



**UNIVERSIDADE FEDERAL DO TOCANTINS
CAMPUS UNIVERSITÁRIO DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL DE SISTEMAS**

WYLLKER SOUSA CRUZ

**INTELIGÊNCIA COMPUTACIONAL APLICADA À ANÁLISE TEXTUAL PARA
DETECÇÃO DE LITIGÂNCIA PREDATÓRIA NO TRIBUNAL DE JUSTIÇA DO
TOCANTINS**

**Palmas, TO
2025**

Wylker Sousa Cruz

**Inteligência Computacional Aplicada à Análise Textual para Detecção de Litigância
Predatória no Tribunal de Justiça do Tocantins**

Dissertação apresentada ao Programa de Pós-Graduação em Mestrado Profissional em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins, como pré-requisito parcial para obtenção do título de Mestre em Modelagem Computacional de Sistemas.

Orientadora: Profa. Dra. Daniela Mascarenhas de Queiroz Trevisan

**Palmas, TO
2025**

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

- C957i Cruz, Wylker Sousa.
Inteligência Computacional Aplicada à Análise Textual para Detecção de Litigância Predatória no Tribunal de Justiça do Tocantins. / Wylker Sousa Cruz. – Palmas, TO, 2025.
83 f.
Dissertação (Mestrado Profissional) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado Profissional) em Governança e Transformação Digital - PPGGTD, 2025.
Orientadora : Daniela Mascarenhas de Queiroz Trevisan
1. Litigância Predatória. 2. Processamento de Linguagem Natural. 3. Inteligência Artificial. 4. Tribunal de Justiça do Tocantins. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

Wylker Sousa Cruz

**Inteligência Computacional Aplicada à Análise Textual para Detecção de Litigância
Predatória no Tribunal de Justiça do Tocantins**

Dissertação apresentada ao Programa de Pós-Graduação em Mestrado Profissional em Modelagem Computacional de Sistemas da Universidade Federal do Tocantins, como pré-requisito parcial para obtenção do título de Mestre em Modelagem Computacional de Sistemas.

Data de aprovação: 30 / 06 / 2025

Banca Examinadora

Doutora. Daniela Mascarenhas de Queiroz Trevisan, PPGMCS/UFT – Membro Interno
(Orientadora)

Doutor. Gentil Veloso Barbosa – PPGMCS/UFT – Membro Interno

Doutor Vilson Soares de Siqueira, Membro Externo

AGRADECIMENTOS

Agradeço, primeiramente, à minha esposa Larissa, pelo apoio incondicional, pela paciência nos momentos difíceis e por sempre acreditar no meu potencial. Estendo minha gratidão à minha família, que é a base de tudo o que sou e realizo, fonte constante de força e inspiração. Aos meus colegas de trabalho, que me incentivaram desde o início desta jornada, deixo meu sincero reconhecimento — em especial ao amigo Ângelo Stacciarini, cuja colaboração foi fundamental para o desenvolvimento deste projeto.

RESUMO

O presente trabalho apresenta o desenvolvimento da ferramenta Tanatose, concebida para apoiar a detecção automatizada de litigâncias predatórias no sistema judicial eletrônico eProc, utilizado pelo Tribunal de Justiça do Estado do Tocantins. A crescente judicialização e a facilidade proporcionada pelo processo eletrônico viabilizaram práticas abusivas, como o ajuizamento massivo de ações com petições iniciais substancialmente idênticas. Diante desse cenário, foi proposta uma solução baseada em técnicas de Inteligência Artificial, Mineração de Dados e Processamento de Linguagem Natural, com foco na comparação textual entre documentos judiciais. A ferramenta Tanatose realiza a baixa automatizada de petições iniciais via Modelo Nacional de Interoperabilidade (MNI), extrai e processa os textos, aplica algoritmos de similaridade textual e gera relatórios de apoio à auditoria processual. A validação experimental, realizada sobre um conjunto de 304 processos reais, demonstrou redução significativa do tempo de triagem, aumento da assertividade na identificação de padrões repetitivos e aprimoramento da qualidade das informações auditadas. A pesquisa evidencia a viabilidade da aplicação de técnicas de inteligência artificial no ambiente jurídico, respeitando os princípios da transparência, rastreabilidade e segurança de dados. Além disso, aponta caminhos para futuras melhorias, como o armazenamento histórico das auditorias e a adoção de modelos avançados de aprendizado de máquina, ampliando o potencial de combate à litigância predatória no Poder Judiciário.

Palavras-chave: Litigância Predatória. Processamento de Linguagem Natural. Inteligência Artificial. Tribunal de Justiça do Tocantins

ABSTRACT

This study presents the development of the Tanatose tool, designed to support the automated detection of predatory litigation within the eProc electronic judicial system, used by the Court of Justice of the State of Tocantins. The increasing judicialization and the ease provided by electronic proceedings have enabled abusive practices, such as the mass filing of lawsuits with substantially identical initial petitions. In this context, a solution based on techniques of Artificial Intelligence, Data Mining, and Natural Language Processing was proposed, focusing on textual comparison between judicial documents. The Tanatose tool automates the retrieval of initial petitions via the National Interoperability Model (MNI), extracts and processes the texts, applies textual similarity algorithms, and generates reports to support procedural audits. Experimental validation, conducted on a set of 304 real cases, demonstrated a significant reduction in screening time, increased accuracy in identifying repetitive patterns, and improvement in the quality of audited information. The research highlights the feasibility of applying artificial intelligence techniques in the legal environment, while respecting the principles of transparency, traceability, and data security. Additionally, it suggests future enhancements such as the historical storage of audits and the adoption of advanced machine learning models, expanding the potential for combating predatory litigation within the Judiciary.

Keywords: Predatory Litigation. Natural Language Processing. Artificial Intelligence. Court of Justice of Tocantins

LISTA DE FIGURAS

Figura 1 - Fluxo de funcionamento da ferramenta Tanatose.....	33
Figura 2 - Diagrama de Caso de Uso.....	39
Figura 3 - Diagrama de Sequência	40
Figura 4 - Diagrama de Banco de Dados.....	43
Figura 5 - Baixa de Processo	51
Figura 6 - Extração de Texto	51
Figura 7 - Análise de Dados	53
Figura 8 - Resultados - ranking de pares	53
Figura 9 – Filtro Customizado.....	54
Figura 10 - Baixa dos processos	56
Figura 11 - Extração Textual.....	57
Figura 12 - Análise Dados.....	58
Figura 13 - Resultado Ranking.....	59
Figura 14 - Resultados Filtros	61

LISTA DE GRÁFICOS

Gráfico 1 - Distância Euclidiana	27
Gráfico 2 - Coeficiente de Jaccard	27
Gráfico 3 - Similaridade do Cosseno.....	28
Gráfico 4 - Faixa de Similaridade.....	54
Gráfico 5 - Resultado Faixa Similaridade	60
Gráfico 6 - Distribuição de Processos por Similaridade Textual	62
Gráfico 7 - Comparativo de Tempo de Análise de Petições Iniciais em Horas	63
Gráfico 8 - Comparativo de Custo da Análise de Petições Iniciais	63

LISTA DE QUADROS

Quadro 1 - Etapas executadas.....	55
-----------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

AJUFESC	Associação dos Juízes Federais de Santa Catarina
BI	Business Intelligence
BoW	Bag of Words
CINUGEP	Central de Inteligência da Unidade de Gestão Estratégica de Processos
CNJ	Conselho Nacional de Justiça
eProc	Sistema Eletrônico de Processos Judiciais
LGPD	Lei Geral de Proteção de Dados Pessoais
MNI	Modelo Nacional de Interoperabilidade
NLP	Natural Language Processing (Processamento de Linguagem Natural)
OCR	Optical Character Recognition (Reconhecimento Óptico de Caracteres)
PDF	Portable Document Format
PLN	Processamento de Linguagem Natural
RF	Requisito Funcional
RNF	Requisito Não Funcional
SEI	Sistema Eletrônico de Informações
TJDFT	Tribunal de Justiça do Distrito Federal e dos Territórios
TJTO	Tribunal de Justiça do Estado do Tocantins
TRF4	Tribunal Regional Federal da 4ª Região

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Proposta Da Pesquisa	13
1.2	O Processo Judicial Eletrônico e Seus Impactos No Judiciário	13
1.3	Litigância Predatória: Conceito e Implicações	14
1.4	Litigância Predatória no Tribunal de Justiça do Tocantins	15
1.5	Problema de Pesquisa	15
1.6	Hipóteses	16
1.7	Delimitação da Pesquisa	16
1.8	Justificativa.....	16
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	O Processo Judicial Eletrônico (Eproc).....	18
2.2	Litigância Predatória.....	19
2.3	Inteligência Computacional Aplicada ao Direito	21
2.4	Mineração de Dados em Documentos Jurídicos	22
2.5	Jurídicos Processamento de Linguagem Natural (PLN)	23
2.6	Reconhecimento Óptico De Caracteres (OCR)	25
2.7	Similaridade Textual.....	26
2.8	Técnicas de Agrupamento e Detecção de Padrões Em Textos.....	29
2.9	Soluções Existentes Para Análise Automatizada de Documentos Jurídicos	30
2.10	Tecnologias Utilizadas no Projeto Tanatose	31
3	METODOLOGIA.....	33
3.1	Tipo e Abordagem da Pesquisa	33
3.2	Análise de Requisitos	34
3.2.1	Necessidade do Usuário.....	34
3.2.2	Objetivo do Sistema	34
3.2.3	Requisitos Funcionais Esperados	35
3.2.4	Requisitos Não Funcionais Esperados.....	35
3.2.5	Restrições Identificadas.....	35
3.3	Estratégia de Desenvolvimento da Ferramenta	36
3.3.1	Requisitos do Sistema.....	37
3.3.2	Diagramas.....	39
3.3.3	Etapas de Funcionamento da Ferramenta.....	41
3.3.4	Projeto Lógico	42
3.3.5	Modelo de Desenvolvimento Adotado	43
3.4	Tecnologias, Ferramentas e Ambiente de Desenvolvimento	44
3.5	Validação da Ferramenta	45
3.6	Limitações e Considerações Éticas	46
4	RESULTADOS E DISCUSSÃO	48
4.1	Visão Geral Da Ferramenta Tanatose.....	48
4.2	fluxo De Funcionamento Da Ferramenta	49
4.2.1	Etapa 1: Baixa de Processos	50
4.2.2	Etapa 2: Extração de Textos	51
4.2.3	Etapa 3: Análise de Similaridade Textual	52
4.2.4	Etapa 4: Geração e Organização dos Resultados.....	53
4.3	Estudo de Caso: Aplicação Prática Com Dados Reais	54
4.3.1	Baixa de Petições Iniciais	56
4.3.2	Extração de Textos	56

4.3.3	Análise de Similaridade Textual.....	57
4.3.4	Relatórios e Visualização dos Resultados	58
4.4	Avaliação dos Ganhos Operacionais e Considerações Finais	61
5	CONSIDERAÇÕES FINAIS	65
5.1	Problema e Solução Proposta	65
5.2	Contribuições da Pesquisa	65
5.3	Limitações do Estudo	66
5.4	Aspectos Éticos	66
5.5	Propostas de Trabalhos Futuros.....	66
	REFERÊNCIAS	68
6	APÊNDICE	72
6.1	Especificação de Casos de Uso	72
6.2	Dicionário de Dados	77
6.3	Comparação de Dados de Técnicas de Similaridade.....	80

1 INTRODUÇÃO

O avanço das tecnologias da informação transformou a dinâmica do Poder Judiciário brasileiro, especialmente com a adoção do processo judicial eletrônico. Se, por um lado, essa inovação trouxe agilidade e eficiência à prestação jurisdicional, por outro, abriu espaço para práticas processuais abusivas, como a litigância predatória. Neste cenário, a inteligência computacional desponta como uma ferramenta estratégica capaz de apoiar a detecção automatizada de padrões repetitivos, promovendo maior racionalidade e segurança no fluxo processual.

Para organizar o raciocínio que fundamenta esta pesquisa, a introdução está estruturada em seções que contextualizam o tema, conceituam o problema de pesquisa, delimitam o escopo do estudo, apresentam as hipóteses formuladas e justificam a relevância da investigação proposta.

1.1 Proposta da Pesquisa

Este trabalho propõe o desenvolvimento de uma ferramenta computacional capaz de identificar automaticamente petições iniciais com elevado grau de similaridade textual no sistema judicial eletrônico eProc, utilizado pelo Tribunal de Justiça do Estado do Tocantins. A proposta surge como resposta à dificuldade enfrentada pelos servidores na identificação manual de demandas repetitivas, característica típica da litigância predatória.

A solução concebida adota técnicas de inteligência computacional, processamento de linguagem natural e mineração de dados para realizar a análise textual automatizada das petições. Ao aplicar algoritmos de similaridade, a ferramenta permite agrupar processos com padrões repetitivos, fornecendo relatórios que apoiam auditorias e triagens processuais. O estudo abrange o desenvolvimento, aplicação e validação prática dessa ferramenta, avaliando sua eficácia no enfrentamento de práticas abusivas no ambiente judicial eletrônico.

