Relatório final de atividades

Projeto: "ServiçosPI: inovação e eficiência na recuperação de informações de propriedade intelectual do INPI."

Coordenador do Projeto:	**Washington Luís Ribeiro de Carvalho Segundo**		
Nome do pesquisador:	**Renan Carneiro Batista**		
Vigência da Bolsa de Pesquisa:	**02/01/2026**	**Até:**	**30/06/2026**
Encerramento em:	**30/06/2026**		
Ref. Fundação:	**31849**	**Ref. Ibict:**	**0677/2025**

Descrição das Atividades da Bolsa de Pesquisa:

1-Prova de conceito (PoC) e protótipos dos sistemas de busca (interna e externa) com desempenho avaliado em testes de escalabilidade, segurança e integração de IA (M4),
2- Relatório das atividades desenvolvidas.

1 INTRODUÇÃO

O projeto ServiçosPI investiga formas de tornar mais eficiente a recuperação de informações mantidas pelo Instituto Nacional da Propriedade Industrial (INPI). Seu escopo abrange as naturezas de proteção disponibilizadas pelo instituto, entre elas patentes, marcas, desenhos industriais, programas de computador e contratos de tecnologia. Em cada natureza há vocabulário próprio, estruturas de classificação distintas e volumes de registro elevados, o que impõe desafios específicos às camadas de recuperação e de apresentação dos resultados.

A presente pesquisa concentrou-se na prova de conceito e na prototipação da interface de busca dos sistemas de recuperação, avaliando seu comportamento frente aos dados reais das diferentes naturezas de proteção. Durante a vigência da bolsa foram identificados problemas de desempenho e de consistência na camada de apresentação do protótipo, o que orientou o esforço de investigação para o refinamento iterativo dessa camada. Em paralelo, a frente de integração de inteligência artificial foi conduzida em caráter exploratório, servindo de base teórica para etapas futuras do projeto.

2 MÉTODO

A investigação seguiu abordagem de prototipação iterativa, adequada à natureza de prova de conceito prevista na atividade da bolsa. Essa escolha alinha-se ao paradigma de *design science*, no qual o conhecimento sobre um problema de projeto e sua solução se constrói na elaboração e na avaliação de um artefato (HEVNER *et al.*, 2004). Peffers *et al.* (2007) organizam esse processo em atividades encadeadas, da identificação do problema à comunicação dos resultados, com entrada possível a partir do próprio problema observado. O presente trabalho adota dessas referências a lógica iterativa de construção e avaliação e o caráter de prova de conceito que orientam a atividade da bolsa. O objeto de estudo foi o protótipo da interface de recuperação do projeto, versionado em repositório no *GitHub*. Cada ciclo de investigação partiu da observação do comportamento da interface frente aos dados reais das naturezas de proteção, seguido de uma hipótese de ajuste, de sua incorporação ao protótipo e da verificação do

resultado. Os ciclos foram registrados por meio de controle de versão, o que permitiu documentar a evolução do artefato ao longo do período.

O trabalho organizou-se em cinco eixos de investigação. O primeiro tratou da padronização terminológica e da internacionalização da interface, com uniformização dos rótulos de filtros, de botões de ação e de menus de ordenação nas línguas de exibição. O segundo eixo abordou o comportamento dos filtros facetados, incluindo a expansão e a ocultação de filtros, os filtros por classificação e a padronização do reinício dos campos entre as diferentes buscas. O terceiro eixo tratou da visualização dos resultados, comparando exibição em tabela, em grade e em lista, e ajustando a apresentação conforme a natureza. O quarto eixo examinou a busca estruturada e a integração com a interface de programação de aplicações, com correções de endereço da API, de nomenclatura de *endpoints* e de tipagem da paginação. O quinto eixo cuidou de aspectos de manutenção e segurança, entre eles a atualização de dependência motivada por correção de segurança.

Para a frente de inteligência artificial adotou-se revisão exploratória de literatura, orientada a identificar abordagens de recuperação assistida por modelos de linguagem aplicáveis a acervos de propriedade intelectual.

