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 da pesquisadora:	**Tainá Regly de Moura Souza**		
Vigência da Bolsa de Pesquisa:	**01/06/2026**	**Até:**	**31/08/2026**
Encerramento em:	**31/08/2026**		
Ref. Fundação:	**31849**	**Ref. Ibict:**	**0677/2025**

Descrição das Atividades da Bolsa de Pesquisa:

1 - Monitoramento técnico-científico e sistematização da execução da pesquisa. Metas relacionadas: M1, M2 e M3; 2 - Levantamento, análise documental e apoio à produção técnico-científica do projeto. Metas relacionadas: M2 e M7; 3 - Elaboração do relatório das atividades desenvolvidas.

1 INTRODUÇÃO

A recuperação da informação ocupa posição central nos sistemas informacionais ao articular processos de organização, representação, armazenamento e acesso aos registros de conhecimento. Harold Borko, em 1968, elaborou uma definição do conceito de Ciência da Informação (CI) que faz frente a esses processos e que impactam diretamente no acesso e disseminação de informações. Para o referido autor, a CI compreende o estudo das propriedades e do comportamento da informação, das forças que governam seu fluxo e dos meios empregados para otimizar seu processamento e uso.

Quando aplicada às patentes, a recuperação de informações apresenta problemáticas concernentes ao seu próprio domínio, tais como: documentos extensos, vocabulário especializado e diferentes formas de descrição de uma mesma solução técnica tornam a recuperação uma tarefa que exige métodos próprios e sensíveis às características desse corpus (Shalaby; Zadrozny, 2019). É nesse contexto que se insere o projeto ***ServiçosPI: inovação e eficiência na recuperação de informações de propriedade intelectual do INPI***, desenvolvido a partir da parceria entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e o Instituto Nacional da Propriedade Industrial (INPI), com o propósito de pesquisar e implementar uma solução voltada à organização, ao tratamento e à recuperação das informações de propriedade intelectual.

No decorrer da execução do projeto ServiçosPI, o projeto Patente.IA se constituiu como uma frente diretamente articulada aos avanços já produzidos no projeto, concentrando-se na estruturação dos dados de patentes e em sua preparação para mecanismos avançados de recuperação. A proposta investiga métodos para a estruturação e o enriquecimento de uma base de dados de informação tecnológica composta por documentos de patentes em formato XML, em conformidade com o padrão ST.36 da Organização Mundial da Propriedade Intelectual (OMPI) e seu Tipo de Documento (DTD) associado, de modo a viabilizar estudos sobre modelos de indexação e o desenvolvimento de funcionalidades avançadas de sistemas de busca baseados em inteligência artificial, incluindo mecanismos de análise de similaridade textual e visual

aplicados a textos, imagens e desenhos técnicos. A interface entre os dois projetos pode ser identificada a partir da incorporação de modelos, protótipos, componentes e experiências anteriormente desenvolvidos no ServiçosPI, buscando preservar a continuidade da pesquisa e evitar a duplicação de esforços.

Uma vez que a possibilidade de desenvolver mecanismos de recuperação relacionados às patentes depende da existência de dados consistentes e estruturados, o padrão ST.36 foi adotado como referência pela OMPI (OMPI, 2004) e, consequentemente, pelo INPI. Essa padronização garante que a extração de informações permita a representação das relações semânticas presentes nos textos, das reivindicações, das imagens e dos demais componentes possam ser efetivamente explorados.

A atuação da bolsista de junho a agosto de 2026 ocorreu na interface entre a dimensão técnico-científica e a organização necessária à sua execução. Concentrou-se nas atividades no acompanhamento das reuniões de trabalho, no registro e na sistematização das discussões, na gestão da comunicação entre os participantes, no apoio à integração de novos bolsistas e pesquisadores e no acompanhamento dos encaminhamentos definidos pela equipe. Nesse sentido, as atividades realizadas deram suporte à memória técnica e à continuidade do trabalho coletivo, vinculando-se especialmente às metas do ServiçosPI relacionadas ao levantamento, à sistematização e à análise das informações que subsidiam o desenvolvimento da solução. A seguir será apresentado o método adotado, bem como os resultados obtidos com a pesquisa.