1.2 O Processo Judicial Eletrônico e seus impactos no Judiciário

A informatização dos processos judiciais no Brasil foi consolidada com a promulgação da Lei nº 11.419, de 2006, que dispõe sobre o uso de meios eletrônicos na tramitação de ações judiciais. No Estado do Tocantins, essa transformação se concretizou com a implantação do

sistema eProc, regulamentado pela Instrução Normativa nº 5 de 2011 no âmbito do Tribunal de Justiça do Tocantins.

O eProc foi concebido para permitir que todo o fluxo processual, desde o ajuizamento até a prolação da sentença e execução das decisões, transcorresse integralmente de forma digital.

Essa modernização trouxe muitos benefícios. Advogados, servidores, magistrados, membros do Ministério Público e partes passaram a acessar e movimentar processos de qualquer localidade, desde que tenham conexão à internet, o que promoveu significativo ganho de eficiência e ampliou o acesso à justiça. A atuação simultânea e descentralizada tornou-se realidade, superando limitações impostas pela tramitação física.

Contudo, a mesma facilidade tecnológica também abriu margem para práticas processuais abusivas. A litigância predatória, caracterizada pelo ajuizamento massivo e repetitivo de ações com petições substancialmente idênticas, tornou-se um desafio crescente. Este fenômeno compromete a celeridade e a racionalidade do sistema judicial, sobrecarregando a estrutura existente e ameaçando o próprio direito de acesso à justiça. Esse problema será detalhado na próxima seção.

1.3 Litigância Predatória: conceito e implicações

As litigâncias predatórias são caracterizadas por serem “ações ajuizadas em massa, em grande quantidade e, geralmente, em várias comarcas ou varas, sempre com um mesmo tema, com petições quase todas idênticas” (TJDFT, 2022).

A expansão desse fenômeno tem gerado preocupações no meio jurídico. Segundo a revista *Conjur*, os impactos incluem o aumento dos custos processuais e o uso indevido do Judiciário como estratégia comercial por alguns escritórios de advocacia (CONJUR, 2024).

Esse tipo de prática compromete a garantia constitucional do acesso à justiça, pois congestiona o sistema judicial e pode dificultar o julgamento de casos legítimos. O Conselho Nacional de Justiça (CNJ) tem implementado medidas para monitorar e coibir a litigância predatória no Brasil (CNJ, 2024).

1.4 Litigância Predatória no Tribunal de Justiça do Tocantins

No caso específico do Tribunal de Justiça do Estado do Tocantins (TJTO), a litigância predatória se apresenta como um problema concreto e recorrente, conforme identificado pela equipe do Núcleo de Inteligência da Secretaria de Gestão Estratégica de Processos (CINUGEP). A facilidade proporcionada pelo sistema eProc, somada ao grande volume de processos ajuizados anualmente, contribuiu para o surgimento de demandas com elevado grau de repetição textual entre petições iniciais.

Durante as atividades de auditoria realizadas pela CINUGEP, observou-se a dificuldade de detecção manual desses padrões, uma vez que o processo de triagem exige a leitura e análise individualizada de centenas de petições. A limitação de recursos humanos e o tempo demandado para essa atividade tornaram inviável a realização de auditorias completas sem apoio tecnológico, comprometendo a capacidade institucional de enfrentamento das práticas abusivas.

Esse cenário evidenciou a necessidade de desenvolvimento de uma solução automatizada capaz de apoiar a identificação de processos repetitivos, proporcionando maior agilidade, eficiência e segurança na gestão estratégica do acervo processual.

1.5 Problema de Pesquisa

O crescimento do número de ações judiciais eletrônicas no estado do Tocantins, aliado à facilidade de peticionamento garantida pelo sistema eProc, tem permitido o surgimento de práticas conhecidas como litigâncias predatórias. Essas práticas consistem no ajuizamento em massa de ações com o mesmo objeto jurídico, distribuídas entre diversas varas e comarcas, dificultando sua identificação por servidores e magistrados. Diante disso, o problema de pesquisa que se impõe é: É possível usar inteligência computacional para identificar automaticamente, no ambiente do sistema eProc do Tribunal de Justiça do Estado do Tocantins, processos judiciais com elevado grau de similaridade textual, apoiando a detecção ágil e eficaz de litigâncias predatórias?

1.6 Hipóteses

Parte-se da hipótese de que o uso de técnicas de inteligência computacional, como mineração de dados e processamento de linguagem natural, pode ser eficaz na identificação automatizada de petições iniciais que apresentem elevado grau de similaridade textual, caracterizando possíveis litigâncias predatórias.

Supõe-se ainda que, com a utilização da ferramenta proposta, será possível apoiar o trabalho dos servidores do Tribunal de Justiça do Tocantins, proporcionando maior celeridade na triagem processual, redução do retrabalho, melhoria na uniformidade de análise de casos repetitivos e fortalecimento da gestão estratégica do acervo judicial.

1.7 Delimitação da Pesquisa

A proposta apresentada nesta pesquisa está restrita à construção de uma ferramenta computacional com o objetivo de auxiliar os servidores do Poder Judiciário do Estado do Tocantins na identificação de possíveis litigâncias predatórias, com base na similaridade textual entre petições iniciais. A ferramenta, portanto, não realiza julgamento de mérito nem substitui a análise jurídica por parte do magistrado ou servidor, mas se propõe a ser um instrumento de apoio à triagem e análise prévia, reduzindo o tempo necessário para identificar padrões repetitivos de litígios.

Ademais, o escopo deste trabalho está limitado à realidade do sistema eProc em uso no Estado do Tocantins, não incluindo análises aplicadas a outros sistemas processuais ou contextos judiciais de outros entes federativos. Contudo, ressalta-se que a arquitetura da solução permite, futuramente, adaptações para o uso em outros tribunais que operem sob o modelo do eProc Nacional.

1.8 Justificativa

O volume crescente de processos judiciais e o uso sistemático do sistema eletrônico de tramitação trouxeram avanços para o Poder Judiciário brasileiro, como maior agilidade, transparência e economia de recursos. No entanto, esse mesmo avanço tecnológico facilitou o ajuizamento em massa de ações com conteúdo repetitivo, estratégia muitas vezes associada a práticas abusivas, conhecidas como litigâncias predatórias. Esses litígios provocam não apenas a sobrecarga dos servidores e a morosidade processual, mas também colocam em risco a

uniformidade das decisões judiciais, uma vez que ações idênticas podem ser julgadas separadamente, por diferentes juízes, com entendimentos distintos.

Diante desse cenário, a presente pesquisa justifica-se por buscar uma solução tecnológica aplicável ao contexto real do Judiciário tocantinense, oferecendo uma ferramenta que pode contribuir diretamente com a melhoria da gestão e da celeridade processual. O uso de algoritmos de inteligência computacional e análise textual, aplicado à realidade de um tribunal estadual, representa uma contribuição prática e relevante, alinhada aos objetivos constitucionais de eficiência e racionalização dos serviços públicos.

Além disso, a ferramenta desenvolvida é passível de reaproveitamento institucional, com possibilidade de integração ao ambiente de trabalho dos servidores do Tribunal de Justiça do Tocantins. Caso comprovada sua eficácia, abre-se espaço para sua replicação em outras unidades da federação, especialmente aquelas que também adotam o sistema eProc como plataforma processual.

Por fim, a relevância científica desta pesquisa reside na exploração da interface entre a Ciência da Computação e o Direito, um campo ainda em desenvolvimento no Brasil, mas com grande potencial de impacto institucional e social. A proposta de utilização de técnicas de mineração de dados e processamento de linguagem natural em ambiente jurídico demonstra não apenas a viabilidade tecnológica, mas também a preocupação com a explicabilidade dos resultados, elemento essencial no setor público, onde a transparência e a rastreabilidade das decisões são princípios basilares.

Ao desenvolver uma ferramenta que alia inovação tecnológica à preservação de valores institucionais como eficiência, uniformidade e segurança jurídica, esta pesquisa contribui para fortalecer a capacidade analítica do Judiciário e aponta caminhos para novas práticas no enfrentamento da litigância predatória.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O Processo Judicial Eletrônico (eProc)

A informatização do processo judicial no Brasil consolidou-se com a promulgação da Lei nº 11.419, de 19 de dezembro de 2006, que regulamentou o uso de meios eletrônicos na tramitação de processos judiciais, comunicação de atos e transmissão de peças processuais (BRASIL, 2006). Essa legislação representou um marco no esforço de modernização do Judiciário, ao estabelecer bases normativas para a adoção de sistemas eletrônicos de tramitação processual.

O Sistema Eletrônico de Processos Judiciais (eProc) surgiu nesse contexto como uma solução inovadora, desenvolvida pelo Tribunal Regional Federal da 4ª Região (TRF4) em 2003, com o objetivo de informatizar e agilizar a tramitação processual no âmbito da Justiça Federal (TRF4, 2019).

No Estado do Tocantins, a implantação do eProc ocorreu em 2011, quando o Tribunal de Justiça do Estado do Tocantins (TJTO) instituiu o sistema como seu meio oficial de tramitação eletrônica de processos, regulamentado pela Instrução Normativa nº 5, de 24 de outubro de 2011 (PODER JUDICIÁRIO DO ESTADO DO TOCANTINS, 2011).

O eProc é caracterizado por sua interface amigável, pela tramitação integralmente eletrônica dos processos, pela integração com sistemas externos e pela capacidade de automação de tarefas judiciais (TJTO, 2024). Seu modelo modular permite a incorporação de inovações, como recursos de inteligência artificial, integração com bases de dados externas e ferramentas de Business Intelligence (BI), promovendo constante modernização tecnológica.

A arquitetura do sistema garante a interoperabilidade entre órgãos públicos e assegura elevados padrões de segurança da informação e governança de dados, aspectos fundamentais para a tramitação eletrônica de processos judiciais.

A digitalização processual, impulsionada pela implantação do eProc, proporcionou significativa ampliação do acesso à justiça, ao eliminar barreiras físicas, reduzir o tempo de tramitação dos processos e democratizar o acesso às informações processuais (CNJ, 2020). A possibilidade de atuação remota por advogados, partes e servidores trouxe celeridade e transparência à atividade jurisdicional.

Entretanto, a informatização também trouxe novos desafios, como o aumento expressivo do volume de processos, a necessidade de investimentos em infraestrutura tecnológica e o surgimento de fenômenos como a litigância predatória (CNJ, 2020). Tais

desafios impõem a necessidade de aprimoramento contínuo na gestão processual e no uso de tecnologias emergentes.

Em 9 de julho de 2019, o TJTO assinou um novo Termo de Cessão com o TRF4, formalizando a adesão à versão nacional do eProc, que integra diversos tribunais brasileiros em um modelo colaborativo de desenvolvimento e evolução contínua do sistema (TRF4, 2019). Essa integração permitiu ao TJTO o acesso a inovações conjuntas, melhorias na interoperabilidade dos sistemas, reforço na segurança da informação e participação ativa nos grupos de trabalho que discutem e implementam melhorias na plataforma.

Assim, a evolução do eProc no Estado do Tocantins reflete não apenas a modernização tecnológica do Judiciário, mas também seu compromisso institucional com a eficiência, a economicidade e a transparência na prestação jurisdicional.

2.2 Litigância Predatória

A atuação do Poder Judiciário brasileiro tem sido desafiada por uma crescente judicialização de conflitos, intensificada nas últimas décadas em razão das políticas públicas de ampliação do acesso à justiça. Contudo, o avanço do acesso ao Judiciário também foi acompanhado por práticas litigiosas abusivas, destacando-se a chamada litigância predatória (FUX, 2021).

Litigância predatória é o termo utilizado para descrever o ajuizamento massivo de ações judiciais com petições praticamente idênticas, alterando-se apenas dados específicos como o nome da parte e o endereço. Tais ações são frequentemente protocoladas simultaneamente em diversas comarcas e, em geral, apresentam temáticas comuns, notadamente no campo das relações de consumo (FUX, 2021).

Essas características demonstram um uso estratégico e massivo do sistema judicial, com o objetivo de obtenção de vantagens, muitas vezes em prejuízo da efetiva prestação jurisdicional e em detrimento do interesse público.

O fenômeno tem despertado a atenção de diversos tribunais e do Conselho Nacional de Justiça (CNJ), uma vez que impacta diretamente o funcionamento do Judiciário, comprometendo o princípio constitucional da razoável duração do processo e gerando dificuldades operacionais para a entrega célere da justiça (TJDFT, 2022; CNJ, 2020).

Exemplos práticos de litigância predatória incluem escritórios de advocacia que ajuízam milhares de ações consumeristas similares, sobrecarregando os sistemas de justiça e dificultando a análise adequada de cada demanda individual (TJDFT, 2022).

O impacto dessa prática é significativo: a litigância predatória contribui para a sobrecarga dos tribunais, o aumento dos custos processuais, o congestionamento das pautas judiciais e a erosão da confiança pública no sistema judicial. Além disso, essas demandas abusivas podem ser utilizadas como mecanismo de pressão sobre empresas e entes públicos, configurando desvio de finalidade processual e prejudicando a credibilidade institucional do Poder Judiciário (FUX, 2021).