3 RESULTADOS

A atividade da bolsa previa dois componentes, a prova de conceito e a prototipação dos sistemas de busca com avaliação de desempenho, e a integração de inteligência artificial. O primeiro componente foi cumprido. O segundo foi conduzido em caráter exploratório. A concentração de problemas de desempenho e de consistência na camada de apresentação do protótipo exigiu a realocação do esforço de pesquisa para essa frente, de modo que a integração de inteligência artificial permaneceu em nível exploratório, conforme detalhado na subseção 3.3.

3.1 Levantamento dos ciclos de prototipação

O período compreendeu trinta e cinco ciclos de refinamento do protótipo, registrados entre janeiro e maio de 2026, com cerca de 3.700 linhas incorporadas e

1.900 linhas revisadas ou removidas. O Quadro 1 consolida os ciclos por eixo de investigação.

Quadro 1 – Ciclos de prototipação por eixo de investigação

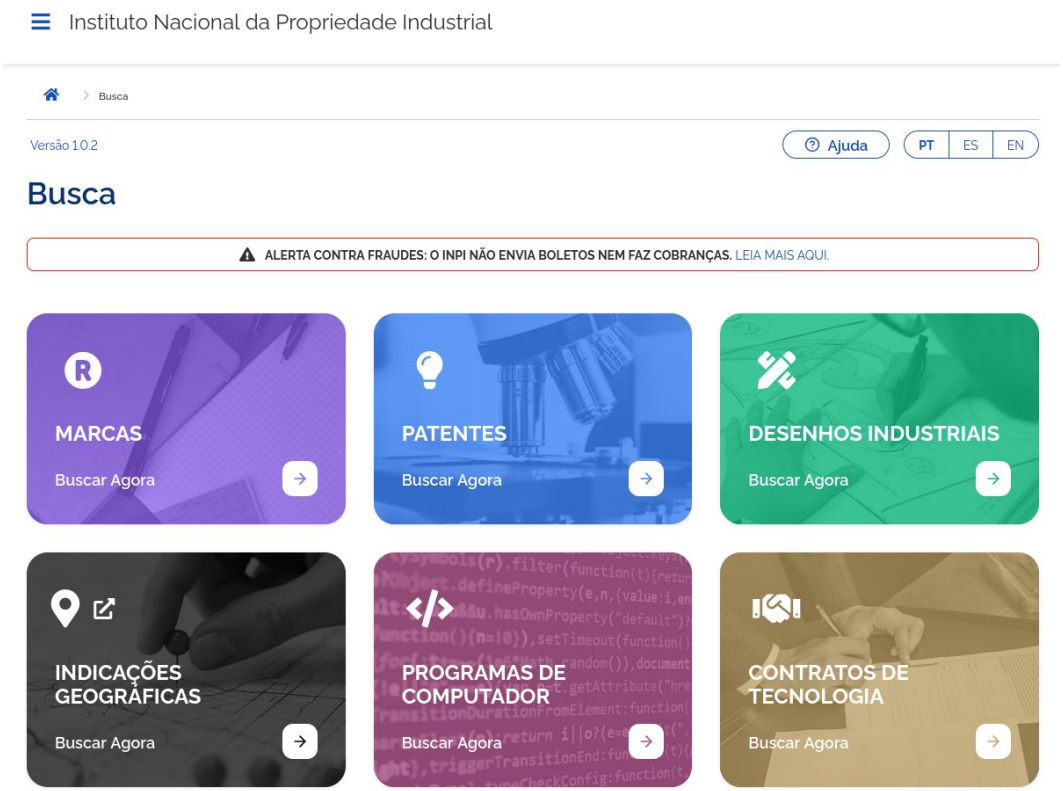
Eixo de investigação	Ciclos	Período
Padronização terminológica e internacionalização	9	jan. a mai.
Filtros facetados e navegação	7	jan. a abr.
Visualização de resultados	8	fev. a mai.
Busca estruturada e integração com a API	5	mar. a abr.
Manutenção e segurança	6	jan. a mar.

Fonte: elaborado pelo autor (2026).

3.2 Desenvolvimento e avaliação do protótipo da interface

O protótipo avaliado organiza o acesso às naturezas de proteção em uma página inicial única, a partir da qual o usuário seleciona o domínio de busca desejado, conforme a Figura 1.

