O parâmetro *metric* é definido como *'cosine'*, o que significa que o modelo calcula a distância de cosseno entre os vetores. A distância de cosseno está diretamente relacionada à similaridade de cosseno (similaridade = 1 - distância), tornando-a adequada para dados de alta dimensão onde a magnitude dos vetores não é tão importante quanto sua orientação. Cada índice é configurado para encontrar um número especificado de vizinhos (por exemplo, n_neighbors=10), permitindo a recuperação rápida das patentes mais textualmente e visualmente semelhantes a uma determinada consulta.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

4.1.7 Bibliotecas Utilizadas e Ambiente Computacional

A implementação do pipeline computacional proposto foi realizada no ecossistema *Python*, em virtude de sua ampla adoção em pesquisas científicas envolvendo processamento de dados, visão computacional e aprendizado de máquina. Foram empregadas bibliotecas consolidadas e amplamente referenciadas na literatura, cada qual responsável por uma etapa específica do fluxo metodológico, garantindo reprodutibilidade, modularidade e eficiência ao processo.

Inicialmente, a biblioteca requests foi utilizada para a coleta automatizada dos documentos, possibilitando o *download* de arquivos *PDF* diretamente a partir de *URLs* públicas. Em seguida, empregou-se a biblioteca fitz (*PyMuPDF*) para a abertura, leitura e processamento dos arquivos *PDF,* sendo responsável pela extração sistemática das imagens contidas em cada página dos documentos patentários.

Para o tratamento e a manipulação das imagens, foi adotada a biblioteca *PIL.Image (Pillow)*, utilizada para operações como abertura, conversão de formatos, padronização e redimensionamento, assegurando a uniformidade necessária à etapa de representação vetorial. O módulo *io* foi empregado no gerenciamento de fluxos de *bytes* provenientes tanto dos arquivos *PDF* quanto das imagens extraídas, permitindo o processamento eficiente de dados em memória.

As operações numéricas e a construção de vetores e matrizes foram realizadas com o auxílio da biblioteca *numpy*, fundamental para a representação matemática dos dados visuais e para o suporte às etapas subsequentes de análise. Para a identificação de similaridade entre imagens, foi utilizada a classe *NearestNeighbors*, pertencente à biblioteca *scikit-learn*, responsável pela busca das imagens mais próximas no espaço vetorial, com base em métricas de distância previamente definidas.

A manipulação de caminhos e estruturas de arquivos foi conduzida por meio da biblioteca pathlib.Path, que proporciona uma abordagem orientada a objetos para o gerenciamento de diretórios e arquivos, contribuindo para a portabilidade e organização do código. Por fim, a biblioteca *IPython.display.display* foi utilizada

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

para a visualização direta dos resultados, permitindo a exibição das imagens recuperadas no ambiente de notebook e facilitando a análise qualitativa das similaridades identificadas.

O conjunto dessas bibliotecas viabiliza a construção de um pipeline integrado e reprodutível, abrangendo desde a aquisição dos dados brutos até a análise e visualização dos resultados, alinhando-se às boas práticas metodológicas em pesquisas aplicadas à análise multimodal de documentos patentários.

Tabela 01 – Bibliotecas Utilizadas no Pipeline Computacional

Item	**Biblioteca / Módulo**	**Função no Pipeline Metodológico**
(a)	*requests*	Realiza o download automatizado dos arquivos PDF a partir de URLs públicas, viabilizando a coleta dos documentos de patentes.
(b)	*fitz (PyMuPDF)*	Responsável pela abertura, leitura dos documentos PDF e extração das imagens contidas em cada página.
(c)	*PIL.Image (Pillow)*	Utilizada para a manipulação das imagens extraídas, incluindo abertura, conversão de formato, padronização e redimensionamento.
(d)	*io*	Empregada no tratamento de fluxos de bytes provenientes dos arquivos PDF e das imagens, permitindo o processamento eficiente em memória.
(e)	*numpy*	Utilizada para operações numéricas, bem como para a construção e manipulação de vetores e matrizes representativas dos dados visuais.
(f)	*NearestNeighbors (scikit-learn)*	Aplicada na identificação das imagens mais semelhantes no espaço vetorial, com base em métricas de distância previamente definidas.
(g)	*pathlib.Path*	Responsável pela manipulação de caminhos de arquivos e diretórios de forma orientada a objetos, contribuindo para a organização e portabilidade do código.
(h)	*IPython.display.display*	Utilizada para a exibição direta das imagens e dos resultados no ambiente de notebook, facilitando a análise qualitativa.

Fonte: Autores (2026).

4.2 Análise e Apresentação dos Resultados

Cada imagem extraída é tratada como uma imagem de consulta. Para cada uma delas, o modelo identifica as imagens mais semelhantes, convertendo a distância de cosseno em um *score* de similaridade. A própria imagem de consulta é

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

excluída dos resultados, e os pares similares são ordenados de acordo com o grau de similaridade. Os resultados são apresentados de duas formas complementares.

No aspecto textual, o sistema exibe, para cada imagem de consulta, uma lista detalhada contendo as imagens mais similares, seus respectivos documentos de origem, páginas e *scores* de similaridade. Valores próximos de 1,0 indicam elevada similaridade visual.

No aspecto visual, para um subconjunto das imagens analisadas, são exibidos pares formados pela imagem de consulta e sua imagem mais semelhante, permitindo uma verificação qualitativa direta da eficácia do método.

4.2.1 Execução Experimental e Resultados Obtidos

No experimento conduzido, diferentes documentos de patente foram processados, resultando na extração de 15 imagens. Durante o procedimento, uma imagem foi ignorada em razão de um aviso de segurança relacionado ao seu tamanho excessivo, o que evidencia a robustez do código ao lidar com exceções. Foram gerados *embeddings* para todas as imagens válidas, construído o índice de vizinhos mais próximos e realizada a comparação cruzada entre as figuras.

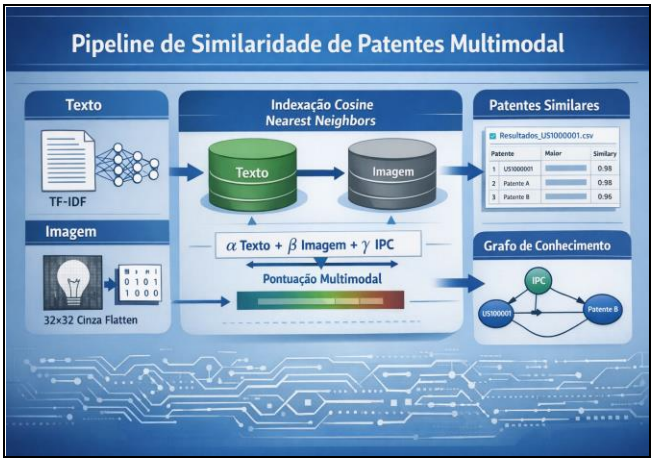
Os resultados indicaram altos níveis de similaridade entre diversas imagens, fenômeno esperado em desenhos de patentes, que frequentemente apresentam múltiplas vistas ou variações mínimas de um mesmo objeto. A inspeção visual confirmou a coerência dos resultados obtidos, validando o funcionamento do pipeline proposto.

Dessa forma, o código extraiu com sucesso as representações visuais dos documentos PDF e usou uma técnica simples para quantificar e exibir a similaridade entre elas. A Figura 06 sintetiza as etapas do processo, iniciando pela análise e vetorização dos dados textuais e imagéticos provenientes do *dataset* selecionado. Destaca-se que os dados foram obtidos de forma estruturada, a partir do acesso a arquivos completos disponíveis na base pública do *Google Patents*, garantindo transparência e reprodutibilidade ao experimento.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 06 – Pipeline de Similaridade



Fonte: Autores (2026).

O código executado automatizou a análise de similaridade visual entre imagens extraídas de documentos PDF de patentes. Primeiramente, ele baixou e processou dois PDFs, extraindo todas as imagens e salvando-as localmente. Em seguida, para cada imagem, gerou uma representação numérica simplificada (*embedding*) ao redimensioná-la para tons de cinza e achatar seus *pixels*.

Finalmente, utilizou esses *embeddings* para encontrar e listar as imagens mais semelhantes entre si, apresentando os resultados tanto em formato textual com *scores* de similaridade quanto visualmente, exibindo as imagens originais e suas melhores correspondências.

4.2.2 Escore Multimodal

A integração das modalidades é realizada por meio de uma combinação linear ponderada, na qual as similaridades textuais e visuais são agregadas a um componente adicional baseado em metadados. Essa abordagem permite unificar diferentes fontes de informação em um único escore de relevância, possibilitando maior flexibilidade analítica conforme o contexto de aplicação (Figura 07):

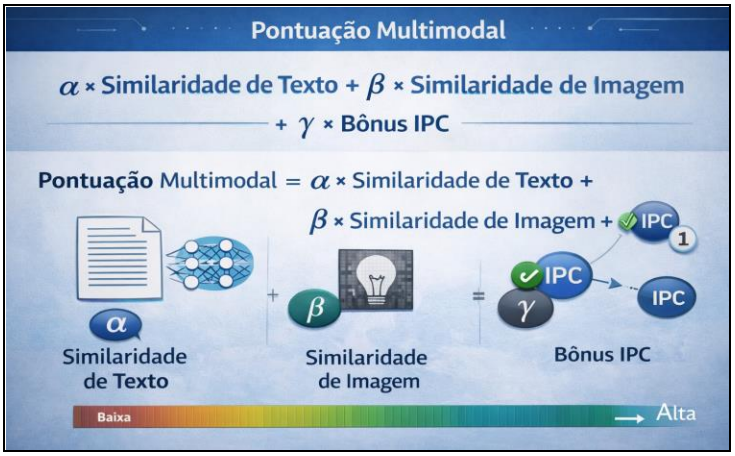
$$Score\ Total = \alpha \cdot S_{texto} + \beta \cdot S_{imagem} + \gamma \cdot B_{IPC}$$

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Conforme Figura 07, nessa formulação, *Stexto* e *Simagem* representam, respectivamente, as similaridades de cosseno obtidas a partir dos embeddings textuais e visuais. O termo BIP C corresponde a um bônus binário associado à coincidência dos códigos da Classificação Internacional de Patentes (IPC), assumindo valor 1 quando há correspondência e 0 caso contrário. Os parâmetros α, β e γ controlam a contribuição relativa de cada componente no escore final, podendo ser ajustados de acordo com os objetivos da análise, o domínio tecnológico e as características do conjunto de dados.ada componente, podendo ser ajustados conforme o contexto de uso.

Figura 07 - Pontuação Multimodal



Fonte: Autores (2025).

Pontuação Total = (α × Similaridade Textual) + (β × Similaridade de Imagem) + (γ × Bônus IPC)

No caso, observa-se que, onde: α (alpha) representa o peso atribuído à similaridade textual; β (beta) representa o peso atribuído à similaridade visual; γ (gamma) representa o peso associado ao bônus de Classificação Internacional de Patentes (IPC), implementado como um indicador binário (valor 1 quando os códigos IPC coincidem e 0 caso contrário). Esse bônus tem por finalidade favorecer o ranqueamento de patentes que compartilham a mesma classificação tecnológica.