2 MÉTODO

Os procedimentos metodológicos adotados durante a vigência desta bolsa consistiram na articulação entre a gestão da memória documental, a gestão dos de fluxos comunicacionais e o apoio à modelagem arquitetural do projeto Patente.IA, vinculado à iniciativa ServiçosPI. A pesquisa, que é voltada para a inovação tecnológica para a recuperação de informações de propriedade industrial, exigiu um suporte metodológico que garantisse a integridade das decisões e a sincronia entre as equipes do Ibict e do INPI.

A preservação da memória técnica do projeto foi operacionalizada por meio da gravação das reuniões semanais de alinhamento e da elaboração de atas[1]. Esse procedimento funciona como um recurso para a recuperação de informações acerca do projeto, permitindo o monitoramento de suas fases de execução. Já a gestão dos convites e calendários institucionais atuou como o eixo ordenador desses encontros, assegurando a participação de especialistas nas áreas críticas do projeto.

O monitoramento das atividades e o acompanhamento dos prazos estabelecidos no cronograma de execução permitiram mitigar gargalos operacionais e garantir o fluxo das demandas entre o Ibict e a diretoria do INPI. Esse processo contribuiu também para a contribuição para a elaboração e desenho do fluxograma da base do projeto Patente.IA, cuja modelagem permitiu alinhar a estruturação dos dados ao padrão internacional ST36 da OMPI, reforçando o rigor científico e a aplicabilidade prática das soluções desenvolvidas.

Além disso, a ampliação da equipe ao longo do período demandou atividades voltadas à integração dos novos bolsistas e pesquisadores às rotinas de trabalho já estabelecidas. Nesse sentido, foram prestadas orientações relacionadas aos canais de comunicação adotados pela equipe, aos ambientes e sistemas necessários para o desenvolvimento das atividades e à localização da documentação compartilhada. Também foram acompanhados os procedimentos referentes aos termos de sigilo e responsabilidade exigidos para o acesso a determinados recursos e informações vinculados ao projeto.

A seguir são apresentados os resultados da pesquisa.

3 RESULTADOS

As atividades desenvolvidas durante o período permitiram acompanhar a evolução do projeto, preservando o registro das decisões tomadas pela equipe no desenvolvimento da pesquisa. A atuação da bolsista concentrou-se na documentação do processo, no acompanhamento das atividades definidas coletivamente, na integração de

[1] Considerando o caráter sensível da pesquisa relacionada às patentes e à assinatura de um termo de sigilo e responsabilidade, as atas redigidas não serão disponibilizadas neste relatório.

novos participantes e na representação do fluxo informacional discutido ao longo das reuniões.

3.1 Acompanhamento e documentação das atividades

Durante o período da bolsa, realizou-se o acompanhamento de reuniões de alinhamento entre as equipes do Ibict e do INPI, realizadas entre junho e agosto de 2026. Para cada encontro, foi feita a gravação da reunião e elaborada uma ata que continha informações acerca dos participantes, da síntese das discussões com os principais encaminhamentos e, quando pertinente, a indicação das ações e de seus respectivos responsáveis.

A produção dessa documentação permitiu acompanhar o amadurecimento das questões tratadas pela equipe. As discussões passaram pela definição da estrutura dos dados de patentes e sua adequação ao padrão ST36, pelo tratamento dos documentos históricos, pelos processos de extração e validação dos PDFs e, posteriormente, pela articulação dessas etapas com mecanismos de busca. Também foi possível acompanhar as discussões sobre integração com o Data Lake e sobre as estratégias consideradas para o tratamento do passivo histórico de patentes. As atas produzidas também serviram como apoio ao acompanhamento dos encaminhamentos definidos nas reuniões, pois a partir desses registros, foram retomadas atividades pendentes e manteve-se contato com os integrantes responsáveis, dando continuidade às decisões tomadas e sua execução nas semanas seguintes.