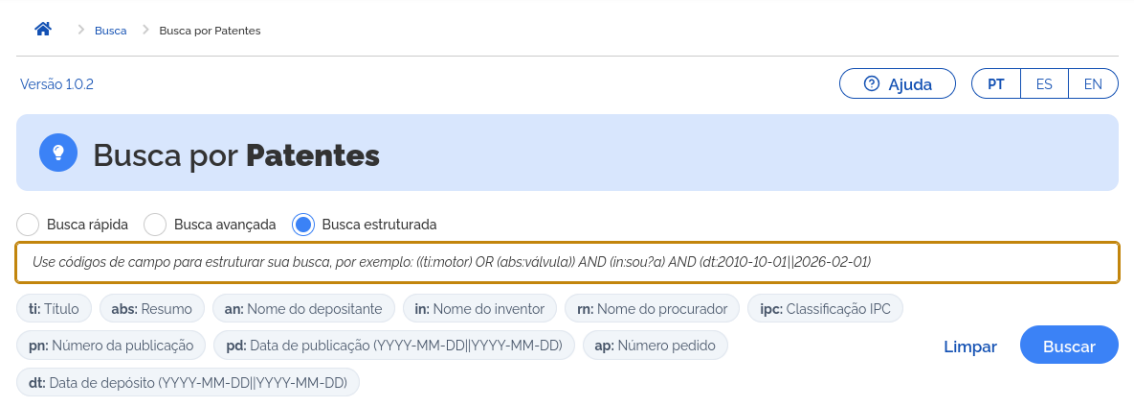
Figura 1 – Página inicial de busca do protótipo, com as naturezas de proteção



Fonte: elaborado pelo autor (2026).

Em cada domínio, a interface disponibiliza modos de busca distintos. A busca estruturada permite a composição de consultas por códigos de campo e operadores lógicos, apoiando a avaliação da expressividade da recuperação, como ilustra a Figura 2.

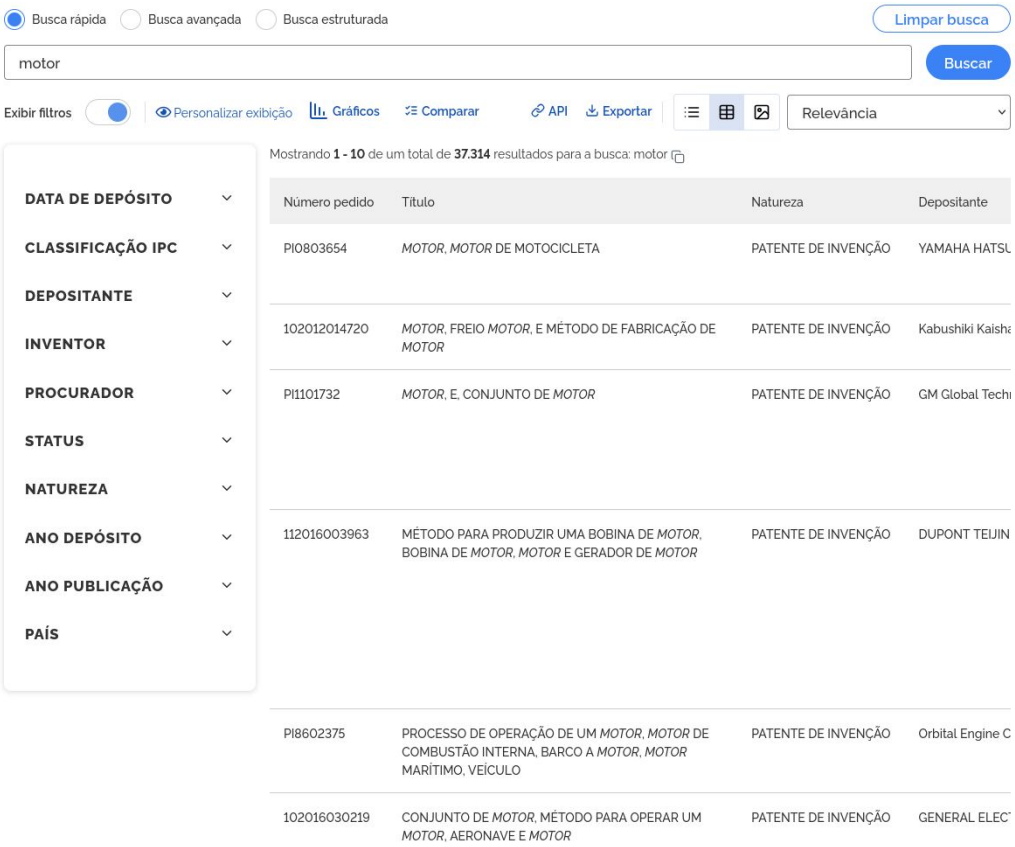
Figura 2 – Busca estruturada de patentes por códigos de campo