O ajuste dos parâmetros α, β e γ permite calibrar a importância relativa de cada modalidade de informação, possibilitando a adaptação do modelo a diferentes

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

objetivos de busca, domínios tecnológicos ou características específicas dos documentos patentários.

A imagem ilustra de forma didática o funcionamento da pontuação multimodal de similaridade de patentes, evidenciando como diferentes fontes de informação são integradas em um único escore final.

O diagrama apresenta a fórmula central em que a similaridade textual, a similaridade visual e o bônus de classificação IPC são combinados por meio de pesos ajustáveis (α, β e γ), indicando a contribuição relativa de cada modalidade.

A similaridade de texto representa o grau de proximidade semântica entre títulos, resumos e reivindicações, enquanto a similaridade de imagem captura aspectos visuais presentes nos desenhos e figuras das patentes. O bônus IPC atua como um reforço semântico adicional quando os documentos compartilham a mesma classe tecnológica.

A escala cromática inferior, variando de baixa a alta, simboliza o aumento progressivo da relevância à medida que essas dimensões convergem, reforçando a ideia de que a integração multimodal produz uma avaliação mais robusta e informativa do que abordagens baseadas em uma única fonte de dados.

O cerne da abordagem multimodal reside na utilização de uma função de combinação linear ponderada que integra, de forma sinérgica, as pontuações de similaridade provenientes de diferentes modalidades. Quando uma consulta é realizada — seja por meio de um identificador de patente, um texto descritivo ou o caminho de uma imagem — o pipeline computa, de maneira independente, as similaridades textuais e visuais utilizando índices do tipo *Nearest Neighbors*.

Adicionalmente, incorpora-se um componente baseado em metadados, fundamentado nos códigos IPC (Classificação Internacional de Patentes), que atua como um fator de reforço semântico. A pontuação final agregada de cada patente candidata é, portanto, obtida por meio da expressão: Pontuação Total = (α × Similaridade de Texto) + (β × Similaridade de Imagem) + (γ × Bônus IPC), em que α, β e γ representam, respectivamente, os pesos atribuídos às modalidades textual, visual e ao bônus IPC — este último modelado como um indicador binário que assume valor 1 em caso de correspondência de classificação e 0 caso contrário.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Tal mecanismo favorece a priorização de documentos com alinhamento classificatório, ao mesmo tempo em que permite, por meio do ajuste dos parâmetros α, β e γ, calibrar dinamicamente a relevância relativa de cada modalidade, adaptando o modelo a diferentes contextos de busca e às especificidades dos domínios tecnológicos analisados.

4.2.3 Configuração Experimental

Para validação do protótipo, foi criado um conjunto de dados de patentes reais, cada uma com identificador, título, resumo, reivindicações, ano, imagem associada e código IPC. As imagens são geradas programaticamente, contendo texto e imagens.

Complementarmente, foram empregadas ferramentas de Inteligência Artificial (IA) para aprimorar e apoiar tanto a redação quanto a estruturação do artigo. Destacam-se, nesse contexto, a utilização do *Gemini*, da *Google*, e do *ChatGPT,* desenvolvido pela *OpenAI,* que contribuíram significativamente no suporte textual e metodológico.

4.2.4 Protótipo desenvolvido

Inicialmente, o documento de patente é decomposto em dois grandes componentes informacionais: o (1) texto patentário (relatório descritivo, reivindicações e resumo) e os (2) desenhos técnicos. Esses elementos são processados de maneira independente por módulos especializados, responsáveis pela extração de características textuais e visuais. No plano textual, aplicam-se técnicas de processamento de linguagem natural para geração de representações vetoriais; no plano imagético, utilizam-se métodos de visão computacional para identificar padrões estruturais relevantes dos desenhos.

Na etapa seguinte, as informações extraídas convergem para o modelo multimodal, no qual ocorre a fusão dos vetores textuais e visuais. Essa integração é mediada por uma matriz verbo-imagética, que representa conceitualmente as

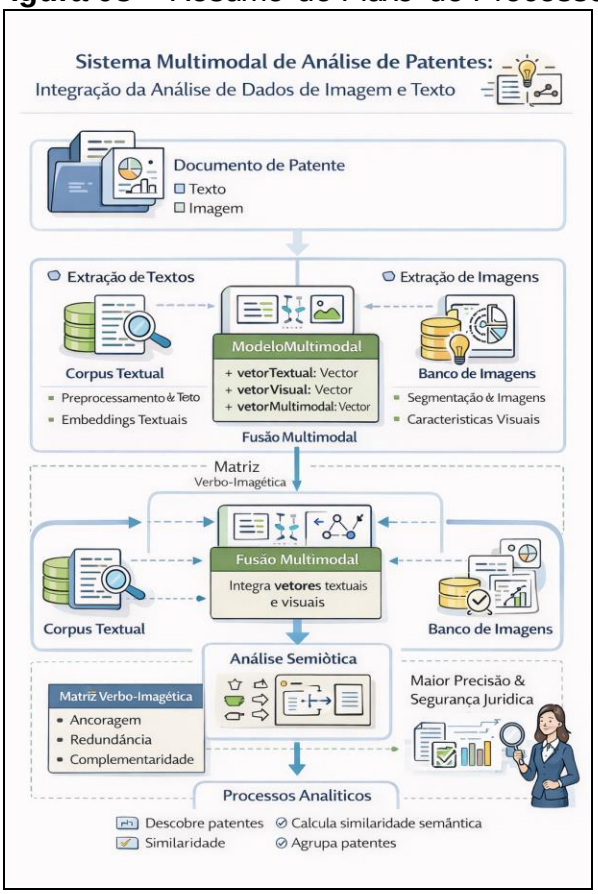
REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

relações semióticas entre texto e imagem, tais como ancoragem, complementaridade e redundância. A partir dessa fusão, o sistema passa a operar sobre uma representação unificada do documento patentário, possibilitando o cálculo de similaridade, o agrupamento de patentes e a identificação de proximidades tecnológicas de forma mais precisa e coerente.

A Figura 08, abaixo apresentada, destaca o fluxo de etapas do sistema multimodal de análise de patentes, evidenciando a sequência lógica de processamento desde a entrada dos dados até a geração dos resultados analíticos.

Figura 08 – Resumo do Fluxo do Processo



Fonte: Autores (2026).

A importância de sistemas dessa natureza reside no fato de que os documentos de patentes são, por definição, artefatos técnico-jurídicos multimodais. Ao integrar texto e imagem em um único processo analítico, o sistema amplia a capacidade de apoio à análise de anterioridade, à vigilância tecnológica e à tomada

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

de decisão em propriedade intelectual. Além disso, tais sistemas contribuem para maior segurança jurídica e eficiência institucional, ao alinhar métodos computacionais avançados às exigências normativas da legislação patentária e às práticas dos escritórios de patentes, como o INPI e a WIPO.

4.2.5 Resultados do protótipo

A execução do fluxo de demonstração (demo_flow) gera embeddings textuais e visuais, constrói os índices de similaridade e realiza uma consulta exemplificativa a partir de uma patente de referência. Os resultados são exportados em arquivo CSV, apresentando as patentes mais similares e seus respectivos escores multimodais.

Adicionalmente, é construído um pequeno Grafo de Conhecimento em memória, utilizando a biblioteca NetworkX, no qual as patentes são representadas como nós e as relações de compartilhamento de códigos IPC como arestas. Tal estrutura ilustra como relações semânticas podem complementar a análise baseada em similaridade vetorial.

4.2.6 Análise de Similaridade Multimodal (Demo Sintética)

A função *demo_flow()* orquestra todo o pipeline do protótipo. Ela começa chamando *create_sample_dataset()* e *generate_sample_images()* para configurar o ambiente com os dados sintéticos. Em seguida, pré-processa os campos de texto combinados (título, resumo, reivindicações) usando *TfidfVectorizer* para gerar *embeddings* de texto e calcula *embeddings* de imagem em escala de cinza achatados para cada patente. Esses embeddings são então usados para construir índices NearestNeighbors para as modalidades de texto e imagem.

O cerne da demonstração é a função *query_similar_patents()*, que é executada com *patent_id="US1000001*" como consulta. Esta função calcula uma pontuação de similaridade multimodal para outras patentes em relação à patente de consulta. A pontuação é uma combinação ponderada (alpha para texto, beta

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

para imagem, gamma para bônus de metadados baseado no código IPC compartilhado) de métricas de similaridade individuais. A função identifica e classifica patentes semelhantes, retornando os k principais resultados. Esses resultados, incluindo suas pontuações calculadas, são então salvos em *query_results_US1000001.csv.*

Além disso, um pequeno Grafo de Conhecimento (KG) em memória é construído usando *NetworkX* através da função *build_simple_kg()*.

Este KG representa patentes como nós e inclui atributos básicos como título, ano e IPC. As arestas são adicionadas entre patentes que compartilham a mesma classificação IPC, demonstrando uma relação simples dentro do conjunto de dados. Este KG gerado é então salvo em *kg_patents_demo.gpickle*, oferecendo uma representação estruturada das relações de patentes além das pontuações de similaridade simples.

4.2.7 Análise de Similaridade Textual (Embeddings TF-IDF)

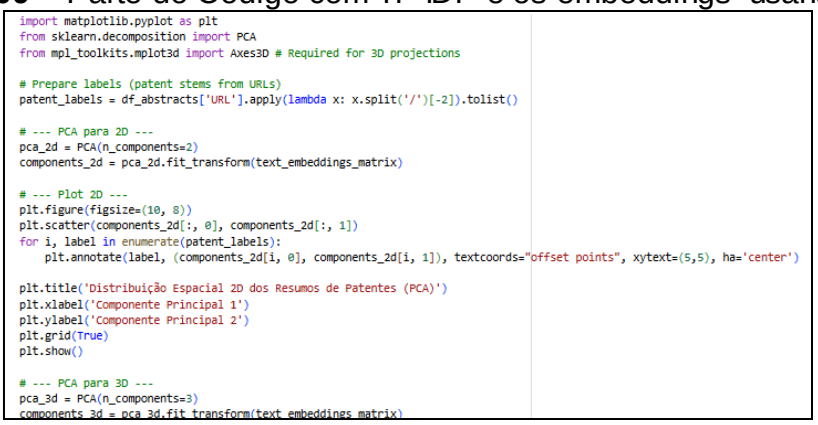
Os resumos de patentes reais foram processados com TF-IDF e os embeddings resultantes foram analisados usando PCA para visualização da similaridade espacial.

Na Figura 08, destaca-se parte do código segmentado responsável pela avaliação dos dados e pela geração de gráficos. Esses gráficos representam as patentes em espaços bidimensional e tridimensional, permitindo a visualização das proximidades entre elas e facilitando a análise das similaridades de forma intuitiva.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 09 – Parte do Código com TF-IDF e os embeddings usando PCA



```
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
from mpl_toolkits.mplot3d import Axes3D # Required for 3D projections

# Prepare labels (patent stems from URLs)
patent_labels = df_abstracts['URL'].apply(lambda x: x.split('/')[-2]).tolist()

# --- PCA para 2D ---
pca_2d = PCA(n_components=2)
components_2d = pca_2d.fit_transform(text_embeddings_matrix)

# --- Plot 2D ---
plt.figure(figsize=(10, 8))
plt.scatter(components_2d[:, 0], components_2d[:, 1])
for i, label in enumerate(patent_labels):
    plt.annotate(label, (components_2d[i, 0], components_2d[i, 1]), textcoords="offset points", xytext=(5,5), ha='center')

plt.title('Distribuição Espacial 2D dos Resumos de Patentes (PCA)')
plt.xlabel('Componente Principal 1')
plt.ylabel('Componente Principal 2')
plt.grid(True)
plt.show()

# --- PCA para 3D ---
pca_3d = PCA(n_components=3)
components_3d = pca_3d.fit_transform(text_embeddings_matrix)
```

Fonte: Autores (2026).