O Tema 1198 do Superior Tribunal de Justiça (STJ) representa um marco importante no enfrentamento da litigância predatória no ordenamento jurídico brasileiro. Ao reconhecer a possibilidade de incidência de multa por litigância de má-fé e a fixação de honorários advocatícios em desfavor de escritórios de advocacia que ajuízam ações repetitivas com petições padronizadas, o STJ sinaliza uma postura institucional firme contra o uso abusivo da máquina judiciária. A tese firmada estabelece que, na hipótese de ajuizamento de demandas em massa, com uso de peças processuais idênticas e sem a devida análise individualizada do caso concreto, o juízo poderá aplicar penalidades não apenas à parte formalmente autora da ação, mas também aos advogados que patrocinam condutas predatórias, quando demonstrada sua participação ativa na estratégia processual abusiva (BRASIL, 2023).

Essa diretriz jurisprudencial busca preservar a integridade do sistema de justiça e coibir o desvirtuamento do acesso à tutela jurisdicional, evitando que o Judiciário seja instrumentalizado como meio de obtenção de vantagens indevidas em prejuízo da coletividade. O julgamento do Tema 1198 reforça a importância da responsabilidade ética na atuação profissional e reconhece que a repetição sistemática e acrítica de demandas compromete os princípios da boa-fé processual, da celeridade e da eficiência da prestação jurisdicional (BRASIL, 2023). No contexto da presente pesquisa, a posição do STJ dialoga diretamente com a necessidade de ferramentas que possibilitem a detecção automatizada de padrões repetitivos, contribuindo com os esforços institucionais para o enfrentamento da litigância predatória.

Nesse cenário, torna-se imprescindível o desenvolvimento e a adoção de mecanismos capazes de identificar padrões repetitivos e suspeitos de litigância. Ferramentas tecnológicas baseadas em inteligência artificial surgem como alternativas promissoras para apoiar a magistratura e os servidores na detecção e na gestão de demandas massificadas, em consonância com os princípios da eficiência, moralidade administrativa e com as diretrizes do uso ético da tecnologia no setor público (CNJ, 2020).

A compreensão aprofundada da lógica que sustenta a litigância predatória é, portanto, fundamental para a construção de soluções que fortaleçam a efetividade da prestação jurisdicional e protejam o sistema judicial de práticas abusivas.

Casos concretos de litigância predatória têm sido cada vez mais identificados pelo Poder Judiciário brasileiro, revelando o impacto sistêmico dessas práticas abusivas. Um exemplo emblemático ocorreu no Tribunal de Justiça de São Paulo (TJSP), onde o Núcleo de Monitoramento de Perfis de Demandas (Numopede) identificou, entre 2016 e 2021, aproximadamente 337 mil processos por ano relacionados a litigância predatória, muitos deles com petições padronizadas, especialmente em ações sobre empréstimos consignados. O volume gerado por essas ações levou a um prejuízo estimado em R\$ 2,7 bilhões anuais aos cofres públicos, além de sobrecarregar significativamente o Judiciário. Em resposta, o TJSP passou a aplicar sanções por litigância de má-fé, além de implementar estratégias de identificação e agrupamento de ações repetitivas, com o objetivo de mitigar os efeitos negativos dessa prática (ANGELLO; VITAL, 2023).

Outro caso notório envolveu uma microempresa que foi alvo de 303 ações trabalhistas, todas com petições semelhantes e representadas pelo mesmo escritório de advocacia. As ações alegavam vínculos empregatícios inexistentes, configurando uma estratégia padronizada e coordenada. Diante da repetição e do padrão das demandas, o Judiciário reconheceu a configuração de litigância predatória, aplicando multas por má-fé processual e encaminhando o caso à Ordem dos Advogados do Brasil (OAB) para apuração disciplinar dos profissionais envolvidos (JOTA, 2024). Esses exemplos evidenciam a urgência de mecanismos automatizados para identificação de padrões repetitivos, como a ferramenta proposta nesta pesquisa, que busca apoiar a triagem técnica de ações potencialmente predatórias.

2.3 Inteligência Computacional aplicada ao Direito

A Inteligência Computacional (IC) constitui um subcampo da Inteligência Artificial (IA) dedicado ao desenvolvimento de sistemas capazes de resolver problemas complexos por meio de métodos inspirados em processos naturais e fenômenos biológicos. Segundo Poole, Mackworth e Goebel (1998), a IC busca a construção de agentes inteligentes, aptos a perceber seu ambiente, tomar decisões e agir racionalmente para alcançar objetivos definidos.

As principais técnicas associadas à Inteligência Computacional incluem redes neurais artificiais, algoritmos genéticos, lógica fuzzy e sistemas de enxame. Essas abordagens modelam o comportamento adaptativo encontrado na natureza para solucionar tarefas que envolvem incerteza, alta dimensionalidade e dados imprecisos (KRUSE et al., 2013).

A aplicação da IC estende-se a diversas áreas do conhecimento, como o reconhecimento de padrões, diagnósticos médicos, previsão de séries temporais e controle de sistemas

dinâmicos complexos (WANG; LIU, 2006). A capacidade adaptativa e a habilidade de aprendizado a partir de dados tornam essas técnicas particularmente adequadas para ambientes onde soluções tradicionais de programação se mostram limitadas.

No contexto jurídico, a Inteligência Computacional tem sido utilizada de forma crescente para otimizar a gestão da informação e apoiar a tomada de decisões judiciais. Aplicações típicas incluem sistemas de recomendação de jurisprudência, análise preditiva de resultados processuais e detecção de padrões em grandes bases de dados judiciais (SILVA; PEREIRA, 2021).

Ferramentas baseadas em redes neurais e algoritmos de aprendizado são empregadas para classificar documentos jurídicos, prever desfechos de processos com base em históricos anteriores e auxiliar na triagem de demandas de massa. A lógica fuzzy, por sua vez, tem sido aplicada para modelar a interpretação de normas jurídicas que envolvem conceitos vagos ou indeterminados.

Além disso, a análise preditiva vem ganhando destaque no ambiente judicial, permitindo identificar, por meio de modelos computacionais, probabilidades de concessão de pedidos, riscos de inadimplência em execuções e tendências jurisprudenciais em diferentes ramos do direito.

O uso de Inteligência Computacional no meio jurídico contribui não apenas para o aumento da eficiência e da produtividade, mas também para a transparência e a previsibilidade das decisões judiciais, elementos essenciais para o fortalecimento do Estado Democrático de Direito.

2.4 Mineração de Dados em Documentos Jurídicos

A Mineração de Dados (MD) é o processo de descoberta de padrões relevantes, correlações e tendências ocultas em grandes volumes de dados, utilizando técnicas combinadas de estatística, aprendizado de máquina e gestão de bancos de dados (HAN; KAMBER; PEI, 2012). O principal objetivo da mineração de dados é extrair informações úteis que não são imediatamente perceptíveis, proporcionando insights valiosos para a tomada de decisões.

Segundo Han, Kamber e Pei (2012), a mineração de dados vai além da simples recuperação de informações, envolvendo a análise exploratória e a construção de modelos que representam o conhecimento embutido nos dados. Essa abordagem é aplicada em diversos setores, como saúde, finanças, marketing, administração pública e no campo jurídico.

O Aprendizado de Máquina (AM) é um dos pilares fundamentais da mineração de dados. Trata-se do desenvolvimento de algoritmos que permitem aos sistemas computacionais aprender e evoluir com base em dados, sem necessidade de reprogramação explícita. Conforme Mitchell (1997), um programa é considerado capaz de aprender quando seu desempenho em determinada tarefa melhora com a experiência adquirida.

A interseção entre MD e AM constitui a base para o desenvolvimento de sistemas preditivos e adaptativos, possibilitando a análise de grandes volumes de dados de forma automatizada e inteligente (WITTEN; FRANK; HALL, 2011).

No âmbito jurídico, a mineração de textos (text mining) tem se consolidado como uma ferramenta estratégica para a otimização da gestão judicial e a análise preditiva de informações processuais.

A aplicação de técnicas de mineração de dados em documentos jurídicos permite, por exemplo, a identificação de padrões de decisões judiciais, a detecção de similaridades entre peças processuais, a classificação automática de processos e a extração de informações relevantes para a construção de bases de precedentes (WITTEN; FRANK; HALL, 2011).

Essas técnicas viabilizam análises complexas em grandes repositórios de processos, auxiliando tribunais e escritórios de advocacia na gestão de seus acervos documentais. Com o uso de mineração de dados, é possível prever tendências jurisprudenciais, identificar perfis de litigância predatória, automatizar a triagem de processos e aprimorar a gestão de recursos humanos e materiais dentro das instituições judiciais.

Além disso, a capacidade de extrair conhecimento estruturado a partir de documentos textuais amplia as possibilidades de implementação de sistemas de suporte à decisão judicial, promovendo maior celeridade, eficiência e transparência na prestação jurisdicional

2.5 Jurídicos Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial (IA) voltada à interpretação, compreensão e geração da linguagem humana por sistemas computacionais. Segundo Jurafsky e Martin (2020), o PLN combina conhecimentos de linguística, ciência da computação e estatística para enfrentar tarefas complexas como tradução automática, análise de sentimentos, resposta a perguntas e sumarização de textos.

O PLN busca construir modelos computacionais capazes de entender não apenas a estrutura superficial da linguagem, mas também seus significados implícitos e variações contextuais. No entanto, enfrenta desafios específicos como a ambiguidade semântica, a

polissemia das palavras e a necessidade de adaptação a contextos variados de uso da linguagem (MANNING; SCHÜTZE, 1999).

Com os avanços em redes neurais profundas, especialmente por meio dos modelos baseados em arquiteturas de transformadores, como o BERT e o GPT, o campo do PLN evoluiu significativamente, proporcionando maior precisão na compreensão e geração automática de textos complexos.

No contexto jurídico, o PLN tem aplicações relevantes na análise de grandes volumes de documentos processuais e jurídicos. Entre as técnicas básicas de PLN aplicáveis à análise de documentos judiciais, destacam-se:

- **Tokenização:** processo de segmentação de um texto em unidades menores, como palavras ou sentenças, facilitando o processamento subsequente.
- **Lematização:** redução das palavras à sua forma canônica (exemplo: "correndo" → "correr"), padronizando os vocábulos para melhorar a análise textual.
- **Extração de Entidades Nomeadas (NER - Named Entity Recognition):** identificação de informações específicas em um texto, como nomes de pessoas, organizações, datas e locais, úteis para catalogação e indexação jurídica.
- **TF-IDF (Term Frequency-Inverse Document Frequency):** Técnica que representa cada texto como um vetor ponderado pela frequência das palavras no documento em relação à frequência das mesmas palavras no conjunto total de documentos. A aplicação do TF-IDF permite capturar a importância relativa dos termos em cada texto, reduzindo a influência de palavras muito comuns que não discriminam bem os documentos (RANGANATHAN, 2008).
- **Bag of Words (BoW):** modelo de representação textual que transforma um texto em um vetor baseado na frequência das palavras que ele contém, desconsiderando a ordem e a estrutura gramatical. Embora simples, o BoW é amplamente utilizado em tarefas de classificação e agrupamento de textos, funcionando como etapa preliminar ao uso de algoritmos de aprendizado. Sua eficácia reside na capacidade de representar grandes volumes de dados textuais de forma estruturada e compatível com métodos estatísticos e computacionais (AGGARWAL, 2018).
- **Análise de Similaridade Textual:** técnica empregada para medir o grau de semelhança entre dois ou mais textos, fundamental para a identificação de peças processuais repetitivas ou litigâncias predatórias.

Essas técnicas permitem não apenas a automação de tarefas repetitivas, mas também a melhoria da eficiência na triagem, classificação e análise de documentos judiciais. Ao aplicar PLN no contexto do Judiciário, busca-se acelerar o tratamento de informações, reduzir o tempo de análise manual e contribuir para uma justiça mais célere e eficiente.

2.6 Reconhecimento Óptico de Caracteres (OCR)

O Reconhecimento Óptico de Caracteres (OCR, do inglês Optical Character Recognition) é uma tecnologia que permite a conversão de documentos digitalizados, imagens ou fotografias de textos manuscritos e impressos em arquivos digitais editáveis e pesquisáveis. Conforme Smith (2007), o OCR transforma imagens de texto em dados estruturados, viabilizando a manipulação e o armazenamento eletrônico das informações.

A importância do OCR reside na sua capacidade de automatizar a digitalização de grandes volumes de documentos físicos, eliminando a necessidade de transcrição manual e ampliando significativamente a eficiência dos processos de gestão documental. Essa tecnologia facilita a indexação, recuperação, análise e compartilhamento de informações, promovendo ganhos expressivos em produtividade e acessibilidade a dados anteriormente restritos a arquivos físicos.

No âmbito jurídico, o OCR desempenha papel fundamental na modernização da gestão processual. A tecnologia é amplamente utilizada para converter autos de processos físicos, petições, despachos e sentenças em formatos digitais compatíveis com sistemas eletrônicos de tramitação, como o eProc.