Fonte: elaborado pelo autor (2026).

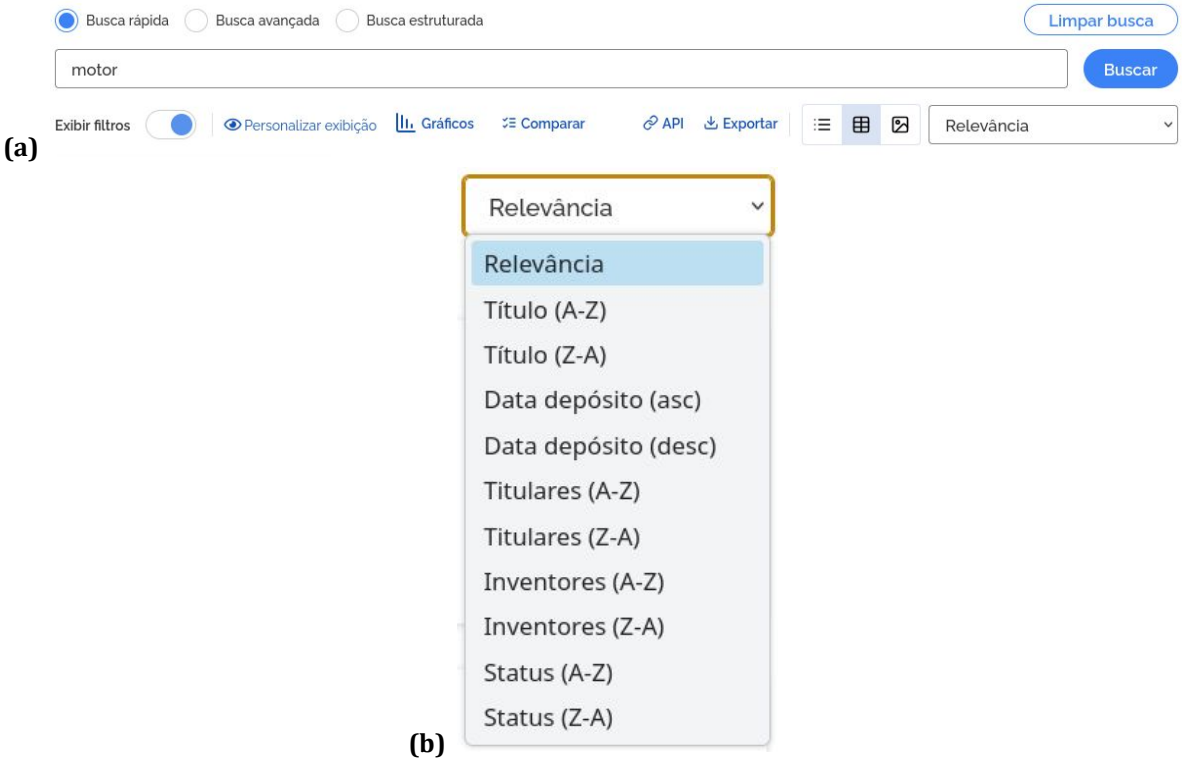
A apresentação dos resultados foi avaliada em visualização de tabela, acompanhada de filtros facetados à esquerda e de uma barra de ferramentas superior que reúne as ações sobre o conjunto recuperado. A Figura 3 apresenta essa configuração e a Figura 4 detalha os controles da barra de ferramentas e o menu de ordenação.

Figura 3 – Visualização em tabela com filtros facetados e barra de ferramentas



Fonte: elaborado pelo autor (2026).