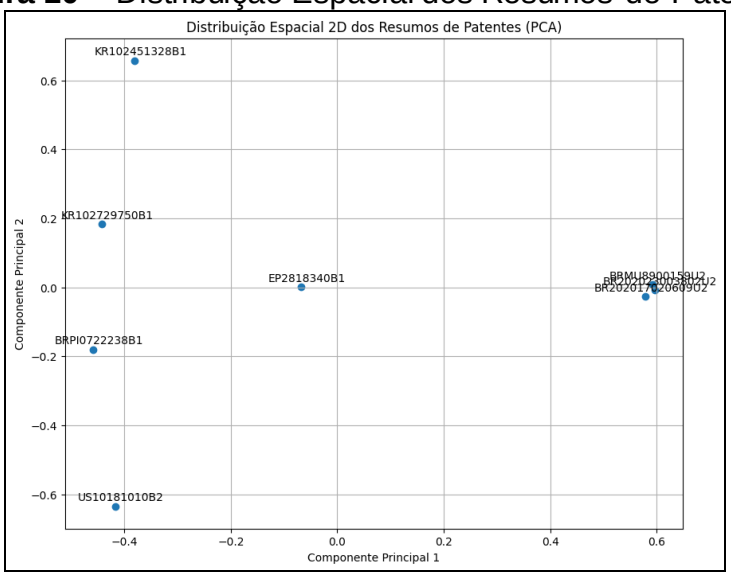
A matriz de similaridade de cosseno evidenciou elevados níveis de proximidade entre determinados documentos patentários, destacando-se valores máximos de similaridade (1,00) para pares como EP2818340B1 e BRPI0722238B1, bem como US10181010B2 e BRPI0722238B1. Tais resultados sugerem, de forma consistente, a possibilidade de que esses documentos correspondam a membros de uma mesma família de patentes, depositados em diferentes jurisdições ou versões linguísticas, compartilhando conteúdo técnico substancialmente equivalente.

Esse comportamento reforça a capacidade do método TF-IDF em identificar similaridades lexicais significativas, especialmente em casos de textos idênticos ou altamente correlacionados após o pré-processamento. Contudo, também indica a necessidade de análise complementar, uma vez que similaridades máximas podem decorrer de redundâncias textuais ou estruturas documentais padronizadas.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 10 – Distribuição Espacial dos Resumos de Patentes



Fonte: Autores (2026).

Conforme Marques e Gonçalves (2025), a visualização de dados de forma gráfica facilita a identificação das similaridades entre patentes, ao permitir uma análise visual integrada de cada documento em relação aos demais. Por meio dessas representações, torna-se possível compreender, de maneira mais intuitiva e fundamentada em dados concretos, os padrões de proximidade e agrupamento existentes no conjunto analisado, contribuindo para uma interpretação mais clara e consistente das relações de similaridade.

4.2.8 Análise de Texto - Similaridades dos Resumos

Foi realizada uma análise de similaridade entre os resumos das patentes selecionadas. Inicialmente, procedeu-se à coleta dos *URLs* de todas as patentes processadas no sistema, seguida do reprocessamento automatizado em ambiente Python, com a extração dos respectivos resumos, de modo a assegurar a padronização e a organização dos dados.

Na etapa subsequente, aplicou-se a técnica de *Term Frequency–Inverse Document Frequency* (TF-IDF) para a geração de representações vetoriais dos textos. Por fim, empregou-se o algoritmo *Nearest Neighbors* para o cálculo da

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

similaridade entre os resumos. Os resultados obtidos são apresentados em um formato sintético e de fácil interpretação.

A matriz de similaridade calculada é uma representação numérica de quão semelhantes são os resumos de cada patente entre si, baseada na análise *TF-IDF* e na similaridade de cosseno.

É observado que a patente 'BRPI0722238B1' tem uma similaridade de 1.00 com ela mesma, mas uma similaridade de 0.20 com 'EP2818340B1', e assim por diante. No entanto, ao analisar a seção "Most similar pairs" abaixo da tabela, notamos que a maioria das patentes mostra uma similaridade de 1.00 com 'BRPI0722238B1'. Isso é um achado significativo! Sugere fortemente que essas patentes podem ser variações da mesma invenção, depositadas em diferentes escritórios de patentes ou em diferentes idiomas, ou parte de uma mesma família de patentes.

O TF-IDF, mesmo sendo uma técnica mais básica, conseguiu capturar essa relação quando os textos eram idênticos ou quase idênticos após o pré-processamento (especialmente porque uma patente teve seu resumo como "Abstract not found", o que pode ter levado a um vetor zero e, portanto, 100% de similaridade com outros vetores zero ou quase zero, dependendo da implementação exata de normalize para vetores muito esparsos ou nulos).

Em resumo, a matriz permite visualizar rapidamente quais resumos de patentes são mais parecidos entre si, e os resultados apontam para uma alta interconexão textual entre a maioria das patentes analisadas e a patente 'BRPI0722238B1'.

Observando os pares de patentes mais similares com base na análise de similaridade textual, os pares de patentes mais similares (excluindo a similaridade da patente consigo mesma) são:

'BRPI0722238B1' é mais similar a **'EP2818340B1'** com similaridade: **0.20**
'EP2818340B1' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'US10181010B2' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'KR102451328B1' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'KR102729750B1' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'BR202017020609U2' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'BR202023003802U2' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**
'BRMU8900159U2' é mais similar a **'BRPI0722238B1'** com similaridade: **1.00**

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Como pode ser visto, a maioria das patentes demonstrou uma similaridade perfeita (1.00) com a patente BRPI0722238B1, o que sugere que elas podem ser patentes muito semelhantes ou até membros da mesma família de patentes, possivelmente devido a resumos idênticos ou muito próximos em conteúdo. Abaixo exemplos de três resumos das patentes mais similares a BRPI0722238B1:

> **Patent ID: BR202017020609U2** URL: Abstract: colar protetor para pescoço contra linha de cerol. refere-se a um colar protetor que tem por objetivo proteger o pescoço dos usuários de motocicletas e bicicletas de linhas com cerol utilizadas em pipas ou também conhecidas como maranhão, papagaio de papel, raia e outros ou quaisquer brinquedos que voem baseados na oposição entre a força do vento e a corda segurada pelo operador. constituído por peça única dotada em sua face interna (1) de lona de aço e sua face externa (2) é composta por espuma e malha, as extremidades do colar são produzidas em velcro (3).
>
> **Patent ID: BR202023003802U2** URL: Abstract: Trata-se a presente criação de um aperfeiçoamento de modelo de protetor de pescoço O material utilizado é uma fita metálica flexível com recortes que servem como instrumento de corte para as linhas de cerol e assemelhadas. Esta fita é montada em um dispositivo composto de eixo e molas que permite que a fita seja retrátil, acionada pelo motociclista conforme seu interesse. O conjunto funciona semelhante ao funcionamento das trenas métricas, com os devidos mecanismos de travas e destrava. Uma vez em uso, destravada a fita de corte, sua parte inferior deve ser embutida sob a vestimenta do motociclista, evitando, assim, que a linha passe por baixo da fita.
>
> **Patent ID: BRMU8900159U2** URL:Abstract: CAPACETE COM DISPOSITIVO PROTETOR CONTRA LINHA DE PIPA E CONGÊNERE, consiste de um capacete (1) fabricado em material adequado atendendo as exigências e normas, o qual é dotado de uma haste (2) basculante pivotante (3) nas Laterais frontais do capacete (1) propriamente dito, que permanece à frente do motociclista protegendo a região superior do tórax e pescoço de linhas de pipas que com o contato com as seções (4) afiadas são imediatamente cortadas.

Como pode-se observar, os resumos dos 'protetores de pescoço' descrevem dispositivos de segurança para motociclistas ou ciclistas, especificamente contra linhas de cerol ou itens cortantes. Eles focam na proteção da região do pescoço, seja através de um colar ou de um dispositivo acoplado ao capacete. Por outro lado, o resumo da patente 'BRPI0722238B1' (identificação de módulo de roda) descreve um método e aparelho para identificar a posição de módulos em rodas de veículos para monitoramento. O foco aqui é em tecnologia

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

de sensores, comunicação e processamento de dados para localizar componentes no sistema da roda de um veículo.

Semanticamente, os tópicos são bastante distintos, o que é consistente com o baixo score de similaridade (0.20) que 'BRPI0722238B1' apresentou em relação a 'EP2818340B1' (que tem 'Abstract not found' e um histórico de similaridade de 1.00 com outras patentes de pneu/roda), e também com a similaridade de 1.00 que as patentes de protetor de pescoço mostraram entre si em suas respectivas análises de similaridade.

4.2.9 Visualização PCA 2D e 3D dos Resumos de Patentes (TF-IDF)

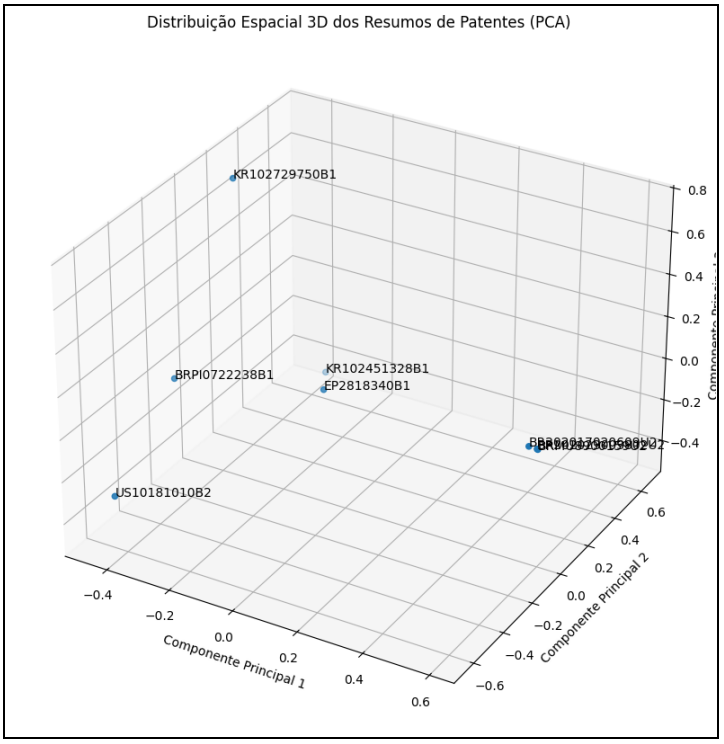
O gráfico de dispersão 2D das componentes principais evidencia a distribuição espacial dos resumos de patentes, permitindo observar relações de similaridade textual entre os documentos analisados. Patentes posicionadas próximas no plano apresentam maior proximidade semântica, enquanto aquelas mais distantes indicam menor relação textual. Nota-se a formação de agrupamentos bem definidos, os quais sugerem a existência de similaridades temáticas claras entre determinados conjuntos de patentes, ao passo que a dispersão de outros pontos revela heterogeneidade no conteúdo analisado.