A aplicação do OCR permite que documentos anteriormente arquivados em meio físico sejam incorporados a bases de dados digitais, possibilitando a realização de buscas textuais, extração de informações relevantes e automação de tarefas de triagem e classificação documental. Além disso, o OCR contribui para a preservação do acervo documental, reduzindo os riscos de deterioração de processos físicos e promovendo maior segurança da informação.

Ferramentas modernas de OCR, integradas a técnicas de Processamento de Linguagem Natural (PLN), vêm sendo utilizadas para realizar análises automáticas de documentos digitalizados, como a identificação de partes, classes processuais e extração de dados estruturados para sistemas de gestão judicial. Tais avanços impulsionam a transformação digital no Judiciário, otimizando fluxos de trabalho, aumentando a transparência e promovendo maior celeridade na prestação jurisdicional.

2.7 Similaridade Textual

A similaridade textual é uma tarefa essencial no campo do Processamento de Linguagem Natural (PLN), consistindo na medição do grau de semelhança entre dois ou mais textos, com o objetivo de avaliar o quanto seus conteúdos convergem semanticamente. Essa técnica é amplamente empregada em aplicações como detecção de plágio, sistemas de recomendação, motores de busca, classificação de documentos e agrupamento temático (MANNING; RAGHAVAN; SCHÜTZE, 2008).

Para realizar essa medição, é necessário representar os textos de maneira estruturada, geralmente por meio de vetores numéricos, de modo que possam ser comparados por métricas matemáticas específicas. Entre os métodos mais consolidados na literatura, destacam-se a **Similaridade do Cosseno**, o **Coefficiente de Jaccard** e a **Distância Euclidiana**, cujas características são brevemente descritas a seguir.

A **Similaridade do Cosseno** calcula o cosseno do ângulo entre dois vetores no espaço vetorial, representando os textos após sua transformação em embeddings numéricos. Essa métrica é especialmente eficaz quando os documentos possuem comprimentos variados, pois considera apenas a direção dos vetores, ignorando sua magnitude. Valores próximos de 1 indicam alta similaridade; valores próximos de 0, baixa similaridade. Devido a essas características, trata-se de uma técnica amplamente utilizada em tarefas de agrupamento textual (MANNING; RAGHAVAN; SCHÜTZE, 2008).

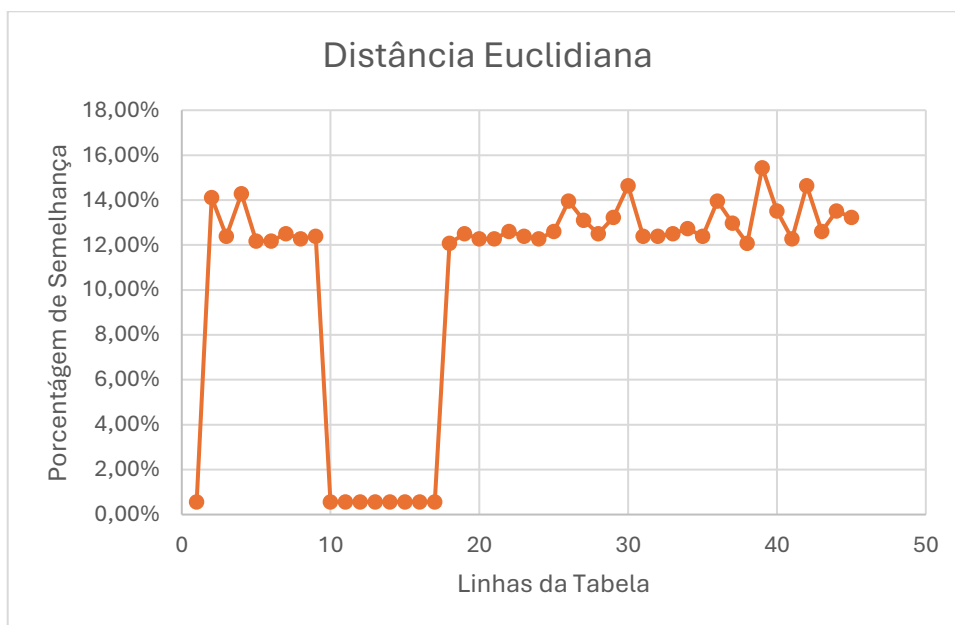
O **Coefficiente de Jaccard**, por sua vez, mede a razão entre a interseção e a união de dois conjuntos. Quando aplicado a textos convertidos em conjuntos de palavras ou n-gramas, o índice de Jaccard revela a proporção de termos compartilhados entre os documentos. Seu valor varia entre 0 e 1, sendo mais adequado para textos curtos e para detectar sobreposições léxicas (JACCARD, 1901; TAN; STEINBACH; KUMAR, 2019).

Já a **Distância Euclidiana** calcula a raiz quadrada da soma das diferenças quadráticas entre os componentes dos vetores que representam os textos. Quanto menor a distância, maior a similaridade. Contudo, essa métrica leva em consideração a magnitude dos vetores, o que pode causar distorções quando há grande variação no tamanho dos textos, como é comum em documentos jurídicos (MANNING; RAGHAVAN; SCHÜTZE, 2008).

A fim de comparar essas três abordagens, foi realizado um experimento utilizando um conjunto composto por 10 petições iniciais judiciais, gerando 45 pares de comparações cruzadas. Os dados foram submetidos às três métricas de similaridade mencionadas, e os

resultados podem ser observados nos gráficos a seguir. Destaca-se que, nos pares compreendidos entre as linhas 2 a 9 e de 18 a 45, as petições são **virtualmente idênticas**, contendo apenas alterações mínimas como o nome da parte e o valor da causa.

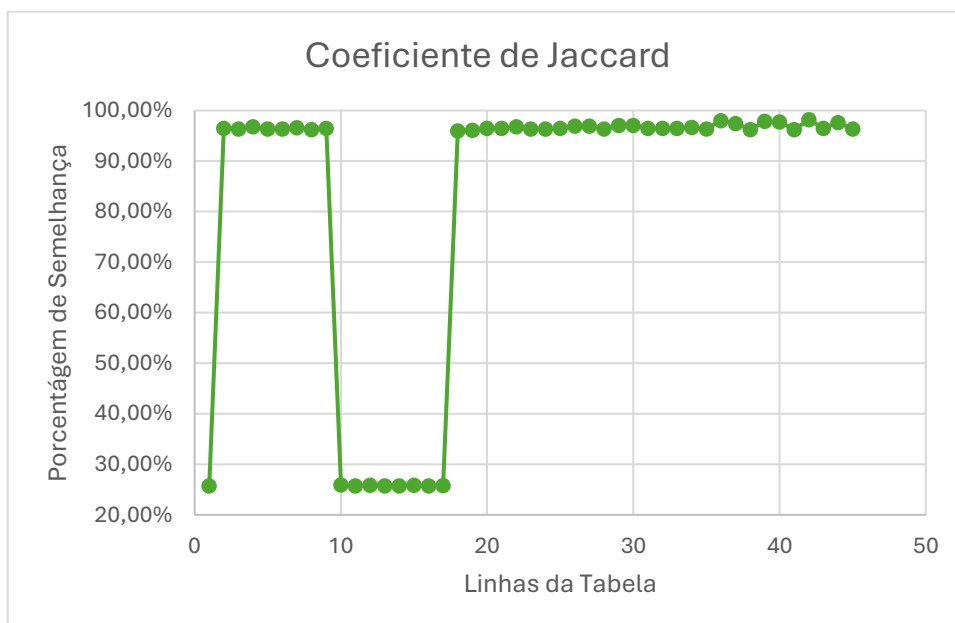
Gráfico 1 - Distância Euclidiana



Fonte: Elaborado pelo autor (2025)

A métrica da Distância Euclidiana não foi eficiente para o propósito em questão, apresentando valores que não refletiram a alta similaridade esperada entre petições praticamente idênticas.

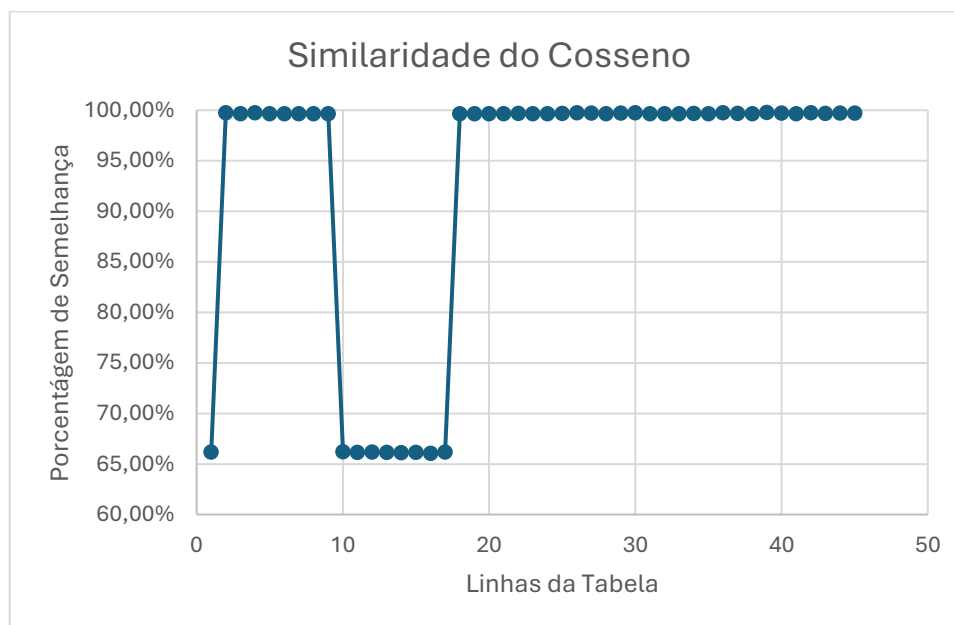
Gráfico 2 - Coeficiente de Jaccard



Fonte: Elaborado pelo autor (2025)

Já a métrica de Jaccard, embora mais precisa que a anterior, ainda apresentou valores ligeiramente abaixo do real esperado, revelando limitações na identificação de similaridade em documentos longos com vocabulário compartilhado parcialmente.

Gráfico 3 - Similaridade do Cosseno



Fonte: Elaborado pelo autor (2025)

A Similaridade do Cosseno foi a técnica que melhor refletiu a realidade dos dados, reconhecendo corretamente os altos níveis de similaridade entre as petições analisadas. Isso confirma sua superioridade técnica no contexto de documentos jurídicos extensos e com padronização estrutural.

Diante dessa análise comparativa, a Similaridade do Cosseno foi a métrica adotada no projeto Tanatose. O sistema processa os textos das petições por meio de técnicas como tokenização, lematização e aplicação do modelo Bag of Words, convertendo-os em vetores de frequência. A seguir, aplica-se o cálculo da similaridade vetorial com base no cosseno do ângulo entre esses vetores.

Essa abordagem revelou-se a mais eficaz para o cenário jurídico em estudo, possibilitando a identificação automatizada de padrões repetitivos, típica da litigância predatória, e oferecendo suporte estratégico à atuação dos servidores responsáveis pela triagem processual. Sua precisão, robustez e interpretabilidade justificam sua escolha como núcleo da análise textual automatizada proposta nesta pesquisa.

2.8 Técnicas de Agrupamento e Detecção de Padrões em Textos

Agrupamento, ou clustering, é a tarefa de organizar um conjunto de objetos em grupos (ou clusters) de itens semelhantes entre si, de forma que objetos pertencentes ao mesmo grupo sejam mais similares uns aos outros do que aos objetos de grupos diferentes (TAN; STEINBACH; KUMAR, 2005).

Entre os métodos tradicionais de agrupamento, destacam-se:

- **K-means:** Técnica de agrupamento particional que divide o conjunto de dados em k grupos, minimizando a distância interna entre os dados do mesmo grupo e maximizando a separação entre grupos diferentes. O número de clusters k deve ser definido previamente, e o algoritmo busca encontrar os centróides ótimos que melhor representem cada grupo (TAN; STEINBACH; KUMAR, 2005).
- **Hierarchical clustering:** Método que constrói uma hierarquia de agrupamentos por meio de uma abordagem bottom-up (aglomeração) ou top-down (divisão). No hierarchical clustering aglomerativo, cada objeto inicia em seu próprio grupo, e os grupos são gradualmente unidos com base em sua similaridade, formando uma estrutura em árvore chamada dendrograma (TAN; STEINBACH; KUMAR, 2005).

Esses métodos são amplamente utilizados em análise exploratória de dados, pois permitem revelar estruturas subjacentes, segmentar bases de dados e facilitar a interpretação de grandes conjuntos de informações.

Quando aplicado à similaridade textual, o agrupamento organiza documentos de acordo com os níveis de semelhança entre seus conteúdos, utilizando métricas como a Similaridade do Cosseno para medir a proximidade semântica entre textos (MANNING; RAGHAVAN; SCHÜTZE, 2008).

A abordagem consiste em representar cada texto como um vetor em um espaço multidimensional, geralmente obtido por técnicas como TF-IDF ou embeddings, e agrupar os textos de acordo com a similaridade de seus vetores. Essa técnica é especialmente útil em bases jurídicas, onde há grande volume de documentos e a necessidade de identificar padrões repetitivos ou agrupar casos similares.