Figura 4 – Detalhe da barra de ferramentas: (a) controles de exibição e ações; (b) menu de ordenação dos resultados



Fonte: elaborado pelo autor (2026).

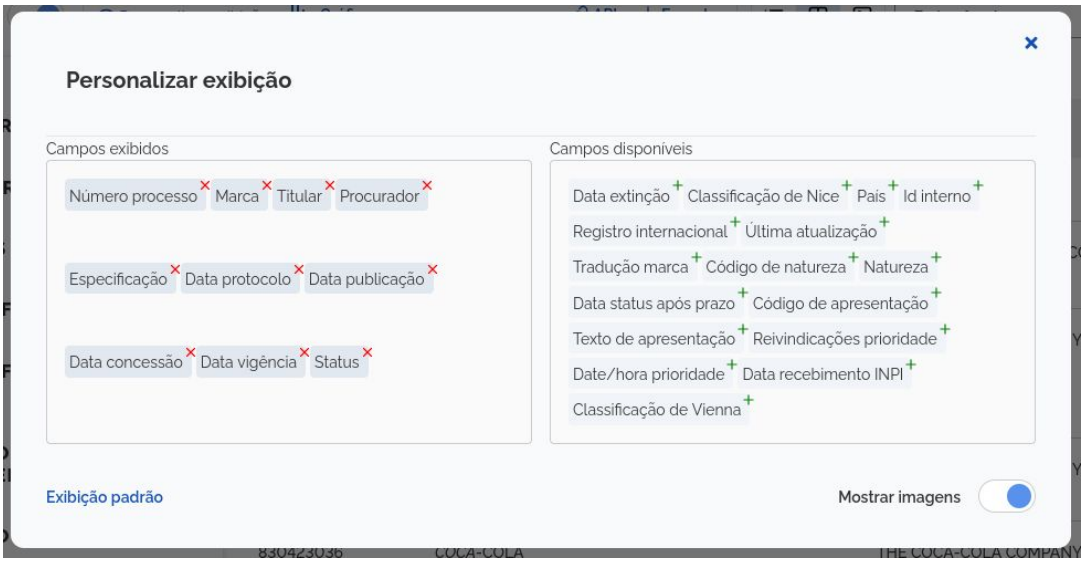
Entre os filtros facetados, avaliou-se a filtragem por classificação, que apresenta os códigos com a respectiva contagem de ocorrências e apoia o refinamento progressivo da recuperação, conforme a Figura 5. A busca facetada apoia a busca exploratória de forma mais efetiva do que a busca convencional por melhor correspondência, e a exibição de facetas com a contagem de ocorrências sustenta o refinamento progressivo da recuperação (TUNKELANG, 2009). A personalização dos campos exibidos nos resultados, ilustrada na Figura 6, foi examinada como recurso de adequação da apresentação às necessidades de cada natureza.

Figura 5 – Filtros por classificação IPC com contagem de ocorrências



Fonte: elaborado pelo autor (2026).