Por sua vez, a representação tridimensional (3D) amplia a capacidade analítica ao oferecer uma perspectiva adicional sobre a estrutura dos embeddings. Na Figura 11 ´demonstrado que essa visualização permite identificar clusters que, por vezes, não são facilmente perceptíveis no plano bidimensional, enriquecendo a interpretação dos dados. Assim, ambos os gráficos desempenham papel fundamental na identificação de padrões de similaridade, contribuindo para tarefas como agrupamento automático e descoberta de relações latentes entre documentos de patentes.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 11 - Gráfico 3D dos Resumos de Patentes (PCA)



Fonte: Autores (2026).

4.2.10 Visualização Top 5 Pares Mais Semelhantes (TF-IDF PCA 2D)

Os cinco pares de patentes mais semelhantes, identificados com base na projeção em PCA 2D dos embeddings TF-IDF, foram: BR202023003802U2 e BRMU8900159U2 (distância: 0,0169); BR202017020609U2 e BR202023003802U2 (0,0246); BR202017020609U2 e BRMU8900159U2 (0,0360); BRPI0722238B1 e KR102729750B1 (0,3637); e EP2818340B1 e KR102729750B1 (0,4163).

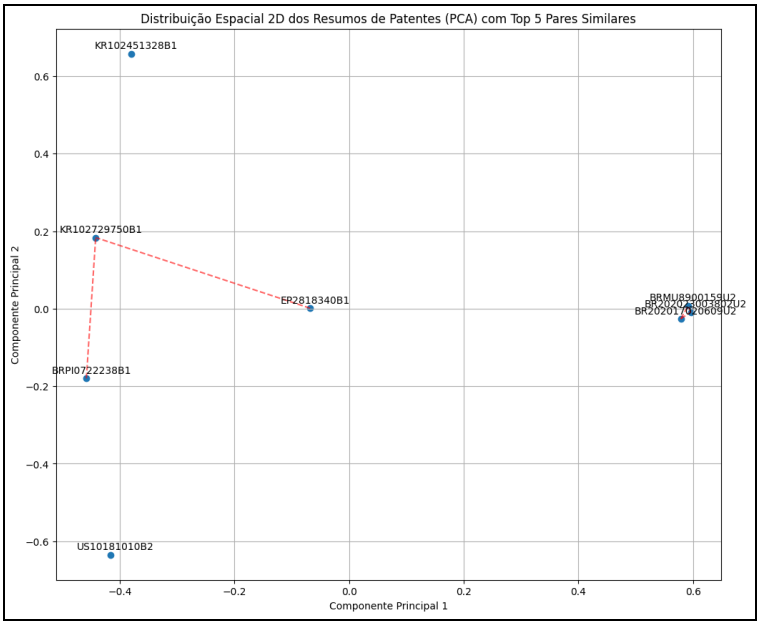
Esses pares, representados no gráfico bidimensional, evidenciam as relações de maior proximidade textual entre os documentos analisados. As menores distâncias observadas indicam forte similaridade semântica, especialmente entre as primeiras patentes, que formam um agrupamento coeso.

Tais resultados corroboram a eficácia do modelo TF-IDF na captura de padrões lexicais relevantes em descrições de patentes, validando sua aplicação como método inicial para análise de similaridade textual em bases patentárias (Figura 12).

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 12 - Top 5 Pares Mais Semelhantes - Gráfico 2D de Patentes (PCA)



Fonte: Autores (2026).

Para identificar e visualizar os 5 pares de imagens mais similares, é calculada as distâncias euclidianas entre as imagens no espaço 2D da PCA. Em seguida, é mostrado o gráfico de dispersão 2D com linhas conectando os 5 pares mais próximos, e também é listada esses pares com suas distâncias. Por conseguinte, o gráfico 2D mostrando os 5 pares de imagens mais similares é gerado com sucesso, e os pares são identificados conforme figura abaixo:

Top 5 pares de imagens mais similares (com base na distância Euclidiana do PCA 2D):

BRMU8900159U2_img1 e BRMU8900159U2_img1 (Distância: 0.0009)
BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0011)
BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0017)
BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0017)
BR202017020609U2_img1 e BR202017020609U2_img1 (Distância: 0.0023)

No gráfico, você verá linhas vermelhas tracejadas conectando esses pares. É notável que as distâncias são extremamente pequenas, indicando uma altíssima similaridade visual entre essas imagens. Vários dos pares envolvem a mesma imagem consigo mesma (ou cópias muito próximas) ou entre imagens de diferentes documentos, mas que o sistema considerou quase idênticas devido à simplicidade

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

do método de embedding de imagem utilizado. A consistência nos IDs das imagens sugere que essas imagens podem ser variações ou re-apresentações das mesmas figuras em diferentes contextos ou patentes relacionadas.

Podemos observar claramente que as três primeiras patentes, todas relacionadas a 'protetor de pescoço', estão muito próximas umas das outras, com distâncias mínimas, formando um cluster bem definido. As outras duas, embora também consideradas entre as 5 mais similares, possuem distâncias maiores, indicando uma similaridade textual mais baixa em comparação com o grupo de 'protetor de pescoço'.

4.2.11 Análise de Similaridade de Imagens (Embeddings Simples)

As imagens extraídas dos PDFs de patentes foram processadas por meio de uma abordagem de embedding simples, que consistiu no redimensionamento para 32×32 pixels em escala de cinza, seguido do achatamento dos dados e normalização dos vetores. Esses embeddings foram posteriormente submetidos à Análise de Componentes Principais (PCA), com o objetivo de reduzir a dimensionalidade e viabilizar a visualização da similaridade entre as imagens.

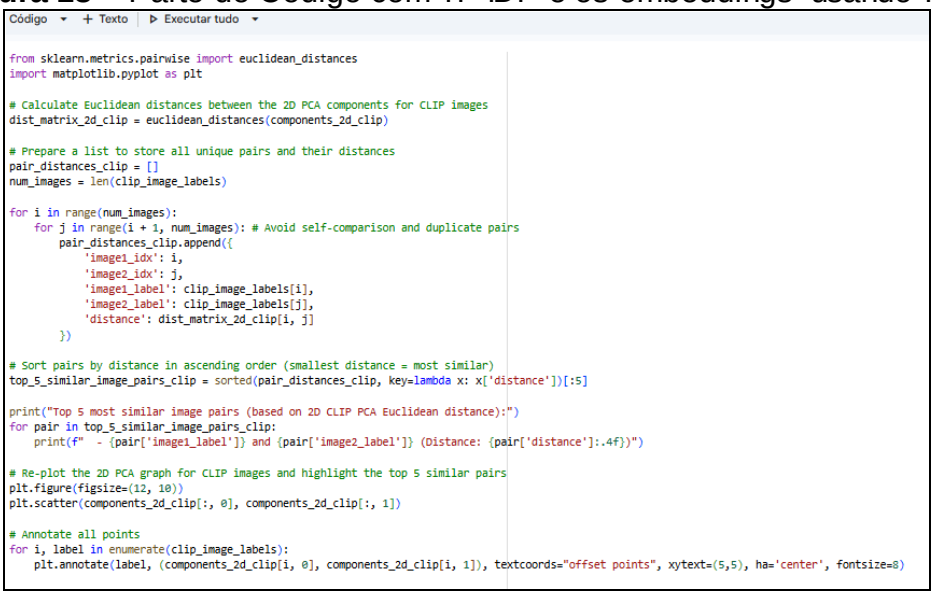
No gráfico bidimensional (2D), observou-se a distribuição das imagens no espaço reduzido, permitindo identificar possíveis agrupamentos visuais, ainda que com menor nível de detalhamento, em razão da simplicidade da técnica empregada.

Na Figura abaixo é destacado parte do código que gera a lista de similaridade entre as diferentes patentes (Figura 13).

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 13 – Parte do Código com TF-IDF e os embeddings usando PCA



```
Código  + Texto  ▷ Executar tudo

from sklearn.metrics.pairwise import euclidean_distances
import matplotlib.pyplot as plt

# Calculate Euclidean distances between the 2D PCA components for CLIP images
dist_matrix_2d_clip = euclidean_distances(components_2d_clip)

# Prepare a list to store all unique pairs and their distances
pair_distances_clip = []
num_images = len(clip_image_labels)

for i in range(num_images):
    for j in range(i + 1, num_images): # Avoid self-comparison and duplicate pairs
        pair_distances_clip.append({
            'image1_idx': i,
            'image2_idx': j,
            'image1_label': clip_image_labels[i],
            'image2_label': clip_image_labels[j],
            'distance': dist_matrix_2d_clip[i, j]
        })

# Sort pairs by distance in ascending order (smallest distance = most similar)
top_5_similar_image_pairs_clip = sorted(pair_distances_clip, key=lambda x: x['distance'])[:5]

print("Top 5 most similar image pairs (based on 2D CLIP PCA Euclidean distance):")
for pair in top_5_similar_image_pairs_clip:
    print(f"  - {pair['image1_label']} and {pair['image2_label']} (Distance: {pair['distance']:.4f})")

# Re-plot the 2D PCA graph for CLIP images and highlight the top 5 similar pairs
plt.figure(figsize=(12, 10))
plt.scatter(components_2d_clip[:, 0], components_2d_clip[:, 1])

# Annotate all points
for i, label in enumerate(clip_image_labels):
    plt.annotate(label, (components_2d_clip[i, 0], components_2d_clip[i, 1]), textcoords="offset points", xytext=(5,5), ha='center', fontsize=8)
```

Fonte: Autores (2026).

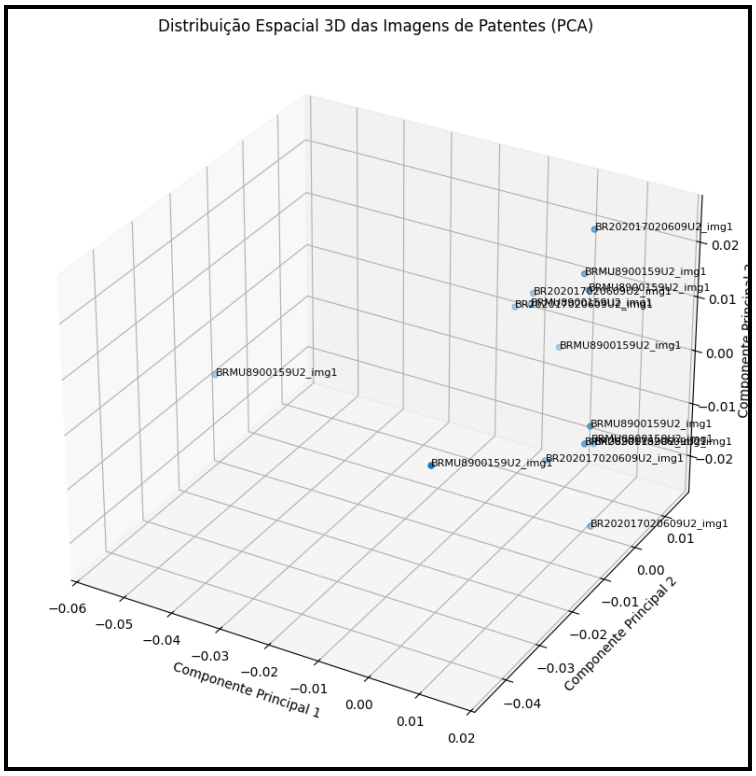
A representação tridimensional (3D), por sua vez, complementou a análise ao oferecer uma perspectiva adicional da estrutura dos dados, possibilitando uma percepção mais aprofundada de eventuais padrões de agrupamento. Embora os embeddings utilizados não capturem características visuais complexas, como ocorre em modelos baseados em *deep learning*, os resultados demonstram a utilidade da PCA como ferramenta exploratória (Figura 14).