No contexto jurídico, o agrupamento baseado em similaridade textual possibilita, por exemplo, a identificação de litigâncias predatórias, a organização automática de peças processuais semelhantes, a detecção de clusters de ações repetitivas e o apoio na gestão de

acervos documentais. Ao agrupar documentos de forma automatizada, facilita-se a análise estratégica, a triagem de processos e o desenvolvimento de políticas judiciais mais eficientes.

Essas técnicas de agrupamento, integradas a métodos de mineração de dados e processamento de linguagem natural, representam ferramentas valiosas para o aprimoramento da atividade judicial, especialmente em tempos de crescente judicialização e massificação de demandas.

2.9 Soluções Existentes para Análise Automatizada de Documentos Jurídicos

A aplicação de técnicas de Inteligência Artificial (IA) no Direito tem avançado de forma significativa, impulsionando o desenvolvimento de soluções comerciais e acadêmicas voltadas para a análise automatizada de documentos jurídicos. Diversas ferramentas de mercado ilustram esse movimento.

Entre as soluções destacam-se o Ross Intelligence, que utiliza processamento de linguagem natural para realizar pesquisas jurídicas e sugerir jurisprudência relevante, e o CaseText, que oferece sistemas de recomendação e análise preditiva de resultados judiciais. Outro exemplo é o IBM Watson Legal, que disponibiliza recursos de classificação de documentos, análise de contratos e suporte à tomada de decisões jurídicas (SURESH; GUTTAG, 2021).

Além das iniciativas comerciais, há um crescente número de projetos acadêmicos dedicados à utilização de IA no meio jurídico, com abordagens voltadas à extração de informações de decisões judiciais, detecção de padrões em demandas repetitivas e desenvolvimento de sistemas de apoio à decisão judicial (SURDEN, 2014).

Essas soluções têm demonstrado a capacidade da Inteligência Artificial em lidar com grandes volumes de dados jurídicos, proporcionando ganhos em eficiência, agilidade e qualidade na gestão da informação jurídica.

Apesar dos avanços na utilização da Inteligência Artificial no Direito, muitas das soluções existentes foram desenvolvidas para atender a ordenamentos jurídicos estrangeiros, especialmente aqueles baseados no sistema anglo-saxão (common law), cuja estrutura, práticas processuais e forma de produção das decisões judiciais diferem significativamente do modelo brasileiro, de tradição romano-germânica (civil law) (SANTOS, 2020; MACHADO, 2021).

Essa diferença cria uma lacuna relevante: poucas ferramentas de análise automatizada de documentos jurídicos estão adaptadas às especificidades linguísticas, procedimentais e estruturais do sistema de justiça brasileiro. Como reconhecido pelo Conselho Nacional de

Justiça (CNJ, 2020), a realidade nacional exige o desenvolvimento de soluções tecnológicas que considerem as peculiaridades locais e a organização normativa vigente.

Além disso, não há, até o momento, soluções amplamente consolidadas no mercado nacional que se dediquem especificamente à detecção automatizada de práticas como a litigância predatória em plataformas de tramitação processual eletrônico, como o eProc.

Nesse contexto, o Projeto Tanatose propõe uma abordagem inovadora e customizada, voltada diretamente para o combate à litigância predatória no sistema judicial eletrônico utilizado pelo Tribunal de Justiça do Estado do Tocantins. Empregando técnicas de mineração de dados, processamento de linguagem natural e análise de similaridade textual sobre documentos extraídos do eProc, o projeto busca preencher uma lacuna crítica, oferecendo suporte tecnológico especializado para a triagem automatizada de demandas abusivas e contribuindo para o fortalecimento da eficiência e da integridade da prestação jurisdicional no Brasil.

2.10 Tecnologias utilizadas no Projeto Tanatose

O Projeto Tanatose incorpora um arcabouço tecnológico multidisciplinar com o propósito de identificar automaticamente padrões de repetição em petições iniciais no sistema judicial eletrônico, colaborando no enfrentamento da litigância predatória no âmbito do Tribunal de Justiça do Estado do Tocantins (TJTO).

A solução articula técnicas de Inteligência Computacional, Mineração de Dados e Processamento de Linguagem Natural (PLN), integrando essas áreas para proporcionar uma análise automatizada eficiente e alinhada às necessidades do sistema judicial brasileiro.

A Inteligência Computacional (IC) é aplicada na modelagem de agentes de análise que operam com base em critérios complexos, simulando processos decisórios adaptativos e flexíveis (KRUSE et al., 2013). Esses agentes são responsáveis por conduzir a comparação textual entre documentos jurídicos, buscando padrões de similaridade que indiquem repetições predatórias.

A Mineração de Dados (MD), por sua vez, é utilizada para a exploração, organização e agrupamento de grandes volumes de processos judiciais, permitindo a identificação de padrões de comportamento litigioso que não são imediatamente perceptíveis (HAN; KAMBER; PEI,

2012). Com isso, é possível segmentar conjuntos de processos similares e subsidiar análises estratégicas para o combate à litigância abusiva.

Já o Processamento de Linguagem Natural (PLN) desempenha papel central no projeto, sendo responsável pela leitura automatizada, extração de informações relevantes e normalização textual das petições iniciais. Técnicas como tokenização, lematização, extração de entidades nomeadas e análise de similaridade textual são aplicadas para transformar os textos jurídicos em representações vetoriais adequadas à comparação automática (JURAFSKY; MARTIN, 2020).

A análise de similaridade textual no Projeto Tanatose é realizada utilizando o **modelo Bag of Words** combinado com a métrica da Similaridade do Cosseno, permitindo medir o grau de semelhança entre as petições com alta precisão (MANNING; RAGHAVAN; SCHÜTZE, 2008).

A integração dessas tecnologias possibilita a geração automatizada de relatórios de similaridade, entregando aos servidores do TJTO ferramentas objetivas para a triagem e identificação de demandas potencialmente abusivas, promovendo a eficiência, a transparência e a racionalização da atividade jurisdicional.

O Projeto Tanatose representa, portanto, uma iniciativa que se insere nas diretrizes contemporâneas de modernização do Poder Judiciário, alinhando-se aos princípios constitucionais da eficiência, da moralidade administrativa e do acesso à justiça, além de atender às recomendações de transformação digital promovidas pelo Conselho Nacional de Justiça (CNJ, 2020).

3 METODOLOGIA

Para auxiliar os servidores do Poder Judiciário do Estado do Tocantins na identificação de litigâncias predatórias, foi desenvolvida a ferramenta denominada *Tanatose*. Essa aplicação automatiza e agiliza significativamente a análise em massa de processos, proporcionando maior eficiência na detecção de padrões repetitivos.

A seguir, apresenta-se uma representação esquemática do funcionamento da ferramenta, destacando as quatro etapas principais do fluxo operacional: entrada dos números dos processos, extração e processamento de textos, análise de similaridade textual e visualização dos resultados.

Figura 1 - Fluxo de funcionamento da ferramenta Tanatose



Fonte: Elaborado pelo autor (2025)

3.1 Tipo e Abordagem da Pesquisa

Esta é uma pesquisa aplicada, uma vez que tem como objetivo desenvolver uma solução tecnológica voltada à resolução de um problema prático enfrentado pelo Poder Judiciário do Estado do Tocantins: a dificuldade na identificação de possíveis litigâncias predatórias em meio ao grande volume de processos judiciais eletrônicos. O conhecimento gerado será diretamente empregado no ambiente institucional, por meio da implementação de uma ferramenta computacional de apoio à análise processual.

A abordagem adotada é mista, combinando elementos **quantitativos e qualitativos**. Do ponto de vista quantitativo, a pesquisa baseia-se na análise de dados processuais extraídos do sistema eProc, utilizando métricas de similaridade textual aplicadas às petições iniciais e

técnicas de mineração de dados. A ferramenta desenvolvida gera indicadores objetivos, como percentuais de semelhança entre textos jurídicos, permitindo mensurar o grau de repetição entre os processos analisados. Já sob a perspectiva qualitativa, o estudo envolveu o **desenvolvimento, validação prática e avaliação da ferramenta Tanatose** no ambiente real do Tribunal de Justiça do Tocantins, considerando a percepção dos usuários sobre sua utilidade, desempenho e aplicabilidade no contexto institucional.

Além disso, trata-se de uma pesquisa de natureza exploratória, uma vez que busca compreender melhor o fenômeno das litigâncias predatórias dentro do contexto digital do Judiciário, ainda pouco explorado sob a perspectiva da inteligência computacional. Essa característica exploratória também se reflete na escolha por utilizar e adaptar métodos de processamento de linguagem natural aplicados ao domínio jurídico, o que exige investigação contínua sobre as melhores técnicas e abordagens a serem empregadas. Além disso, o método utilizado baseia-se no desenvolvimento incremental de sistemas, com etapas bem definidas, seguindo o ciclo de desenvolvimento incremental, desde a elicitação de requisitos até a validação experimental da aplicação Tanatose.

3.2 Análise de Requisitos

Durante a fase inicial do desenvolvimento da ferramenta Tanatose, foi realizada uma análise de requisitos com o objetivo de compreender as necessidades operacionais do Poder Judiciário no que diz respeito à identificação de litigâncias predatórias. Esta etapa consistiu na definição das funcionalidades essenciais do sistema, bem como garantir que a solução proposta fosse viável, eficiente e aderente ao contexto de uso institucional.

3.2.1 Necessidade do Usuário

Foi identificado que servidores do Poder Judiciário enfrentam dificuldades na detecção de litigâncias predatórias, especialmente em cenários com grande volume de processos. A triagem manual consome tempo e recursos, além de estar sujeita a falhas humanas.

3.2.2 Objetivo do Sistema

O sistema deve permitir a análise automatizada de petições iniciais, com foco na identificação de padrões repetitivos entre processos, oferecendo agilidade e apoio à tomada de decisão por parte dos servidores responsáveis.

3.2.3 Requisitos Funcionais Esperados

Com base na demanda apresentada, definiu-se que o sistema deveria:

- Permitir o envio de listas de processos previamente identificados para análise;
- Realizar a extração automática do conteúdo textual das petições iniciais;
- Aplicar técnicas de comparação textual para mensurar a similaridade entre processos;
- Apresentar os resultados de forma organizada e acessível, por meio de relatórios.

3.2.4 Requisitos Não Funcionais Esperados

Além das funcionalidades esperadas, o sistema deve respeitar os seguintes critérios não funcionais:

- **Eficiência:** Capacidade de processar grandes volumes de processos com desempenho aceitável;
- **Usabilidade:** Interface intuitiva e amigável, voltada a usuários não técnicos;
- **Segurança:** Garantia da confidencialidade dos dados judiciais manuseados;
- **Manutenibilidade:** Código organizado e modular, facilitando futuras melhorias;
- **Compatibilidade:** Interoperabilidade com os formatos de documentos utilizados no sistema eProc.

Foi identificado que servidores do Poder Judiciário enfrentam dificuldades na detecção de litigâncias predatórias, especialmente em cenários com grande volume de processos. A triagem manual consome tempo e recursos, além de estar sujeita a falhas humanas.

3.2.5 Restrições Identificadas

Algumas restrições técnicas e operacionais também foram observadas, tais como:

- A dependência do acesso aos documentos processuais por meio do sistema eProc;

- A necessidade de lidar com arquivos de diferentes tamanhos, formatos e estruturas textuais;
- A obrigação de respeitar as diretrizes institucionais de segurança da informação.

Essa análise inicial serviu de base para a modelagem da aplicação, orientando todas as etapas subsequentes do desenvolvimento do sistema Tanatose. Nas próximas seções, serão detalhadas cada uma das etapas desse processo, abrangendo a construção da ferramenta, as soluções tecnológicas adotadas e os métodos empregados para garantir a precisão na identificação de litigâncias predatórias.

3.3 Estratégia de Desenvolvimento da Ferramenta

A concepção da ferramenta Tanatose foi orientada por uma análise prática do fluxo de trabalho dos servidores do CINUGEP (Centro de Inteligência do Núcleo de Gerenciamento de Precedentes) do Tribunal de Justiça do Estado do Tocantins. A partir de conversas informais e observações de rotina, identificou-se que uma das tarefas mais onerosas em termos de tempo e esforço era a leitura manual das petições iniciais, documento que contém a causa de pedir e serve como ponto de partida na identificação de padrões repetitivos entre processos judiciais.

Ficou evidente que a principal necessidade do setor era a detecção de demandas com elevado grau de similaridade textual. Esse trabalho, quando realizado manualmente, limita-se à análise por amostragem, pois exigia grande dedicação da equipe — composta por apenas dois servidores. Diante desse cenário, verificou-se a viabilidade de desenvolver uma ferramenta capaz de automatizar a etapa de comparação textual, liberando os servidores para atuar diretamente sobre os casos mais relevantes identificados pela triagem automatizada.