Figura 6 – Modal de personalização dos campos exibidos nos resultados

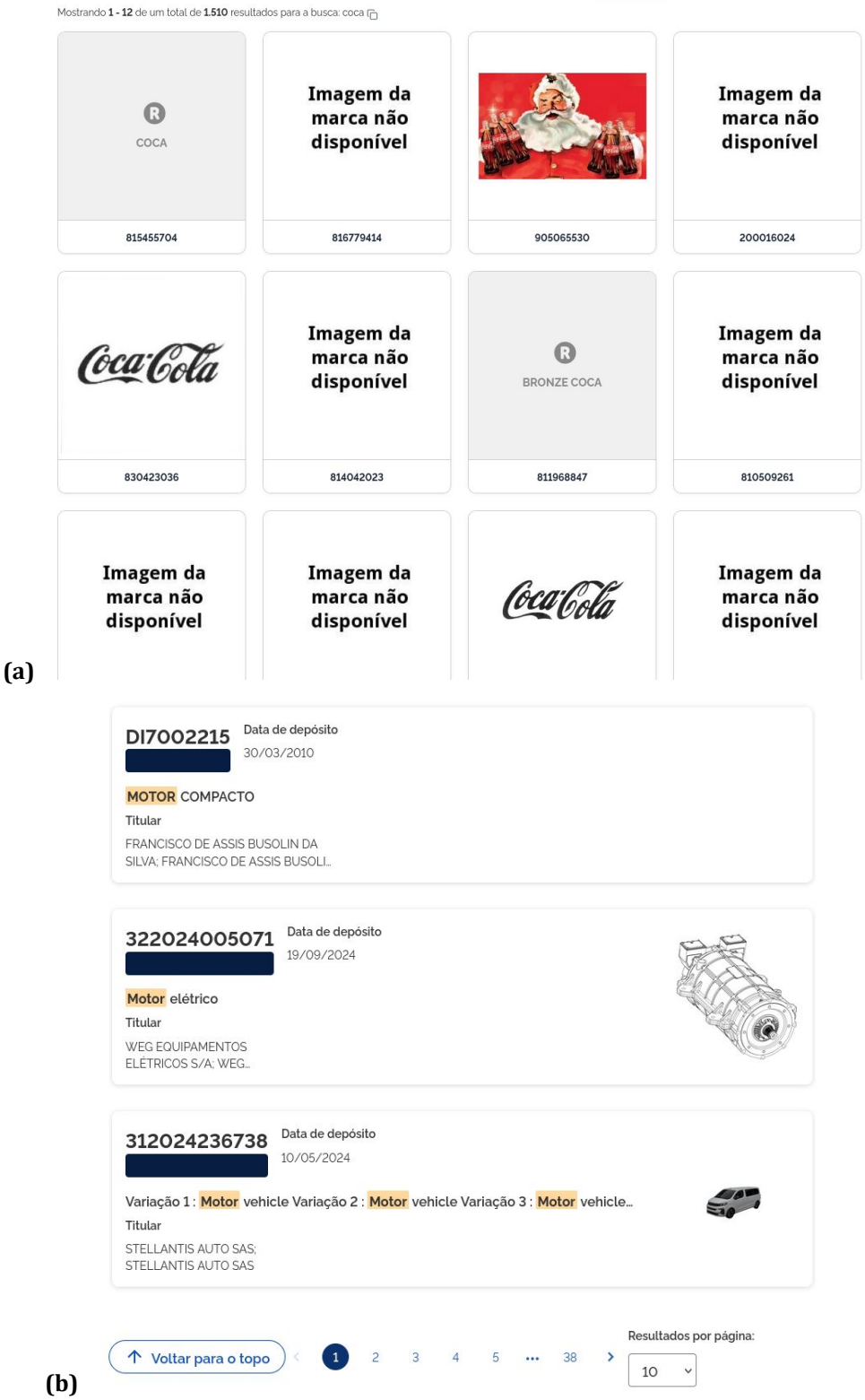


Fonte: elaborado pelo autor (2026).

A apresentação dos resultados foi avaliada também em visualização de grade e em lista, adequadas a naturezas cujo conteúdo é composto predominantemente por

imagens, como marcas e desenhos industriais. A Figura 7 apresenta as duas visualizações.

Figura 7 – Visualizações alternativas de resultados: (a) grade de marcas; (b) lista de desenhos industriais



(a)

(b)

Fonte: elaborado pelo autor (2026).

Quanto à escalabilidade da apresentação, a recuperação sobre conjuntos extensos foi avaliada com paginação e com controle do número de resultados por página, de modo a limitar a carga de renderização, conforme a Figura 8. A utilidade e a qualidade de um artefato de projeto devem ser demonstradas por métodos de avaliação apropriados, entre os quais o exame de atributos como desempenho, usabilidade e consistência (HEVNER *et al.*, 2004). Complementarmente, introduziu-se exibição condensada de campos textuais longos, com recurso de expansão sob demanda.

Figura 8 – Paginação e controle de resultados por página



Fonte: elaborado pelo autor (2026).

Quanto à integração, verificou-se a comunicação do protótipo com a interface de programação de aplicações, com correção do endereço da API, padronização da nomenclatura dos *endpoints* e ajuste da tipagem da paginação, além da disponibilização da consulta em formato de *string* de API para inspeção. Quanto à segurança, atualizou-se dependência da biblioteca de roteamento em razão de correção de vulnerabilidade reportada.