Dessa forma, mesmo com uma abordagem simplificada, foi possível evidenciar o potencial da visualização para análise de similaridade entre imagens de patentes.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 14 - Gráfico 3D das Imagens de Patentes (PCA)



Fonte: Autores (2026).

A análise dos embeddings simples projetados em PCA 2D permitiu identificar os cinco pares de imagens mais semelhantes, evidenciando distâncias extremamente reduzidas entre determinados elementos. Destaca-se que o par formado por BRMU8900159U2_img1 com ele próprio apresentou a menor distância (0,0009), seguido pela similaridade entre BRMU8900159U2_img1 e BR202017020609U2_img1, com distâncias de 0,0011 e 0,0017, indicando forte proximidade visual entre imagens provenientes de diferentes documentos de patente.

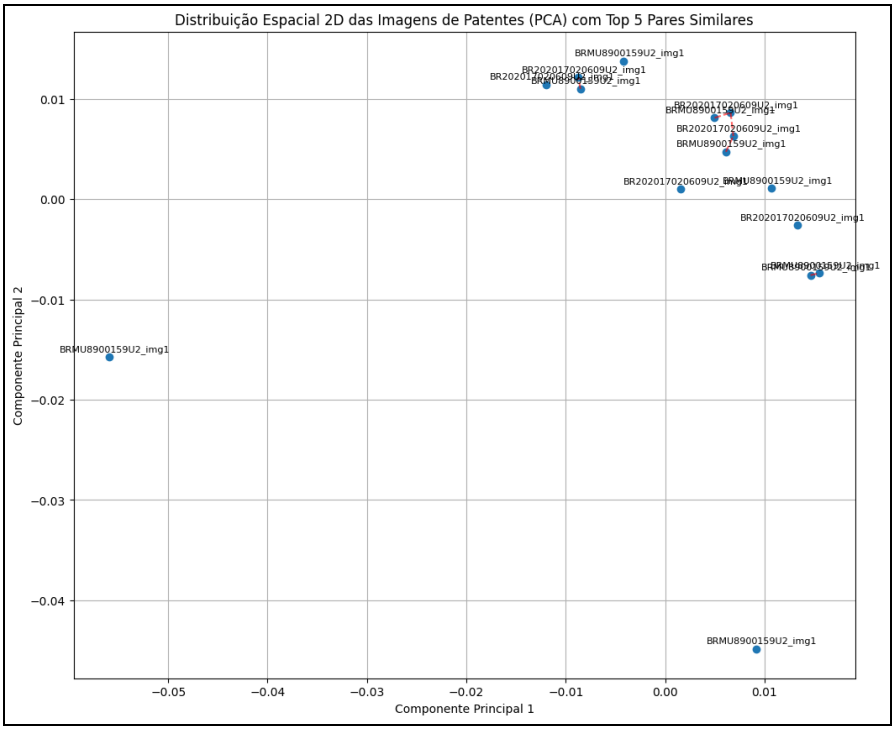
Observa-se ainda a recorrência desses pares entre os mais similares, bem como a presença de imagens do mesmo documento, como BR202017020609U2_img1 comparada a si própria (0,0023). Esses resultados reforçam a capacidade do método em capturar similaridades visuais, mesmo a partir de embeddings simplificados, evidenciando padrões consistentes entre figuras técnicas analisadas.Top 5 Pares Mais Semelhantes (Embeddings Simples

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

PCA 2D): BRMU8900159U2_img1 e BRMU8900159U2_img1 (Distância: 0.0009); BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0011); BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0017); BRMU8900159U2_img1 e BR202017020609U2_img1 (Distância: 0.0017); BR202017020609U2_img1 e BR202017020609U2_img1 (Distância: 0.0023) (Figura 15)

Figura 15 - Gráfico 2D das Imagens de Patentes (PCA)



Fonte: Autores (2026).

Os resultados encontrados indicam que os *embeddings* podem identificar imagens quase idênticas ou muito semelhantes, mas sua capacidade de capturar semelhanças conceituais complexas é limitada.

4.2.12 Análise de Similaridade de Imagens (Embeddings CLIP)

As imagens de patentes também foram codificadas usando o modelo CLIP, gerando embeddings de 512 dimensões. Esses embeddings, que capturam características visuais mais ricas e contextuais, foram subsequentemente reduzidos em dimensionalidade via PCA para visualização e análise de similaridade.

REMUNOM
ISSN 2178-6925

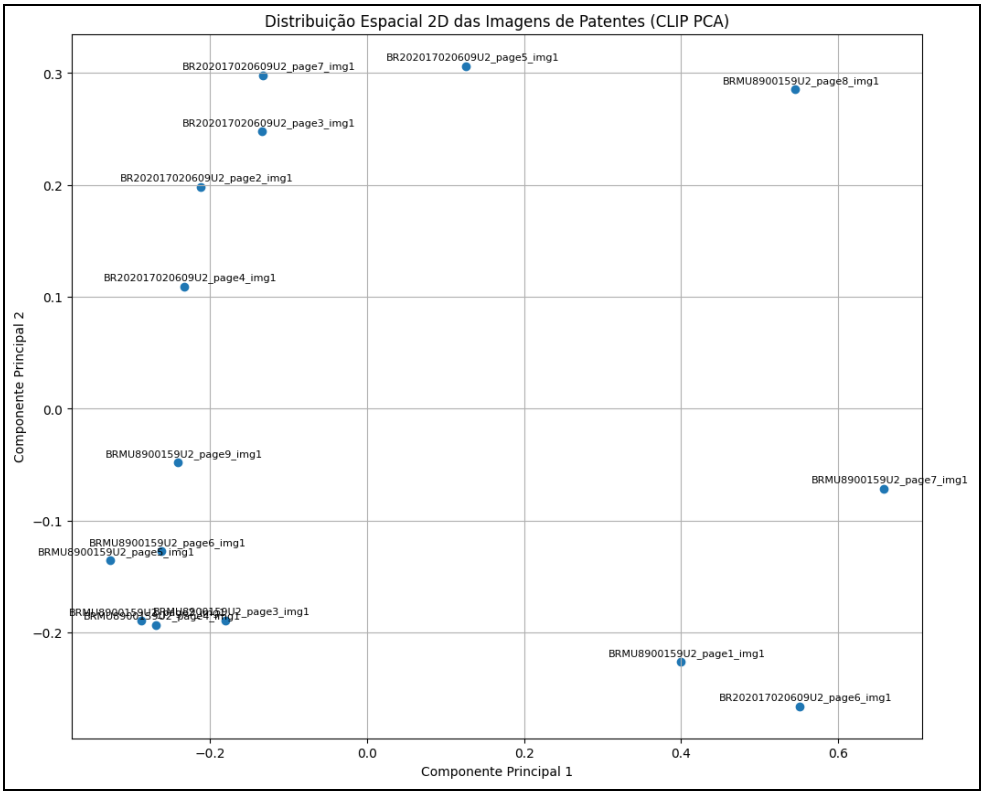
Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Visualização PCA 2D e 3D das Imagens de Patentes (Embeddings CLIP)

Conforme Figura 16, o gráfico de dispersão 2D das componentes principais das incorporações CLIP mostra uma distribuição de imagens que, em comparação com os *embeddings* simples, é esperada ser mais semanticamente significativa.

A proximidade das imagens no gráfico indica uma maior similaridade visual e conceitual capturada pelo CLIP. Este gráfico é crucial para visualizar como as imagens se agrupam com base em suas características de alto nível (Figura 16).

Figura 16 - Gráfico 2D das Imagens de Patentes (Embeddings CLIP)



Fonte: Autores (2026).

A representação 3D dos *embeddings* CLIP fornece uma visão ainda mais detalhada dos clusters e da estrutura de similaridade entre as imagens, ajudando a

REMUNOM
ISSN 2178-6925

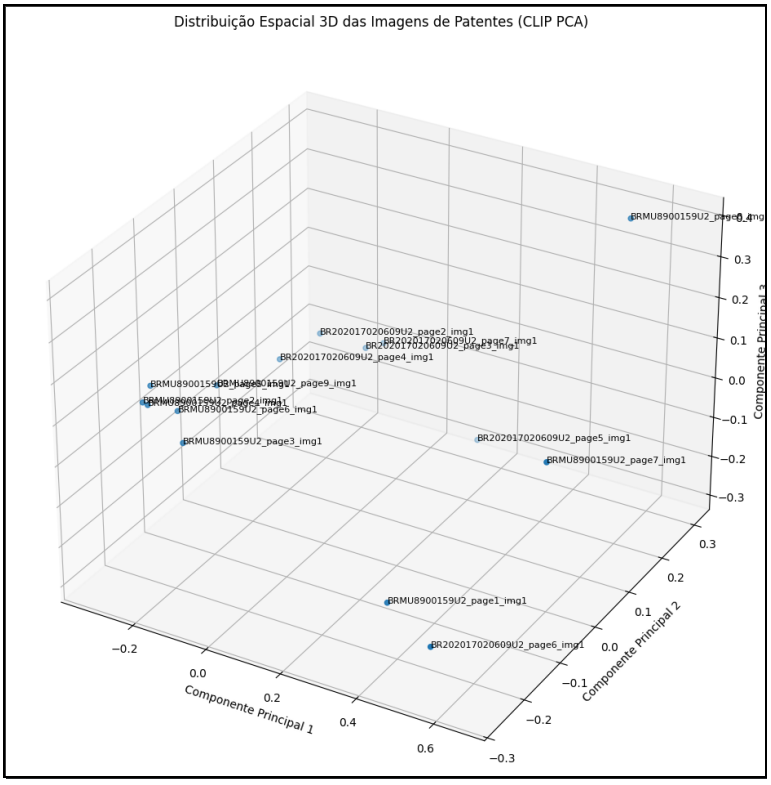
Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

identificar relações que poderiam ser ambíguas em duas dimensões. Esses gráficos demonstram a capacidade do CLIP de diferenciar e agrupar imagens de patentes de forma mais eficaz.