3.2 Apoio à integração da equipe

A ampliação da equipe durante o período exigiu a integração de novos bolsistas às rotinas já estabelecidas. Nesse processo, os pesquisadores foram orientados quanto aos canais de comunicação, à documentação compartilhada, aos sistemas utilizados no projeto e aos procedimentos relacionados aos termos de sigilo e responsabilidade.

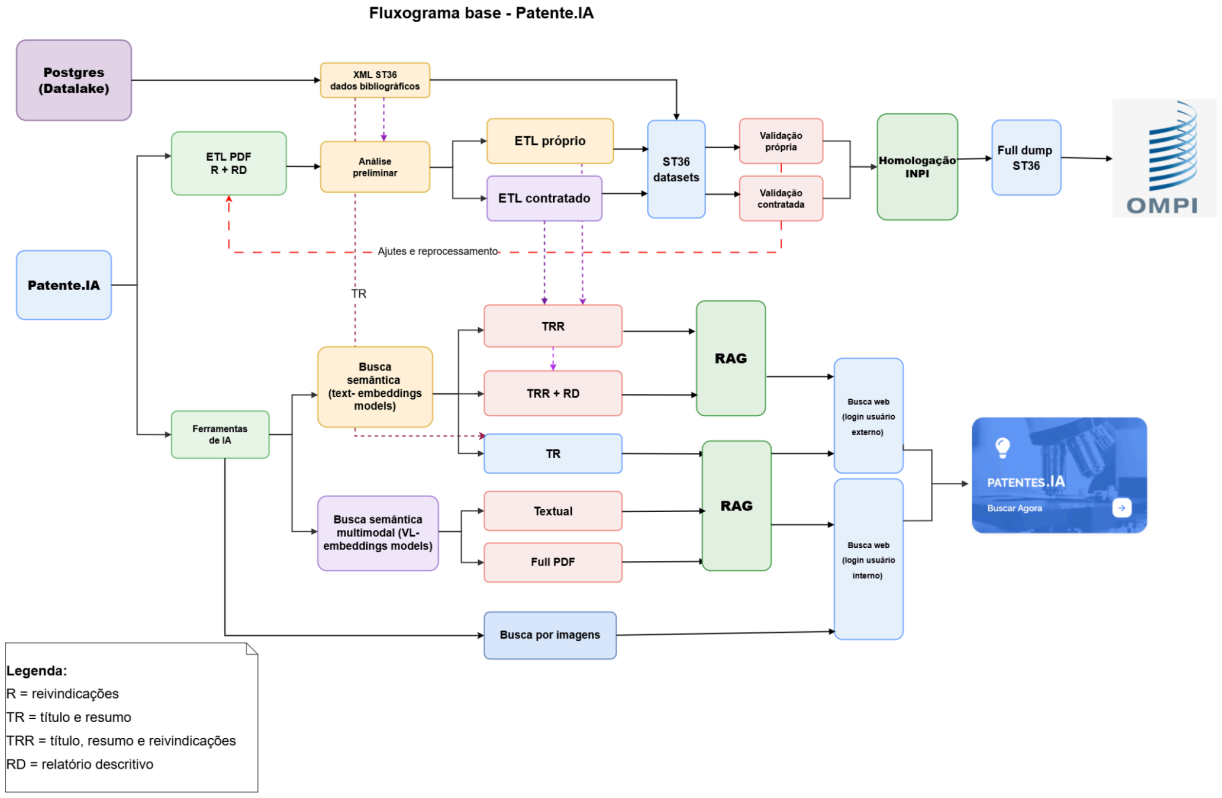
Essa atuação acompanhou um momento de expansão do projeto, marcado pela incorporação de profissionais com diferentes especialidades, entre elas inteligência

artificial e metadados. A organização dos acessos e das informações de referência favoreceu a inserção desses integrantes nas atividades que já se encontravam em desenvolvimento.

3.3 Sistematização do fluxo do Patente.IA

Como resultado da sistematização das discussões realizadas pela equipe, contribuiu-se com a elaboração do fluxograma inicial do projeto Patente.IA, apresentado na Figura 1 a seguir.