A proposta metodológica consistiu, então, no desenvolvimento de uma aplicação computacional capaz de realizar, de forma automatizada, a extração, o tratamento e a comparação do conteúdo textual das petições iniciais. A seleção dos processos a serem auditados continuou a ser feita por ferramentas de Business Intelligence (BI) do próprio Tribunal, de modo que a ferramenta Tanatose foi projetada para atuar exclusivamente na etapa de análise textual automatizada.

Para viabilizar a obtenção automatizada das petições, adotou-se como solução o Modelo Nacional de Interoperabilidade (MNI) — padrão regulamentado pelo Conselho Nacional de Justiça (CNJ), que estabelece diretrizes de interoperabilidade entre sistemas judiciais eletrônicos. Isso se mostrou compatível com o sistema eProc, utilizado pelo Tribunal de Justiça do Tocantins, cuja base legal é a Lei Federal nº 11.419/2006, que dispõe sobre a informatização

do processo judicial no Brasil (BRASIL, 2006; CNJ, 2023). Através do MNI, foi possível estabelecer conexão segura com o sistema eProc e realizar o download automatizado das petições iniciais em lote.

A etapa seguinte consistiu na extração do conteúdo textual dos arquivos por meio de técnicas de Processamento de Linguagem Natural (PLN) e Mineração de Dados. Quando necessário, foi utilizado OCR (Optical Character Recognition) para transformar PDFs baseados em imagem em texto processável. Todo o conteúdo foi armazenado em banco de dados próprio, com estrutura adequada ao processo de comparação cruzada.

Para a análise de similaridade textual, foi adotada uma abordagem baseada em vetores de frequência de palavras e cálculo do ângulo vetorial entre documentos, complementada com verificação de sinônimos por meio da biblioteca NLTK. Essa escolha também responde à necessidade de transparência e explicabilidade em soluções adotadas pelo setor público, permitindo que os resultados sejam compreendidos e auditados por usuários não especializados em ciência de dados.

Os resultados da análise foram organizados em relatórios estruturados por faixas de similaridade, os quais passaram a servir como apoio direto à auditoria dos servidores do TJTO. Assim, a ferramenta não substitui a análise jurídica, mas atua como um sistema de triagem inteligente, aumentando a eficiência institucional, reduzindo o retrabalho e promovendo a uniformidade no tratamento de litígios repetitivos.

3.3.1 Requisitos do Sistema

A ferramenta desenvolvida, denominada Tanatose, tem como objetivo principal auxiliar os servidores do Poder Judiciário na identificação de litigâncias predatórias, promovendo maior eficiência na análise em massa de processos judiciais. Os requisitos do sistema foram divididos em funcionais (o que o sistema deve fazer) e não funcionais (como o sistema deve se comportar), conforme as boas práticas de engenharia de software. A seguir, apresentam-se os requisitos definidos para garantir o alinhamento entre a solução técnica e as necessidades institucionais.

3.3.1.1 Requisitos Funcionais

RF1: Permitir a importação de uma lista de processos previamente selecionados, contendo indícios de anomalias identificados na triagem inicial;

RF2: Realizar automaticamente a extração das petições iniciais dos processos informados na lista;

RF3: Armazenar o conteúdo textual das petições extraídas em banco de dados próprio para análise posterior;

RF4: Aplicar técnicas de Processamento de Linguagem Natural (PLN) para realizar a comparação textual entre as petições iniciais;

RF5: Calcular e classificar os graus de similaridade entre os textos processados;

RF6: Agrupar os processos conforme o percentual de similaridade textual;

RF7: Gerar relatórios organizados com os resultados da análise, separados por faixas de similaridade;

RF8: Disponibilizar uma interface de visualização que permita ao usuário consultar os processos analisados e seus respectivos agrupamentos;

RF9: Executar de forma automatizada todas as etapas do fluxo de trabalho, permitindo o processamento em lote.

3.3.1.2 Requisitos Não Funcionais

RNF1: Garantir desempenho eficiente, com capacidade de processar grandes volumes de processos em tempo adequado;

RNF2: Oferecer uma interface gráfica amigável e intuitiva, voltada a usuários sem conhecimento técnico especializado;

RNF3: Assegurar a confidencialidade dos dados processuais, respeitando os níveis de sigilo estabelecidos pelo sistema eProc;

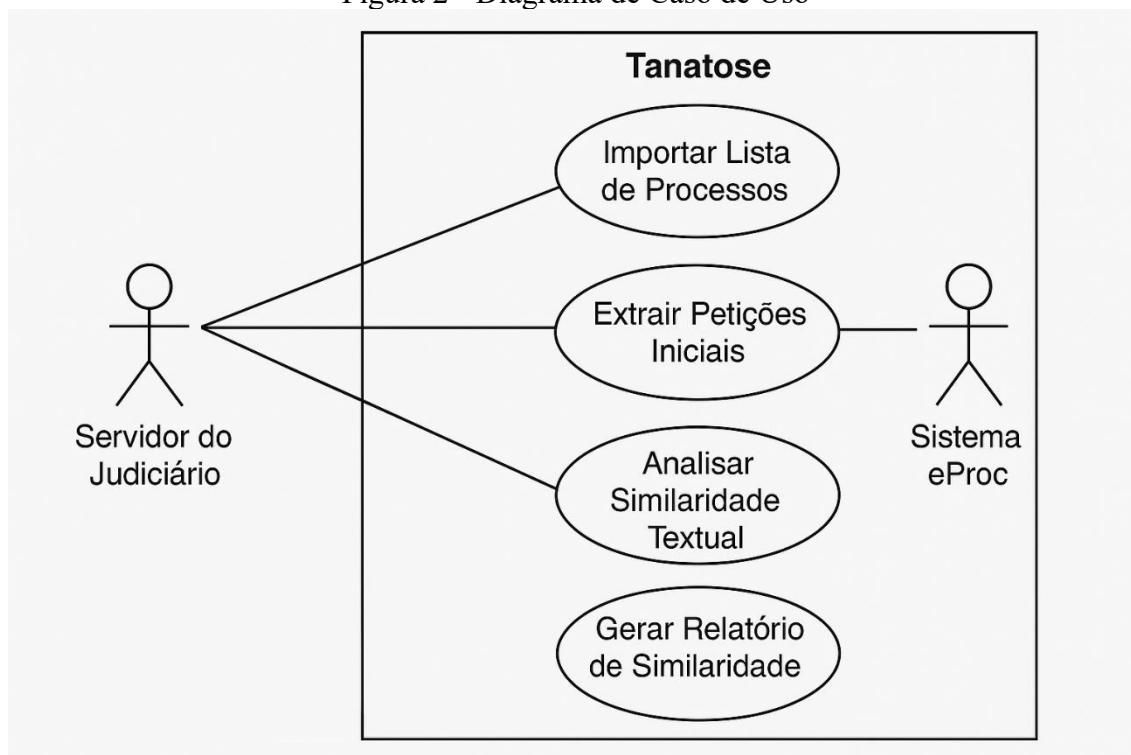
RNF4: Ser compatível com os formatos de petições geradas pelo eProc, mesmo na ausência de recursos de OCR;

RNF5: Possuir portabilidade, permitindo a execução em diferentes ambientes computacionais, sem necessidade de infraestrutura dedicada;

RNF6: Adotar uma arquitetura modular, facilitando a manutenção e futuras atualizações do sistema;

3.3.2 Diagramas

Figura 2 - Diagrama de Caso de Uso



Fonte: Elaborado pelo autor (2025)

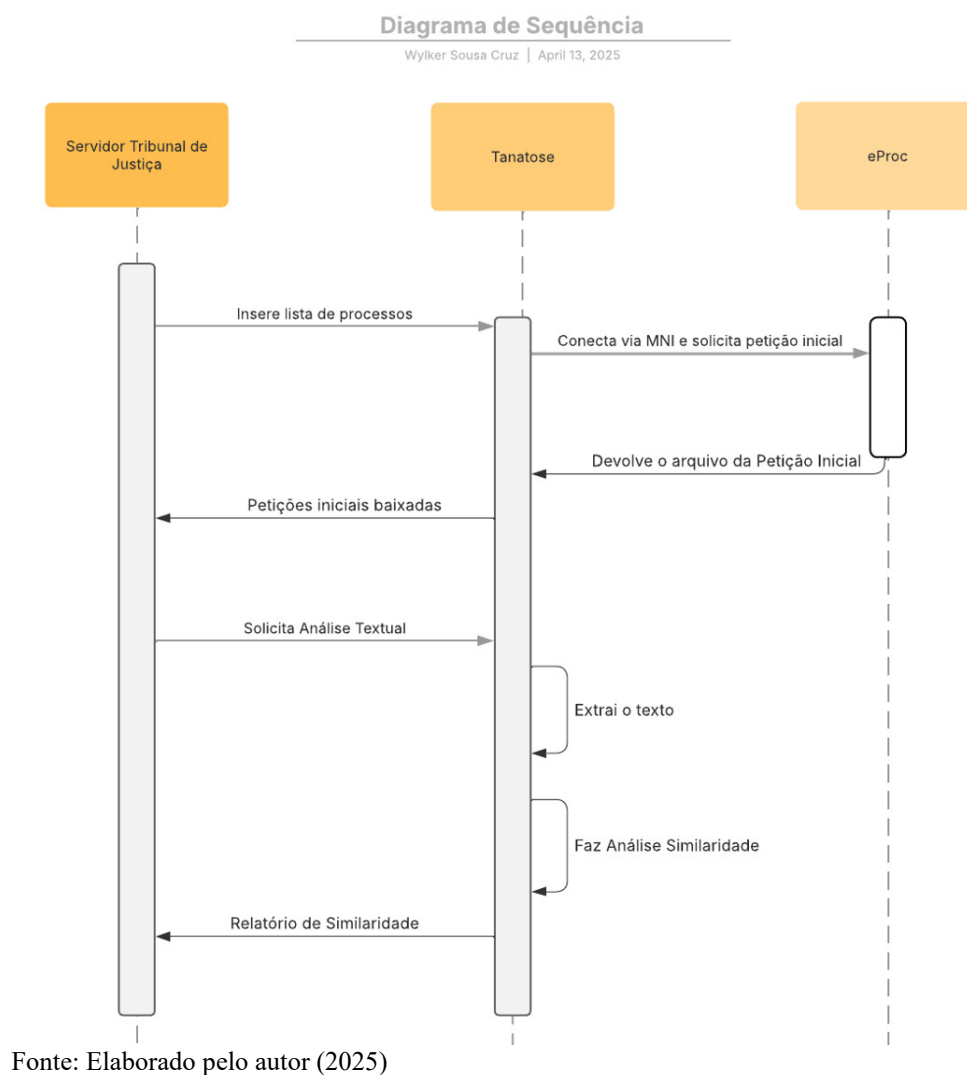
O diagrama de caso de uso da ferramenta Tanatose, demonstrado na Figura 1, apresenta a interação entre os principais atores e as funcionalidades do sistema. Os dois atores identificados são: o Servidor do Judiciário, que representa o usuário responsável por operar a aplicação, e o Sistema eProc, que atua como fonte de dados para a extração das petições iniciais dos processos.

Tanatose foi conduzido em fases distintas, cada uma com um papel fundamental no processo. Essas etapas foram organizadas de maneira sequencial, de modo que cada fase depende diretamente da conclusão da anterior.

- **Importar Lista de Processos:** Permite que o usuário insira uma lista de processos previamente selecionados para análise.
- **Baixar Petições Iniciais:** O sistema acessa o eProc para coletar automaticamente as petições iniciais de cada processo da lista.
- **Analisar Similaridade Textual:** Após a extração, a aplicação realiza uma análise de similaridade entre os textos das petições, identificando padrões repetitivos.
- **Gerar Relatório de Similaridade:** Por fim, o sistema organiza os resultados da análise em um relatório que pode ser visualizado e exportado pelo usuário.

O Sistema eProc aparece como ator auxiliar no caso de uso “Extrair Petições Iniciais”, pois fornece os documentos necessários para a análise. O diagrama reforça a automatização e a simplicidade do fluxo de trabalho, destacando o papel do servidor como operador e do sistema Tanatose como ferramenta de apoio à detecção de litigâncias predatórias.

Figura 3 - Diagrama de Sequência



O Diagrama de Sequência da ferramenta Tanatose, no qual estão representadas as principais interações entre os atores envolvidos no processo de análise automatizada de litigâncias predatórias. O fluxo se inicia com a ação do servidor do Tribunal de Justiça, que realiza a importação da lista de processos a serem analisados.

A partir dessa ação, o sistema Tanatose estabelece conexão com o sistema eProc, por meio do Modelo Nacional de Interoperabilidade (MNI), e realiza o download automatizado das petições iniciais dos processos informados. Após a conclusão do download, o sistema comunica ao servidor que os documentos foram baixados com sucesso.

Em seguida, o servidor solicita ao sistema que inicie a análise de similaridade textual. O Tanatose, então, executa as seguintes etapas de forma automatizada: extração do conteúdo textual das petições iniciais, aplicação de algoritmos de comparação e, por fim, geração do relatório com os percentuais de similaridade entre os processos analisados. Este relatório é devolvido ao servidor, que poderá utilizá-lo como base para aprofundar a auditoria processual de forma mais direcionada e eficiente.