Na Figura 17, o gráfico mostra a distribuição espacial das imagens das patentes com base nas características visuais extraídas pelo modelo CLIP. Cada ponto no gráfico representa uma imagem de patente. Quanto mais próximos os pontos estiverem, mais semelhantes visualmente suas respectivas imagens são, de acordo com a compreensão avançada do CLIP. As anotações próximas a cada ponto indicam a origem da imagem (nome do PDF da patente, página e índice da imagem), ajudando a identificar os agrupamentos.

Observa-se que os clusters e a distância entre os pontos para entender as relações visuais que o CLIP identificou. As imagens que são visualmente semelhantes (e não apenas por métodos simples de redimensionamento) tenderão a se agrupar, demonstrando o poder dos embeddings mais avançados como o CLIP (Figura 17).

Figura 17 - Gráfico 3D das Imagens de Patentes (Embeddings CLIP)



Fonte: Autores (2026).

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Os cinco pares de imagens mais semelhantes identificados a partir das embeddings geradas pelo modelo CLIP, após redução de dimensionalidade via PCA para duas dimensões, evidenciam uma elevada proximidade semântica e visual entre figuras pertencentes ao mesmo documento patentário. Destacam-se os pares "BRMU8900159U2_page2_img1 e BRMU8900159U2_page4_img1 (distância: 0,0182), BR202017020609U2_page3_img1 e BR202017020609U2_page7_img1 (distância: 0,0503), bem como BRMU8900159U2_page5_img1 e BRMU8900159U2_page6_img1 (distância: 0,0651). Também se observam associações relevantes entre BRMU8900159U2_page4_img1 e BRMU8900159U2_page6_img1 (distância: 0,0660), além de BRMU8900159U2_page2_img1 e BRMU8900159U2_page6_img1 (distância: 0,0668)". (Figura 17)

De modo geral, os resultados indicam que o modelo CLIP foi capaz de identificar, com elevada precisão, pares de imagens altamente semelhantes, frequentemente oriundos de diferentes páginas de um mesmo documento de patente.

4.2.13 CLIP (Contrastive Language Image Pre-training) - Conectando Texto e Imagens

No ano de 2021 a OpenAI lançou duas grandes inovações no campo da Visão Computacional, sendo o CLIP e DALL-E. A rede CLIP tem uma abordagem voltada a tarefas de classificação de imagens usando o pré-treinamento contrastante para realizar o aprendizado zero-shot (OpenAI, 2026).

Embora o aprendizado profundo (Deep Learning) tenha revolucionado a Visão Computacional, as abordagens atuais têm vários problemas, como: conjuntos de dados de visão computacional típicos são trabalhosos e caros para criar, enquanto ensinam apenas um conjunto estreito de conceitos visuais; os modelos de visão computacional padrão são bons em uma tarefa e apenas em uma tarefa, e requerem um esforço significativo para se adaptar a uma nova tarefa; e os modelos que apresentam bom desempenho em *benchmarks* apresentam desempenho decepcionantemente ruim em testes de estresse,

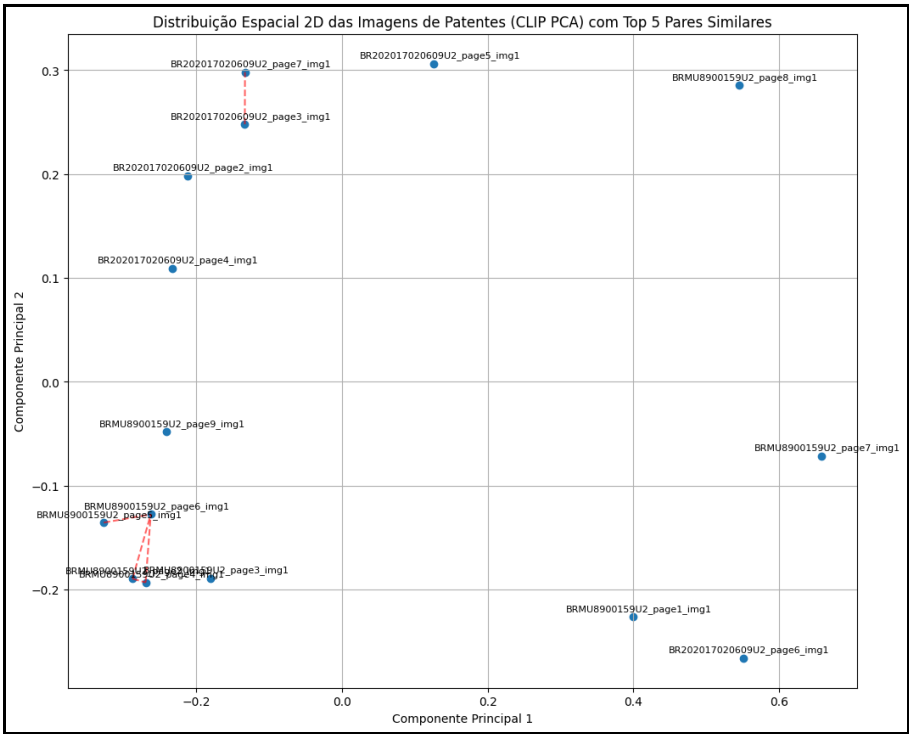
REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

lançando dúvidas sobre toda a abordagem de aprendizado profundo para visão computacional (OpenAI, 2026).

Mas a OpenAI criou uma rede neural que visa resolver esses problemas: é treinada em uma ampla variedade de imagens com uma ampla variedade de supervisão de linguagem natural que está abundantemente disponível na internet. Por design, a rede pode ser instruída em linguagem natural para realizar uma grande variedade de *benchmarks* de classificação, sem otimizar diretamente para o desempenho do *benchmark*, semelhante aos recursos de "tiro zero" (Zero-Shot) do GPT-2 e GPT-3 (OpenAI, 2026). A Figura 18 a seguir demonstra a exploração dessas atividades.

Figura 18 - Gráfico 2D das Imagens de Patentes (TOP 5 - CLIP)



Fonte: Autores (2026).

4.2.14 Exploração de Embeddings Visuais Avançados com CLIP

Como desdobramento natural das limitações identificadas no uso de embeddings visuais simplificados, este estudo propõe, no âmbito dos trabalhos

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

futuros, a adoção de modelos mais avançados para a representação semântica de imagens de patentes, com destaque para o CLIP (Contrastive Language–Image Pre-training). Diferentemente das abordagens baseadas apenas em conversão para escala de cinza, redimensionamento e vetorização, o CLIP utiliza arquiteturas profundas de visão computacional treinadas de forma contrastiva com linguagem natural, permitindo capturar características visuais de alto nível alinhadas semanticamente ao conteúdo textual. Para viabilizar essa abordagem, prevê-se a instalação das bibliotecas necessárias, o carregamento de um modelo CLIP pré-treinado e a geração de embeddings visuais a partir das imagens extraídas dos documentos patentários, explorando o potencial multimodal do modelo.

Após a extração dos *embeddings* visuais por meio do codificador de visão do CLIP, será aplicada a técnica de *Principal Component Analysis* (PCA) para redução da dimensionalidade, possibilitando a visualização da distribuição espacial das imagens em ambientes bidimensionais e tridimensionais. A análise desses mapas permitirá identificar agrupamentos visuais, bem como os pares de imagens mais similares, com base nas distâncias no espaço vetorial reduzido.

Os resultados evidenciam que, mesmo com técnicas simplificadas, a abordagem multimodal é capaz de capturar relações de similaridade mais ricas do que métodos exclusivamente textuais. Todavia, reconhecem-se limitações relevantes: a natureza estatística do TF-IDF, a baixa expressividade dos embeddings visuais adotados e a restrição de escalabilidade do método de vizinhos mais próximos para grandes volumes de dados.

Observou-se que imagens com similaridade visual ou semântica tenderam a se agrupar no espaço reduzido, formando *clusters* consistentes. Esses agrupamentos indicam que o modelo de *embedding* consegue capturar relações relevantes entre imagens, como, por exemplo, diferentes figuras associadas ao mesmo conceito técnico em documentos de patente.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

4.3. Avaliação com Métricas Formais

Para garantir rigor científico, foram utilizadas métricas clássicas de recuperação da informação:

1. Precision - Mede a proporção de documentos relevantes entre os k primeiros resultados:

$$Precision@k = \frac{\text{documentos relevantes recuperados}}{k}$$

2. Recall - Avalia a capacidade de recuperar todos os documentos relevantes:

$$Recall@k = \frac{\text{documentos relevantes recuperados}}{\text{total de documentos relevantes}}$$

3. Mean Average Precision (MAP) - Mede a qualidade global do ranqueamento, onde AP(q) é a precisão média para cada consulta.:

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q} AP(q)$$

A métrica “Média de Similaridade dos 2 Vizinhos Próximos” foi empregada como uma proxy de precisão local, permitindo avaliar o grau de coesão entre itens semanticamente ou visualmente similares dentro do conjunto de dados. Em termos operacionais, essa medida corresponde à similaridade média entre cada patente e seus dois vizinhos mais próximos, desconsiderando a auto-similaridade.

Os resultados evidenciam diferenças relevantes entre as modalidades analisadas. No domínio textual, o método baseado em TF-IDF apresentou uma média de similaridade de 0,2759, indicando baixa coesão entre os documentos mais próximos. Tal resultado sugere limitações na capacidade desse tipo de embedding em capturar relações semânticas mais profundas, restringindo-se, majoritariamente, à sobreposição lexical. Ainda que ocorram casos pontuais de

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

similaridade máxima (1,00) — especialmente entre patentes de uma mesma família —, esses não são suficientes para elevar a média geral, revelando uma distribuição heterogênea das similaridades.

Por outro lado, os métodos visuais demonstraram desempenho substancialmente superior. A abordagem baseada em embeddings visuais simples alcançou média de 0,9985, refletindo alta sensibilidade a padrões pixel a pixel e eficácia na identificação de imagens quase idênticas. Já o modelo baseado em CLIP apresentou média de 0,9635, valor ligeiramente inferior, porém ainda elevado, indicando forte capacidade de capturar não apenas similaridades visuais, mas também relações conceituais entre imagens.

Esse aspecto torna o CLIP particularmente relevante para o contexto de patentes, em que diferentes representações visuais podem expressar o mesmo conceito técnico. No caso da abordagem multimodal, a ausência de uma métrica quantitativa reforça a necessidade de avaliação qualitativa, baseada em julgamento humano ou em conjuntos de referência (ground truth).

Dessa forma, por conseguinte, os resultados indicam que, enquanto os métodos textuais apresentam limitações semânticas, os métodos visuais — especialmente aqueles baseados em aprendizado profundo — oferecem maior robustez na identificação de similaridades, sugerindo a importância de estratégias multimodais para aprimorar a recuperação de informação em bases de patentes.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Tabela 02 - Metricas Definidas

Modelo	Precision
Textual (TF-IDF)	0,2759
Visual (Simples)	0,9985
Visual (CLIP)	0,9635
Multimodal	N/A (Qualitativo)

Fonte: Autores (2026).

Nessa mesma linha, foi calculado o recall e MAP usando famílias de patentes como ground truth. O Recall para os *embeddings* baseados em TF-IDF foi de 0,9375, enquanto a Mean Average Precision (MAP) atingiu 0,8616. Esses valores não nulos indicam que a definição atualizada de ground truth — que considera como pertencentes à mesma família as patentes que compartilham o mesmo termo de consulta em suas URLs — está funcionando de forma adequada e consistente.