3.3 Estudo exploratório para a integração de inteligência artificial

A frente de integração de inteligência artificial foi conduzida em caráter exploratório dentro da vigência da bolsa, por meio de revisão de literatura orientada a etapas futuras do projeto. A prioridade da pesquisa recaiu sobre a estabilização e o refinamento da camada de apresentação, cujos problemas de desempenho comprometiam a prova de conceito da interface de recuperação. Esse direcionamento delimitou o escopo desta frente ao estudo exploratório, cujos resultados subsidiam a incorporação futura de recuperação assistida ao protótipo.

Realizou-se revisão exploratória de literatura sobre recuperação de informação assistida por modelos de linguagem, com atenção à aplicabilidade dessas abordagens a acervos de propriedade intelectual.

A revisão concentrou-se em três frentes. Na recuperação de patentes assistida por representações vetoriais, a avaliação comparativa de vinte e dois modelos de *embeddings* conduzida por Yousefiramandi e Cooney (2026) indica que a combinação de título, resumo e reivindicações tende a ser a representação textual mais eficaz e que a fusão de métodos densos com o BM25 produz ganhos modestos, além de evidenciar forte degradação de desempenho em consultas fora do domínio de treinamento. Na análise de dados de patentes por meio de redes, Kim *et al.* (2018) exploram redes de coocorrência de classificações IPC e medidas de centralidade para caracterizar tecnologias, abordagem próxima da estrutura de classificação empregada pelos filtros da interface avaliada. Quanto aos fundamentos técnicos dos modelos de linguagem, Bolla, Subbaiah e Kaata (2026) reúnem práticas de ajuste fino, otimização e implantação que subsidiam uma eventual etapa de integração. Em conjunto, essas leituras delimitam o espaço de projeto para a incorporação futura de recuperação assistida por modelos de linguagem aos acervos de propriedade intelectual do projeto.

REFERÊNCIAS

BOLLA, Bharath Kumar; SUBBAIAH, Kalpa; KAATA, Sashi Kiran. **Large Language Model Recipes**: a hands-on guide to fine-tuning, optimization, deployment, and real-world applications. New York: Apress, 2026. E-book. ISBN 979-8-8688-2607-8. DOI 10.1007/979-8-8688-2607-8.

HEVNER, Alan R.; MARCH, Salvatore T.; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. **MIS Quarterly**, [*S. l.*], v. 28, n. 1, p. 75-105, 2004.

KIM, Keungoui; JUNG, Sungdo; HWANG, Junseok; HONG, Ahreum. A dynamic framework for analyzing technology standardisation using network analysis and game theory. **Technology Analysis & Strategic Management**, London, v. 30, n. 5, p. 540-555, 2018. DOI 10.1080/09537325.2017.1340639. Disponível em: http://dx.doi.org/10.1080/09537325.2017.1340639. Acesso em: 3 jul. 2026.

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A.; CHATTERJEE, Samir. A design science research methodology for information systems research. **Journal of Management Information Systems**, [*S. l.*], v. 24, n. 3, p. 45-77, 2007. DOI 10.2753/MIS0742-1222240302.

TUNKELANG, Daniel. **Faceted search**. San Rafael: Morgan & Claypool, 2009. (Synthesis Lectures on Information Concepts, Retrieval, and Services, 5).

YOUSEFIRAMANDI, Amirhossein; COONEY, Ciarán. **Benchmarking patent embeddings**: a multi-task evaluation of 22 models across retrieval, classification, and

clustering. [*S. l.*]: arXiv, 2026. Preprint. arXiv:2605.24297. Disponível em: https://arxiv.org/abs/2605.24297. Acesso em: 3 jul. 2026.

gov.br
Documento assinado digitalmente
RENAN CARNEIRO BATISTA
Data: 21/07/2026 13:23:59-0300
Verifique em https://validar.iti.gov.br

Renan Carneiro Batista
Pesquisador

De acordo:

gov.br
Documento assinado digitalmente
WASHINGTON LUIS RIBEIRO DE CARVALHO SEG
Data: 21/07/2026 14:17:51-0300
Verifique em https://validar.iti.gov.br

Washington Luís Ribeiro de Carvalho Segundo
Coordenador do projeto