3.3.3 Etapas de Funcionamento da Ferramenta

O funcionamento da ferramenta Tanatose está dividido em quatro etapas principais: baixa das petições iniciais, processamento textual, análise de similaridade e geração de relatórios. A seguir, cada uma dessas etapas é detalhada.

Baixa das Petições Iniciais

Com o intuito de identificar litigâncias predatórias, adotou-se como estratégia a comparação das petições iniciais dos processos, tendo em vista que, nesse tipo de demanda, tais documentos tendem a apresentar elevado grau de similaridade. O primeiro passo do processo consiste na extração automática dessas petições diretamente do sistema processual eProc.

A aplicação recebe como entrada uma lista contendo os números dos processos a serem analisados, sem limitação quanto ao seu tamanho. Para realizar a extração das petições, é utilizado o Modelo Nacional de Interoperabilidade (MNI), padrão instituído pelo Conselho Nacional de Justiça (CNJ), que define diretrizes para a integração entre os sistemas do Poder Judiciário. Por meio do MNI, é possível obter os dados completos do processo, incluindo peças processuais, movimentações e documentos anexos, desde que respeitadas as permissões de acesso.

Processamento Textual

Nesta etapa, são extraídos os dados das petições iniciais obtidas anteriormente, com o objetivo de prepará-los para análise. O processo de extração e tratamento textual é realizado com o uso de técnicas de Mineração de Dados (Data Mining) e Processamento de Linguagem Natural (PLN), que permitem estruturar as informações em um formato adequado para a comparação textual subsequente.

Nos casos em que as petições estavam em formato de imagem ou apresentavam texto não selecionável, foi utilizado o Reconhecimento Óptico de Caracteres (OCR) para extrair o conteúdo textual. A ferramenta Tanatose aplica o mecanismo **Tesseract OCR**, por meio da biblioteca `pytesseract`, amplamente reconhecida por sua acurácia e código aberto. As páginas em PDF são previamente convertidas em imagens no formato JPEG utilizando a biblioteca `pdf2image`, com resolução de 300 dpi, assegurando boa qualidade para o reconhecimento óptico. Essa abordagem permitiu o tratamento adequado de documentos digitalizados, ampliando a cobertura da análise textual automatizada e garantindo a integridade das informações submetidas à etapa seguinte de similaridade.

Análise de Similaridade

A análise de similaridade consiste na comparação dos textos extraídos na etapa anterior, agrupando os processos conforme o grau de semelhança identificado entre as petições. Essa comparação é realizada por meio de algoritmos que quantificam o percentual de similaridade entre os documentos, permitindo a organização dos processos em grupos coesos e com padrões repetitivos.

Geração de Relatórios

Por fim, a ferramenta gera relatórios com os resultados da análise, organizando os processos por faixas de similaridade, a saber:

- Inferior a 20%
- Entre 20% e 40%
- Entre 40% e 60%
- Entre 60% e 80%
- Superior a 80%

Esses relatórios são apresentados de forma acessível e estruturada, oferecendo aos servidores uma visão clara dos agrupamentos, o que facilita a identificação de possíveis litigâncias predatórias e contribui para a tomada de decisões estratégicas.

3.3.4 Projeto Lógico

O projeto lógico do banco de dados da ferramenta Tanatose foi concebido com o objetivo de organizar e estruturar eficientemente as informações utilizadas e produzidas durante a análise de litigâncias predatórias. Para atender às necessidades do sistema, foi criado um esquema denominado `processos_predatorios`, contendo três tabelas que representam,

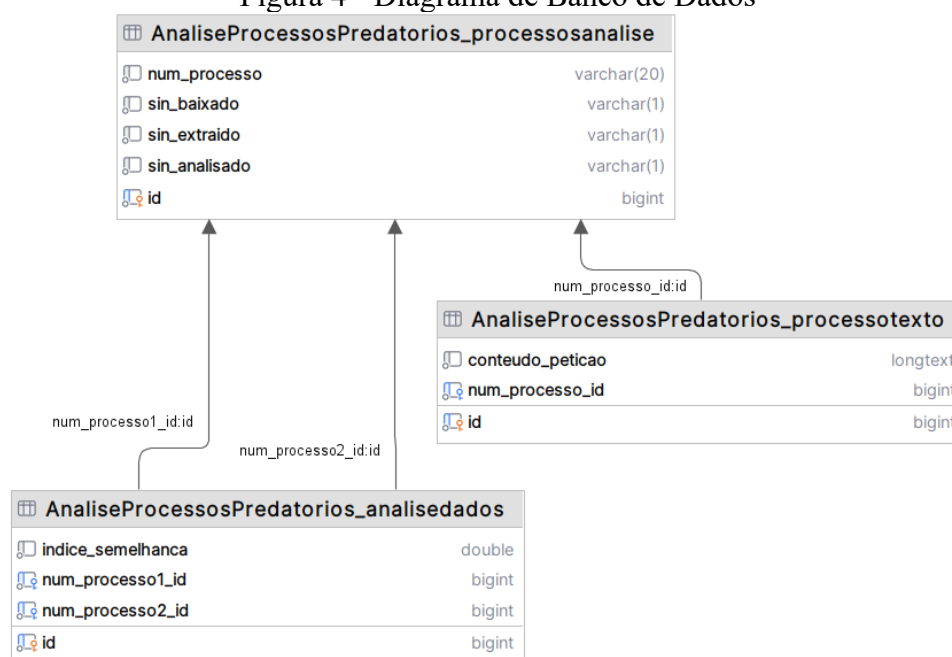
respectivamente, os processos analisados, os textos extraídos das petições iniciais e os resultados das análises de similaridade.

Tecnologias utilizadas:

- Banco de Dados Relacional (ex: MySQL).

Tabelas otimizadas para consultas de agrupamento e similaridade textual.

Figura 4 - Diagrama de Banco de Dados



Fonte: o autor Fonte: Elaborado pelo autor (2025)

3.3.5 Modelo de Desenvolvimento Adotado

O desenvolvimento da ferramenta Tanatose foi conduzido com base nos princípios da Engenharia de Software definidos por Pressman (2016), especialmente no que se refere ao modelo de processo incremental. Esse modelo pressupõe que o sistema seja construído e entregue em partes funcionais (incrementos), com cada versão incorporando funcionalidades adicionais, refinando a solução progressivamente.

Segundo Pressman, o modelo incremental é particularmente adequado para contextos onde os requisitos podem ser parcialmente definidos no início, sendo aperfeiçoados ao longo do tempo com base na validação prática do sistema. Esse modelo permite flexibilidade e adaptação às necessidades do usuário, favorecendo a entrega contínua de valor.

Durante o projeto da ferramenta Tanatose, essa abordagem se mostrou apropriada, uma vez que o sistema foi desenvolvido de forma progressiva, com testes e validações parciais em ambiente real, permitindo a coleta de feedback da equipe de auditoria da CINUGEP e a realização de ajustes contínuos nas funcionalidades implementadas.

A adoção do modelo incremental contribuiu diretamente para a robustez do sistema, pois permitiu que cada componente fosse testado e validado separadamente, reduzindo o risco de falhas e otimizando o uso dos recursos disponíveis. Essa estratégia também favoreceu o alinhamento entre os objetivos técnicos da pesquisa e as demandas operacionais da instituição usuária da aplicação.

3.4 Tecnologias, Ferramentas e Ambiente de Desenvolvimento

Para o desenvolvimento da ferramenta Tanatose, foram utilizadas tecnologias e ferramentas modernas, amplamente consolidadas no mercado de desenvolvimento de software, visando garantir robustez, escalabilidade e facilidade de manutenção da aplicação.

A aplicação foi implementada utilizando a linguagem de programação Python, escolhida por sua versatilidade, grande variedade de bibliotecas voltadas ao processamento de linguagem natural e forte comunidade de apoio. Para estruturação do sistema web e organização da lógica de negócio, foi adotado o framework Django, que segue o padrão de arquitetura Model-View-Template (MVT) e proporciona recursos nativos para segurança, roteamento e integração com banco de dados.

Como sistema gerenciador de banco de dados relacional, foi utilizado o MySQL, devido à sua confiabilidade, desempenho e compatibilidade com a estrutura do projeto, além de sua ampla aceitação nos ambientes corporativo, governamental e acadêmico.

O ambiente de desenvolvimento utilizado para a construção da aplicação consistiu em um computador pessoal com as seguintes especificações técnicas:

Processador: AMD Ryzen 7

Memória RAM: 32 GB

Armazenamento: SSD de 1 TB

Placa de vídeo: NVIDIA GeForce RTX 3060

Sistema operacional: Microsoft Windows 11 Pro

Esses recursos computacionais foram fundamentais para garantir o desempenho durante a fase de desenvolvimento, especialmente nas etapas de análise textual em lote, que exigem significativa capacidade de processamento e manipulação de dados em memória.

A escolha dessas tecnologias visou assegurar a compatibilidade com os padrões utilizados no Poder Judiciário, bem como facilitar futuras integrações com sistemas já existentes, como o eProc. Essa escolha também favorece a replicabilidade da solução em outros tribunais que utilizam o sistema eProc, além de facilitar a manutenção por equipes técnicas institucionais, uma vez que são ferramentas amplamente conhecidas.

3.5 Validação da Ferramenta

A validação da ferramenta Tanatose foi realizada por meio da aplicação prática no ambiente institucional do Tribunal de Justiça do Estado do Tocantins, com base em relatos e observações dos servidores que atuam na auditoria de processos no âmbito da CINUGEP. Após a implantação da ferramenta de maneira experimental, os usuários passaram a utilizá-la em suas rotinas de trabalho e relataram percepções sobre sua utilidade, desempenho e impacto na identificação de litigâncias predatórias.

Essa validação foi de natureza qualitativa e ocorreu de forma contínua, a partir do uso real da ferramenta. Os relatos indicaram redução significativa do tempo de triagem, maior clareza na organização dos processos por grau de similaridade e percepção de que a ferramenta aumenta a confiabilidade das análises manuais subsequentes, ao direcionar a atenção dos servidores para os casos com maior potencial de repetição.

Embora não tenha sido aplicada uma métrica quantitativa formal, os usuários expressaram confiança nos agrupamentos apresentados, destacando que os relatórios gerados pela ferramenta demonstraram elevada coerência com os padrões identificados empiricamente por meio da leitura manual das petições. Essa percepção positiva reforça o valor prático da solução proposta, especialmente em contextos de grande volume processual e equipe reduzida.

Como proposta para trabalhos futuros, sugere-se a implementação de mecanismos formais de avaliação, incluindo a comparação entre os resultados gerados pela ferramenta e auditorias manuais controladas, além da coleta sistematizada de métricas como tempo de análise, taxa de confirmação e índice de falseamento dos agrupamentos.

Esse tipo de validação, ainda que qualitativa, é adequado na etapa inicial de implementação de sistemas computacionais no setor público, sobretudo quando associados a mudanças de fluxo de trabalho.

3.6 Limitações e Considerações Éticas

O desenvolvimento desta pesquisa respeitou os princípios legais, institucionais e éticos aplicáveis ao manuseio de dados sensíveis no âmbito do Poder Judiciário. Para a obtenção dos dados necessários à implementação e validação da ferramenta Tanatose, foi protocolado um pedido formal por meio de processo administrativo no Sistema Eletrônico de Informações (SEI), utilizado pelo Tribunal de Justiça do Estado do Tocantins. O referido processo tramita sob o número SEI nº 23.0.000036039-2, e sua autorização foi concedida mediante o cumprimento rigoroso das normas de sigilo processual e dos dispositivos da Lei nº 13.709, de 14 de agosto de 2018, conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD).

O escopo deste trabalho abrange exclusivamente os processos previamente auditados pela CINUGEP (Centro de Inteligência do Núcleo de Gerenciamento de Precedentes), não havendo qualquer tipo de acesso, análise ou divulgação de dados que extrapolem os limites expressamente autorizados pela Administração do Tribunal. Todos os dados utilizados encontram-se sob salvaguarda institucional e foram tratados de forma anonimizada, sempre que possível, com o objetivo de garantir a preservação da identidade das partes envolvidas e o respeito aos princípios da finalidade, necessidade e minimização do tratamento de dados.

De acordo com o art. 6º da LGPD, o tratamento de dados pessoais deve observar os princípios da boa-fé, da finalidade e da necessidade, entre outros. Conforme estabelece a legislação:

“As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: [...] I – finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades; [...] III – necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento” (BRASIL, 2018).

Adicionalmente, os documentos processuais utilizados possuem diferentes níveis de sigilo no sistema eProc. Assim, foram observadas as devidas permissões de acesso e os parâmetros de segurança da informação previstos no sistema, de modo a evitar qualquer violação ética ou legal durante as etapas de desenvolvimento e execução da aplicação.

Por fim, ressalta-se que este trabalho tem caráter técnico-científico, sem qualquer intenção de julgamento sobre o mérito das demandas judiciais analisadas. Seu objetivo principal é o desenvolvimento de uma ferramenta de apoio à identificação de padrões processuais, com

foco na eficiência administrativa e na melhoria da prestação jurisdicional. Por essa razão, o sistema foi concebido como um instrumento de apoio à triagem técnica, sem a pretensão de substituir a análise interpretativa realizada por profissionais jurídicos.