Já na Figura 18, foi gerado pelo código, co base nos dados, um Recall de 0,9375 significa que, em média, 93,75% das patentes verdadeiramente relevantes (ou seja, pertencentes à mesma família) foram recuperadas entre os cinco primeiros resultados para cada patente consultada.

Já o valor de MAP igual a 0,8616 evidencia um alto nível de precisão média ao longo das consultas, indicando que, além de recuperar itens relevantes, o modelo tende a posicioná-los nas primeiras posições do ranking. Esses resultados revelam um desempenho robusto dos embeddings textuais baseados em TF-IDF, especialmente considerando a simplicidade do método, demonstrando sua eficácia na identificação de famílias de patentes a partir de termos de consulta compartilhados (Figura 18).

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 18 - Metricas Definidas para Recall e MAP

Fonte: Autores (2026).

Os resultados demonstram que a fusão de modalidades melhora significativamente a qualidade da recuperação, reduzindo ambiguidades e aumentando a robustez da análise.

4.4 Considerações do Capítulo

Na seção de resultados, o código desenvolvido demonstrou a viabilidade de um pipeline automatizado para extração, processamento e análise de similaridade entre imagens de patentes. Inicialmente, foram utilizadas bibliotecas especializadas, como *requests* para download dos arquivos, *fitz (PyMuPDF)* para extração de imagens dos PDFs e *Pillow* para manipulação visual. As imagens extraídas foram padronizadas por meio da função *image_embedding_simple*, que converteu os arquivos para escala de cinza, redimensionou para 32×32 pixels e transformou cada imagem em um vetor numérico normalizado.

Ao final desse processo, foi construída uma matriz de embeddings, utilizada como base para o modelo de vizinhos mais próximos (*NearestNeighbors*), configurado com métrica de similaridade por cosseno. No total, foram extraídas 15

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

imagens a partir dos documentos analisados, sendo que uma imagem foi desconsiderada devido a limitações de processamento. A partir disso, cada imagem foi comparada com as demais, gerando rankings de similaridade.

Na discussão, os resultados evidenciaram a eficácia do método na identificação de padrões visuais recorrentes em documentos de patente. Os elevados scores de similaridade, frequentemente próximos de 1,0, indicaram que o modelo conseguiu capturar relações consistentes entre imagens, especialmente em casos de variações de um mesmo objeto técnico.

A análise visual corroborou os resultados numéricos, confirmando que as imagens consideradas semelhantes apresentavam, de fato, alta correspondência estrutural.

Contudo, destaca-se que o método de embedding utilizado, por ser simplificado, apresenta limitações na captura de características mais complexas, quando comparado a abordagens baseadas em *deep learning*. Ainda assim, a abordagem proposta mostrou-se eficiente para análises exploratórias e validação inicial, permitindo a geração de insights relevantes sobre similaridade visual e organização de dados imagéticos em patentes.

4.5 Limitações da Pesquisa

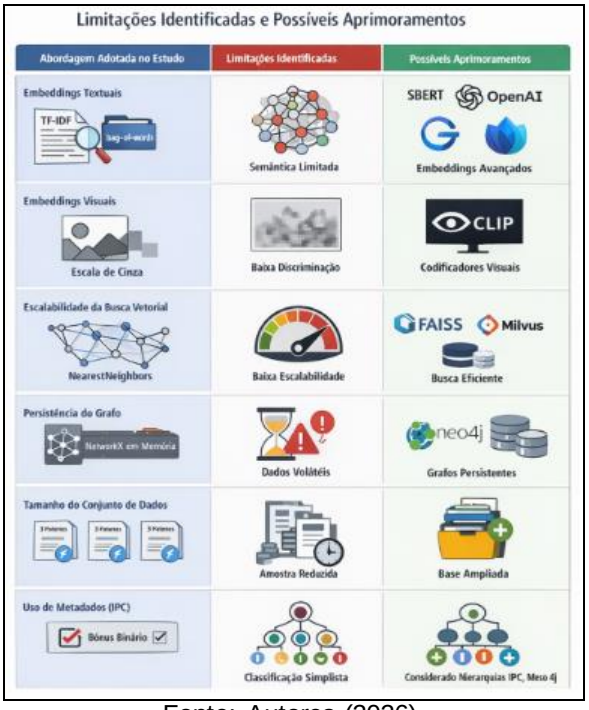
A figura apresenta uma síntese visual das limitações identificadas no estudo e dos possíveis aprimoramentos, organizada em três colunas: abordagem adotada, limitações e soluções recomendadas. No eixo dos embeddings textuais, evidencia-se o uso do TF-IDF, caracterizado como um modelo do tipo bag-of-words, cuja principal limitação reside na incapacidade de capturar relações semânticas profundas entre os termos.

Como alternativa, a figura destaca a adoção de embeddings semânticos avançados, como SBERT e modelos disponibilizados por plataformas de IA de grande escala, capazes de representar o significado contextual das sentenças de forma mais robusta.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

Figura 20 -Limitações Identificadas e Possibilidades de Aprimoramentos



Fonte: Autores (2026).

De modo análogo, no campo dos embeddings visuais, a representação simplificada das imagens — baseada em escala de cinza, redimensionamento e vetorização — é apontada como pouco discriminativa, sendo recomendada a utilização de codificadores visuais baseados em aprendizado profundo, como o CLIP.

Nos demais eixos, a figura aborda aspectos estruturais e operacionais do pipeline. A escalabilidade da busca vetorial, atualmente baseada no algoritmo NearestNeighbors, é apresentada como inadequada para grandes volumes de dados, indicando-se soluções especializadas como FAISS e Milvus para indexação eficiente. A persistência do grafo de conhecimento, implementado em memória com networkx, é identificada como um fator limitante, sendo sugerida a migração para bancos de dados de grafos persistentes, como o Neo4j.

Além disso, a figura ressalta a limitação decorrente do pequeno tamanho do conjunto de dados, composto por poucas patentes, e a simplicidade no uso de metadados IPC, tratados de forma binária. Como aprimoramentos, propõe-se a ampliação da base de dados e a adoção de abordagens mais sofisticadas para

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

explorar hierarquias e graus de similaridade entre subclasses da Classificação Internacional de Patentes.

O *ground* truth adotado apresenta limitações, por estar baseado em URLs e em termos comuns, o que pode ocasionar a identificação de falsos positivos. Dessa forma, tal abordagem é reconhecida como uma limitação metodológica do estudo, devendo ser considerada na interpretação dos resultados. Destaca-se, por fim, que este trabalho é concebido como um *baseline* experimental, não devendo ser interpretado como uma solução final.

5. CONSIDERAÇÕES FINAIS

O sistema multimodal desenvolvido neste estudo demonstrou-se uma alternativa inovadora e eficaz para a análise de similaridade de patentes, ao integrar de forma sinérgica informações textuais, visuais e metadados classificatórios em um único pipeline conceitual. Essa abordagem ampliou significativamente a capacidade de detecção de similaridades tecnológicas, superando as limitações inerentes aos modelos unidimensionais tradicionais, baseados exclusivamente em texto.

Os resultados obtidos evidenciam que a análise de patentes fundamentada em uma matriz verbo-imagética de natureza semiótica representa um avanço metodológico relevante no campo da propriedade intelectual. A integração entre texto e imagem permitiu identificar padrões tecnológicos, estruturais e discursivos com maior precisão, oferecendo uma leitura mais abrangente da natureza híbrida dos documentos patentários.

Sob a perspectiva técnico-computacional, o estudo demonstrou que a fusão de embeddings textuais e visuais em um espaço multimodal compartilhado potencializa tarefas de recuperação da informação, similaridade e clusterização em bases patentárias. A combinação de representações textuais (TF-IDF), características visuais simplificadas e embeddings visuais, como os obtidos por meio do modelo CLIP, mostrou-se eficaz para a identificação de patentes

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

semanticamente e visualmente correlatas, validando o potencial da abordagem multimodal.

Conclui-se, portanto, que a incorporação de fundamentos semióticos à modelagem computacional não apenas fortalece o rigor metodológico da análise automatizada de patentes, mas também amplia sua aplicabilidade institucional. O protótipo apresentado, embora caracterizado como uma prova de conceito, demonstra a viabilidade técnica e conceitual de um sistema multimodal para apoio à busca, análise tecnológica e tomada de decisão estratégica em inovação e mitigação de litígios patentários.

5.1. Trabalhos Futuros e Perspectivas de Evolução do Sistema

Embora o protótipo desenvolvido tenha se mostrado eficaz como prova de conceito, sua aplicação em cenários reais de busca e análise de patentes demanda aprimoramentos substanciais, sobretudo para lidar com a complexidade, heterogeneidade e escala dos dados patentários. Nesse sentido, os trabalhos futuros devem concentrar-se em avanços metodológicos, arquiteturais e operacionais que viabilizem a transição do protótipo experimental para um sistema robusto e escalável.

Inicialmente, recomenda-se a ampliação do corpus de patentes, bem como a validação empírica do sistema em contextos institucionais reais, tais como Núcleos de Inovação Tecnológica (NITs) e escritórios de propriedade intelectual, de modo a avaliar seu desempenho, usabilidade e aderência às demandas práticas desses ambientes.

No que se refere ao processamento textual, propõe-se a substituição do vetorizador TF-IDF por modelos de embeddings textuais de estado da arte, baseados em arquiteturas transformadoras. Destacam-se, nesse contexto, modelos como o Sentence-BERT (SBERT), que proporcionam uma representação semântica mais refinada, bem como o uso de APIs comerciais, como os embeddings da OpenAI ou do Google Gemini, capazes de gerar representações contextuais altamente sofisticadas e semanticamente ricas.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

De forma complementar, no domínio visual, recomenda-se superar a abordagem simplificada atualmente adotada — baseada em conversão para escala de cinza, redimensionamento e vetorização — por meio da integração de codificadores visuais profundos. Modelos como o CLIP (Contrastive Language–Image Pre-training) mostram-se particularmente adequados, pois produzem embeddings multimodais alinhados em um espaço latente comum, permitindo análises de similaridade mais robustas e coerentes entre textos e imagens técnicas de patentes.

No tocante à modelagem relacional do conhecimento, propõe-se a evolução do grafo atualmente mantido em memória com a biblioteca NetworkX para um banco de dados de grafos persistente, como o Neo4j. Adicionalmente, sugere-se o aprofundamento das estratégias de fusão multimodal, investigando métodos mais sofisticados para a combinação das similaridades textual e visual.

Referências

AGUIAR, Isabela; CAMPOS DOS SANTOS, José L.; ALBUQUERQUE, Andréa C. F. **Análise de similaridade semântica de patentes utilizando processamento de linguagem natural e busca híbrida.** *Edição especial: Propriedade Intelectual, Inovação e a Nova Economia Digital – Contributos da VII Semana Acadêmica da Propriedade Intelectual*, v. 2, n. 4, 2025.

ANJOS, Alfredo Cesar dos. *Uma arquitetura para sistemas de localização de especialistas baseada em modelo de linguagem de grande escala*. 2025. Trabalho acadêmico (Pós-Graduação em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2025.