Figura 1 - Fluxograma inicial do Patente.IA



Fonte: dados da pesquisa (2026).

O fluxograma apresentado na Figura 1 reúne os principais processos relacionados à estruturação e à recuperação dos dados de patentes. O desenho relaciona as fontes de dados, os processos de ETL e tratamento dos documentos, a geração e validação dos

conjuntos em ST36, a homologação pelo INPI e as diferentes possibilidades de recuperação das informações. Nesse fluxograma também estão representadas as frentes de busca semântica textual e multimodal, busca por imagens e utilização de RAG, considerando diferentes partes do documento de patente, como título, resumo, reivindicações, relatório descritivo e PDF integral.

A partir da elaboração desse fluxograma buscou-se tornar visíveis as relações entre componentes que haviam sido discutidos progressivamente nas reuniões, facilitando a compreensão do fluxo geral e das dependências existentes nas etapas identificadas.

3.4 Contribuição das atividades às metas do projeto

As atividades desenvolvidas no período da bolsa contribuíram para diferentes metas previstas no plano de trabalho do ServiçosPI. O Quadro 1 sintetiza essa relação, considerando as atividades previstas para a bolsa e os resultados apresentados.

Quadro 1 - Contribuição das atividades desenvolvidas às metas do projeto ServiçosPI.

Meta	Contribuição das atividades
M1 - Identificação e avaliação dos serviços e funcionalidades existentes	Acompanhamento e sistematização das discussões sobre ferramentas, fluxos e procedimentos já utilizados no tratamento e na recuperação das informações de patentes.
M2 - Levantamento do cenário informacional relacionado aos sistemas existentes	Organização das informações produzidas nas reuniões sobre ST36, processamento documental, busca semântica, multimodalidade e integração entre diferentes componentes do sistema.
M3 - Análise informacional das bases de dados do INPI	Acompanhamento das discussões sobre estrutura, tratamento, validação e fluxo dos dados de patentes, com contribuição para a elaboração do fluxograma-base do Patente.IA.

M7 - Disseminação dos conhecimentos produzidos no âmbito do projeto	Produção das atas, organização da memória técnica e consolidação das atividades no relatório final.

Fonte: elaboração própria (2026).

Conforme exemplificado no Quadro 1, as atividades realizadas contribuíram para o monitoramento e a sistematização da execução do projeto, especialmente no que se refere às metas M1, M2 e M3, relacionadas ao levantamento e à organização das informações necessárias ao desenvolvimento da solução, e à M7, associada à documentação e disseminação dos conhecimentos produzidos.

A contribuição ocorreu sobretudo pela construção da memória técnica da pesquisa, pelo acompanhamento dos encaminhamentos e pela organização das informações produzidas no decorrer da execução do projeto.

REFERÊNCIAS

BORKO, Harold. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, 1968. DOI: https://doi.org/10.1002/asi.5090190103

ORGANIZAÇÃO MUNDIAL DA PROPRIEDADE INTELECTUAL (OMPI). WIPO Standard ST.36: recommendation for the processing of patent documents using XML (Extensible Markup Language). Geneva: WIPO, 2004. Disponível em: https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=31992&utm_source=chatgpt.com. Acesso em: 17 ago. 2026.

SHALABY, Walid; ZADROZNY, Wlodek. Patent retrieval: a literature review. **Knowledge and Information Systems**, v. 61, n. 2, p. 631-660, 2019. DOI: https://doi.org/10.1007/s10115-018-1322-7

Documento assinado digitalmente
gov.br
TAINA REGLY DE MOURA SOUZA
Data: 23/09/2026 11:30:24-0300
Verifique em https://validar.iti.gov.br

Taina Regly de Moura Souza
Pesquisadora

De acordo:

Documento assinado digitalmente
gov.br
WASHINGTON LUIS RIBEIRO DE CARVALHO SEG
Data: 23/09/2026 12:17:25-0300
Verifique em https://validar.iti.gov.br

Washington Luís Ribeiro de Carvalho Segundo
Coordenador do projeto