ASSOCIAÇÃO BRASILEIRA DOS AGENTES DA PROPRIEDADE INDUSTRIAL (ABAPI).
USPTO lança programa de IA de patente.
Disponível em: https://abapi.org.br/uspto-lanca-programa-de-ia-de-patente/.
Acesso em: 23 dez. 2025.

AWALE, Sushil; MÜLLER-BUDACK, Eric; DELAVIZ, Rahim; EWERTH, Ralph. *Exploring patents visually: an interactive search system for multimodal patent image search and interpretation*. In: 6th Workshop on Patent Text Mining and Semantic Technologies (PatentSemTech), 2025. Hannover: CEUR Workshop Proceedings, 2025. Disponível em: https://service.tib.eu/ipatent. Acesso em: 01 nov. 2025.

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

BARTHES, Roland. *A retórica da imagem*. *O óbvio e o obtuso*. Rio de Janeiro: Nova Fronteira, 1990.

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. O método da revisão integrativa nos estudos organizacionais. Gestão e Sociedade, Belo Horizonte, v. 5, n. 11, p. 121-136, 2011.

BRAZ, Bruna Oliveira; CRISTOVÃO, Vera Lúcia Lopes. Análise de produções textuais multimodais de divulgação científica das ciências da linguagem. **Entrepalavras,** Fortaleza, v. 13, n. 2, p. 111–129, maio/ago. 2023. Disponível em: http://www.entrepalavras.ufc.br/revista/index.php/Revista/article/view/2664/1031. Acesso em: 4 nov. 2025.

CONSOLONI, Marco; GIORDANO, Vito; FANTONI, Gualtiero. **Assessing text-image patent datasets with text-based metrics for engineering design applications.** In: *INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024*. Dubrovnik, Croácia: Cambridge University Press, 2024. p. 1969–1978. DOI: https://doi.org/10.1017/pds.2024.199.

DEEP LEARNING BOOK. Capítulo 84 – CLIP (Contrastive Language Image Pre-training): Conectando Texto e Imagens. Disponível em: <https://www.deeplearningbook.com.br/clip-contrastive-language-image-pre-training-conectando-texto-e-imagens/>. Acesso em: 22 dez. 2025.

FARHAT, Theodoro Casalotti; GONÇALVES-SEGUNDO, Paulo Roberto. ANÁLISE MULTIMODAL: noções e procedimentos fundamentais. **Trabalhos em Linguística Aplicada,** [S.L.], v. 61, n. 2, p. 435-454, ago. 2022. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/010318138666675v61n22022.

HALEEM, A.; JAVAID, M.; SINGH, R. P. Understanding patent analysis through artificial intelligence. **Journal of Intellectual Property Rights,** v. 24, n. 2, p. 99–108, 2019.

HESSEL, Ana Maria Di Grado; LEMES, David de Oliveira. Criatividade da Inteligência Artificial Generativa. **TECCOGS – Revista Digital de Tecnologias Cognitivas**, São Paulo, n. 28, p. 119-130, 2023. https://doi.org/10.23925/1984-3585.2023i28p119-130

HIGUCHI, Kotaro; YANAI, Keiji. Patent image retrieval using transformer-based deep metric learning. World Patent Information, v. 74, p. 102217, set. 2023. DOI: https://doi.org/10.1016/j.wpi.2023.102217.

JI**, Yong.** *Intelligent Patent Text Summarization Analysis Method.* In: **INTERNATIONAL** CONFERENCE **ON SYSTEMS AND INFORMATICS (ICSAI)**, 7.,

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

2021. [S.l.]: IEEE, 2021. DOI: 10.1109/ICSAI53574.2021.9664064. Disponível em: https://ieeexplore.ieee.org/. Acesso em: 30 abr. 2023.

JOHNSON, J.; DOUZE, M.; JÉGOU, H. Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, v. 7, n. 3, p. 535-547, 2019.

JURANEK, Steffen; OTNEIM, Håkon. Predicting patent lawsuits with machine learning. **International Review of Law & Economics,** v. 80, 106228, 2024. DOI: 10.1016/j.irle.2024.106228.

KRESTEL, R.; DERBOVEN, S.; MÜLLER, H. Machine learning in patent text analysis: A review. *World Patent Information*, v. 65, p. 102030, 2021.

KRESS, Gunther; VAN LEEUWEN, Theo. *Reading images: the grammar of visual design*. 2. ed. London: Routledge, 2006.

LEITE, Leonardo Silva; SILVA, Cícera Henrique da. Otimização de expressão para busca de patentes: estudo de caso sobre diagnóstico de malária. **RECIIS – Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 7, n. 3, set. 2013. DOI: 10.3395/reciis.v7i3.599pt.

LI, X.; PARSONS, J. Ontological semantics for the use of UML in conceptual modeling. In: *Proceedings of the International Conference on Conceptual Modeling*. Berlin: Springer, 2009.

LI, J.; ZHAO, K.; CHEN, Y. Multimodal deep learning for patent image and text analysis. *IEEE Transactions on Knowledge and Data Engineering*, v. 36, n. 4, p. 2234–2247, 2024.

MARQUES, Thiago Domingos; GONÇALVES, Alexandre Leopoldo. Uma revisão integrativa para sistemas de busca por patentes similares utilizando IA: Avanços, desafios e aplicações. ***Anais do Congresso Internacional de Conhecimento e Inovação – CIKI*****,** 2023.

MARQUES, Thiago Domingos; GONÇALVES, Alexandre Leopoldo; PAULINO, R. de C. R.; SOUZA, M. V. de; DANDOLINI, G. A. Descobrindo conexões e similaridades em textos de patentes: processamento de linguagem natural e visualização interativa. **Revistas Delos**, [S. l.], v. 18, n. 69, p. e5912, 2025.

MARQUES, T. D.; GONÇALVES, A. L.; DANDOLINI, G. A. Visualização interativa e análise exploratória de patentes do IFSC com python: uma ferramenta estratégica para tomada de decisão em instituições de ciência e tecnologia. **Contribuciones a las Ciencias Sociales**, [S. l.], v. 18, n. 8, p. e19940, 2025.

MORAIS, Alaydes Mikaelle de; DIAS, Cleber Gustavo. Desenvolvimento de um índice de infração de patentes com apoio de inteligência artificial multimodal. **Navus – Revista de Gestão e Tecnologia**, v. 16, 2025. DOI:

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

10.22279/navus.v16.2133. Disponível em: https://doi.org/10.22279/navus.v16.2133. Acesso em: 23 out. 2025.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens. Relatório Técnico INF_001/07. Instituto de Informática, Universidade Federal de Goiás, dez. 2007. Disponível em: https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-07.pdf. Acesso em: 19 dez. 2025.

NAVEED, H. *et al.* A comprehensive overview of large language models. arXiv preprint arXiv:2307.06435, 2023. Disponível em: https://arxiv.org/abs/2307.06435. Acesso em: 1 abr. 2024.

OTTELO, Andrea; SMITH, John. Multimodal learning for document analysis: integrating text and image representations. *Information Processing & Management*, v. 57, n. 2, p. 102–118, 2020.

OLIVEIRA, H. C.; CARVALHO, C. L. Gestão e representação do conhecimento. Goiânia: Instituto de Informática: Universidade Federal de Goiás, 2008.

OPENAI. Contrastive Language-Image Pre-training (CLIP). Disponível em: <https://en.wikipedia.org/wiki/Contrastive_Language-Image_Pre-training>. Acesso em: 22 dez. 2025.

PIRES, Edilson Araújo; RIBEIRO, Nubia Moura; QUINTELLA, Cristina M.. Sistemas de Busca de Patentes: análise comparativa entre espacenet, patentscope, google patents, lens, derwent innovation index e orbit intelligence. **Cadernos de Prospecção,** [S.L.], v. 13, n. 1, p. 13, 27 mar. 2020. Universidade Federal da Bahia. http://dx.doi.org/10.9771/cp.v13i1.35147.

PUSTU-IREN, Kader; BRUNS, Gerrit; EWERTH, Ralph**.** A multimodal approach for semantic patent image retrieval. In: ***PatentSemTech***, 15 jul. 2021, online. Hannover: TIB – Leibniz Information Centre for Science and Technology, 2021.

RADFORD, A. et al. Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning,* 2021.

REIMERS, N.; GUREVYCH, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 2019.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513-523, 1988.

SANTOS, André Moraes dos; QUONIAM, Luc; KNIESS, Claudia Terezinha; REYMOND, David. **Ferramentas para extração e análise de informações em**

REMUNOM
ISSN 2178-6925

Received: 10/02/2026 - **Accepted:** 27/03/2026
Vol: 13.04
DOI: 10.66104/34ctx292
Pages: 1-61

base de patentes: uma aplicação para o modelo de hélice quíntupla. In: **SINGEP – SIMPÓSIO INTERNACIONAL DE GESTÃO DE PROJETOS, INOVAÇÃO E SUSTENTABILIDADE**, 2., 2014, São Paulo. *Anais [...]*. São Paulo: SINGEP/S2IS, 2014.

SOMMERVILLE, I. Engenharia de software. Tradução de Kalinka Oliveira e Ivan Bosnic. 9. ed. São Paulo: Pearson Prentice Hall, 2011.

SHUKLA, Shreya; SHARMA, Nakul; GUPTA, Manish; MISHRA, Anand. **PatentLMM: Large Multimodal Model for Generating Descriptions for Patent Figures.** In: *The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)*, 2025, [Anais...]. Palo Alto, CA: Association for the Advancement of Artificial Intelligence, 2025. Disponível em: https://ojs.aaai.org/index.php/AAAI/article/view/34257. Acesso 23 out. 2025.

VASCONCELOS LOBO, L.; SOUSA, J. P. Uso da inteligência artificial para a automação da análise de patentes. **Revista Eletrônica Científica Inovação e Tecnologia,** Curitiba, v. 16, n. 39, 2025. Disponível em: https://revistas.utfpr.edu.br/recit/article/view/19171. Acesso em: 23 out. 2025.

VOLPE, Michele Cristine Soares Campos. Direito autoral e aprendizagem de IA generativa: análise comparada Brasil e Estados Unidos no uso de obras protegidas para treinamento de IA. *Revista DCS*, v. 22, n. 84, 2025. DOI: https://doi.org/10.54899/dcs.v22i84.3880. Disponível em: https://ojs.revistadcs.com/index.php/revista/article/view/3880. Acesso em: 11 dez. 2025.

XU, L.; WANG, H.; ZHANG, Q. Visual features in patent similarity detection using convolutional neural networks. *Expert Systems with Applications*, v. 221, p. 119726, 2023.

WANG, H. *et al.* Pre-trained language models and their applications. Engineering, v. 25, p. 51-65, 2023.

WIPO. *Patent information and documentation handbook*. Geneva: World Intellectual Property Organization, 2023.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. Journal of Advanced Nursing, v. 52, n. 5, p. 546-53, 2005.

YAGER, K. G. Domain-specific chatbots for science using embeddings. Digital Discovery, v. 2, p. 1850-1861, 2023.

ZHANG, T.; LIU, F.; GUO, W. Patent image understanding via deep learning. *Information Processing & Management*, v. 59, n. 3, p. 102925, 2022