4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos a partir da implementação da ferramenta Tanatose, desenvolvida como solução tecnológica para ajudar a enfrentar o problema da litigância predatória no sistema judicial eletrônico eProc, utilizado pelo Tribunal de Justiça do Tocantins. A ferramenta tem como principal objetivo automatizar a análise de grandes volumes de processos judiciais, com foco na identificação de demandas potencialmente repetitivas ou predatórias. Essa identificação se dá por meio da comparação e análise de similaridade textual entre as petições iniciais.

Inicialmente, será descrito o funcionamento da ferramenta Tanatose, detalhando sua arquitetura, os fluxos de operação e os principais algoritmos utilizados no processo de agrupamento e classificação das petições. Em seguida, serão apresentados os resultados extraídos a partir da aplicação prática da ferramenta sobre um conjunto real de processos judiciais, com destaque para métricas de desempenho como tempo de processamento, taxa de agrupamentos similares identificados e indicadores de acurácia da análise.

A apresentação dos dados será acompanhada de uma discussão crítica e interpretativa, que visa contextualizar os achados à luz das demandas operacionais do Tribunal de Justiça do Tocantins e da percepção dos servidores envolvidos no processo de validação. Cada etapa do fluxo operacional da aplicação — desde a baixa de petições iniciais até a geração dos relatórios — será discutida individualmente com base nos dados coletados e nas observações práticas.

4.1 Visão Geral da Ferramenta Tanatose

O nome da ferramenta desenvolvida, Tanatose, faz alusão ao termo utilizado na biologia para descrever o fenômeno da imobilidade tônica, também conhecido como "morte aparente", um comportamento adotado por determinados animais como estratégia de defesa contra predadores. Esse mecanismo consiste na simulação da morte, fazendo com que a presa permaneça imóvel, desestimulando o ataque de predadores, que frequentemente ignoram alvos que aparentam estar mortos. Tal resposta comportamental é observada em diversas espécies, como gambás, serpentes e insetos, e é considerada uma adaptação evolutiva que aumenta as chances de sobrevivência diante de ameaças (KALID, 2025).

Essa metáfora foi escolhida por refletir a proposta central do sistema: detectar padrões predatórios ocultos no volume processual do sistema eProc, interrompendo a inércia

institucional frente à litigância predatória, de modo análogo ao rompimento da passividade do organismo promovido pela tanatose como estratégia de sobrevivência.

A ferramenta Tanatose foi criada com o objetivo de automatizar o processo de identificação de demandas repetitivas ou predatórias no âmbito do sistema judicial eletrônico eProc.

O sistema está acessível no endereço interno <https://tanatose.tjto.jus.br>, com acesso restrito à rede do Tribunal de Justiça do Estado do Tocantins, por motivos de segurança institucional.

A interface da ferramenta foi desenvolvida com foco na usabilidade, adotando uma estrutura visual limpa, objetiva e funcional. O sistema guia o usuário por meio de um menu superior, onde as funcionalidades estão organizadas de maneira lógica e sequencial, conforme descrito a seguir:

a) Baixa de Processos: etapa inicial responsável pela importação dos números dos processos judiciais que serão analisados. A ferramenta realiza a busca automatizada das petições iniciais diretamente no sistema eProc, utilizando integração via Módulo Nacional de Interoperabilidade (MNI);

b) Extração de Textos: funcionalidade que executa o processamento das petições obtidas, realizando a extração textual dos documentos em formato PDF, transformando-os em dados legíveis e estruturados para análise;

c) Análise de Dados: módulo central da ferramenta, onde ocorre a aplicação dos algoritmos de similaridade textual e agrupamento de demandas. É nesta etapa que se realiza a identificação de possíveis padrões repetitivos entre as petições iniciais;

d) Visualização de Resultados: disponibiliza ao usuário, de forma clara e interativa, os agrupamentos de processos identificados com alto grau de similaridade. Os resultados são apresentados em relatórios que possibilitam a análise detalhada de cada conjunto processual.

4.2 Fluxo de Funcionamento da Ferramenta

O funcionamento do sistema ocorre em quatro etapas sequenciais:

a) Etapa 1: Baixa de Processos

O usuário insere, por upload, um arquivo do tipo .txt contendo os números dos processos que deseja analisar. O sistema lê o arquivo e registra os processos em uma tabela interna. Em seguida, realiza automaticamente a conexão com o sistema eProc por meio do Modelo Nacional de Interoperabilidade (MNI), e baixa a petição inicial de cada processo.

b) Etapa 2: Extração de Textos

Com os arquivos em PDF baixados, a ferramenta executa a extração do conteúdo textual. Caso o arquivo contenha texto não selecionável (imagem digitalizada), aplica-se OCR (Reconhecimento Óptico de Caracteres) para converter imagem em texto. Todo o texto extraído é salvo no banco de dados.

c) Etapa 3: Análise de Similaridade Textual

A partir dos textos armazenados, o sistema realiza comparações cruzadas entre todos os pares de processos, utilizando um algoritmo de similaridade baseado em frequência de palavras e ângulo vetorial. O índice de semelhança entre cada par é calculado e armazenado.

d) Etapa 4: Geração e Organização dos Resultados

Os resultados da análise são disponibilizados por meio de três formatos principais:

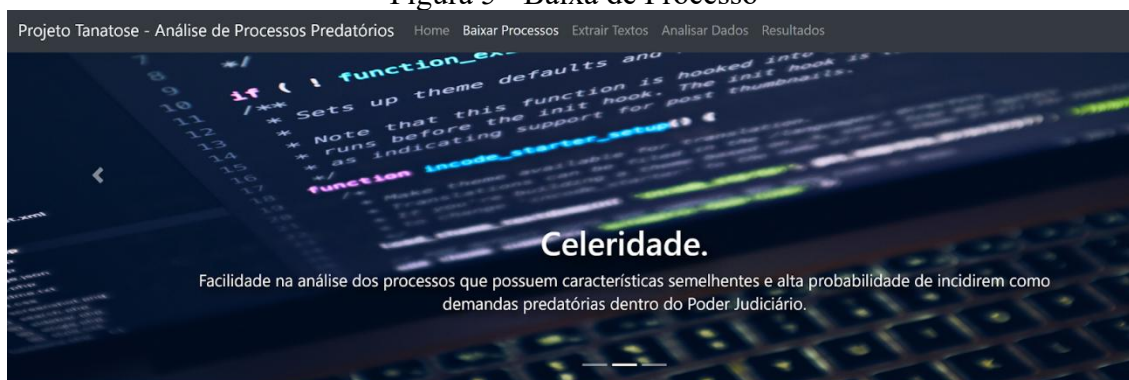
- Ranking dos pares mais e menos semelhantes;
- Gráfico com distribuição dos processos por faixas de similaridade;
- Filtro customizado: permite que o usuário selecione um intervalo de similaridade e visualize os pares de processos correspondentes.

A ferramenta Tanatose foi projetada para operar de maneira automatizada e sequencial, com o objetivo de reduzir o esforço humano na triagem de petições iniciais e identificar padrões de repetição textual que possam indicar litigância predatória. Seu funcionamento está estruturado em quatro etapas, que compreendem da coleta inicial dos processos até a apresentação final dos resultados.

4.2.1 Etapa 1: Baixa de Processos

A primeira etapa consiste na inserção de um arquivo no formato .txt, contendo os números dos processos judiciais que serão submetidos à análise. O sistema realiza a leitura desse arquivo e registra os dados em uma tabela interna. A partir dessas informações, estabelece conexão com o sistema eProc, por meio do Modelo Nacional de Interoperabilidade (MNI), e realiza automaticamente o download das petições iniciais correspondentes, em formato PDF.

Figura 5 - Baixa de Processo



Insira um arquivo com a lista dos processos que serão analisados

Fonte: <https://tanatose.tjto.jus.br/download/> (2025)

4.2.2 Etapa 2: Extração de Textos

Uma vez concluído o processo de obtenção dos documentos, a ferramenta inicia a etapa de extração textual. Os arquivos PDF são processados, e o conteúdo textual das petições é extraído. Caso os documentos apresentem texto não pesquisável, em razão de digitalizações ou imagens, é aplicado o recurso de Reconhecimento Óptico de Caracteres (OCR), que converte as imagens em texto. Todos os dados extraídos são armazenados em banco de dados.

Figura 6 - Extração de Texto



Extrair os Dados das Petições Baixadas

Ao total foram baixadas 4 petições iniciais.

Todos os processos cadastrados no banco de dados já foram baixados para o servidor e estão prontos para extração.

3 of 4 processed. Extraindo Processo 000 2706

Fonte: <https://tanatose.tjto.jus.br/extrair/> (2025)

4.2.3 Etapa 3: Análise de Similaridade Textual

Com os textos das petições iniciais devidamente extraídos e armazenados, o sistema Tanatose inicia a etapa de análise de similaridade textual, núcleo central da proposta para identificação de litígios predatórios. Esta fase consiste na comparação cruzada entre todos os pares de processos, com base em técnicas consolidadas de mineração de textos e representação vetorial.

A técnica de comparação adotada fundamenta-se na Similaridade do Cosseno aplicada sobre vetores de frequência de palavras, conforme o modelo conhecido como Bag of Words (BoW). Inicialmente, cada petição é submetida a um processo de normalização textual, que envolve a remoção de pontuações, conversão para letras minúsculas, tokenização (divisão do texto em palavras), eliminação de palavras irrelevantes (stopwords) da língua portuguesa — utilizando a biblioteca NLTK (BIRD; LOPER, 2002) — e lematização com o auxílio da biblioteca spaCy (HONNIBAL et al., 2020), a fim de reduzir palavras flexionadas à sua forma canônica.

Com os textos normalizados, são construídos vetores de frequência, nos quais cada elemento representa o número de ocorrências de um termo específico no documento. Assim, cada texto é representado por um dicionário de palavras e suas frequências relativas, segundo a abordagem Bag of Words, amplamente reconhecida em tarefas de análise textual (AGGARWAL, 2018).

A Similaridade do Cosseno é então calculada para cada par de vetores de frequência, utilizando a seguinte equação:

$$\text{Similaridade}(A, B) = \cos(\theta) = (A \cdot B) / (\|A\| \times \|B\|)$$

Onde:

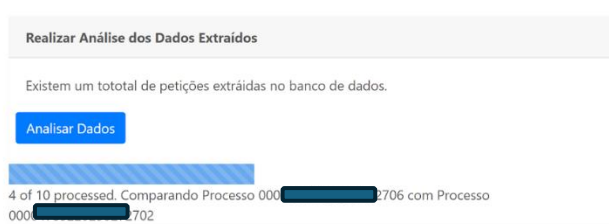
- **A e B** são os vetores de frequência dos textos comparados;
- **A · B** representa o produto escalar entre os vetores;
- **||A|| e ||B||** são as normas (comprimentos) de cada vetor.

Esse método permite comparar textos de diferentes tamanhos, considerando apenas a proporção relativa dos termos compartilhados, o que o torna especialmente útil para detectar similaridades em petições padronizadas (MANNING; RAGHAVAN; SCHÜTZE, 2008).

Os resultados da análise de similaridade são registrados em estrutura própria no banco de dados, possibilitando consultas, cruzamentos e filtrações posteriores. Essa abordagem técnica, aliada à automação do processo, proporciona uma análise textual robusta, escalável e alinhada às necessidades práticas dos servidores da Justiça, permitindo identificar rapidamente

petições com alto grau de semelhança — fator crítico para a detecção de demandas repetitivas ou litigância predatória no sistema eProc.

Figura 7 - Análise de Dados



Fonte: <https://tanatose.tjto.jus.br/analise/> (2025)

4.2.4 Etapa 4: Geração e Organização dos Resultados

A última etapa compreende a apresentação dos resultados da análise, de forma clara e interativa. O sistema disponibiliza três formas principais de visualização:

a) Ranking dos pares mais e menos semelhantes: lista ordenada dos pares de processos conforme seus respectivos índices de similaridade;

Figura 8 - Resultados - ranking de pares



Fonte: <https://tanatose.tjto.jus.br/resultados/> (2025)

b) Gráfico de distribuição: representação gráfica da frequência de pares de processos por faixas de similaridade (acima de 80%, entre 60% e 80%, entre 40% e 20% e menos de 20%);

Para a utilização desses dados no contexto do mestrado profissional, foi protocolado um pedido formal junto à Presidência do TJTO, por meio do Sistema Eletrônico de Informações (SEI), solicitando autorização específica para uso acadêmico. O pleito foi devidamente analisado e **autorizado institucionalmente**, respeitando os princípios da legalidade, da proteção de dados e da finalidade científica da pesquisa.

A execução do estudo de caso foi realizada por mim, pesquisador responsável, diretamente no ambiente de produção do sistema Tanatose, com o objetivo de medir de forma precisa o tempo de execução de cada etapa e avaliar o desempenho da ferramenta sob condições reais de operação. Os **resultados obtidos foram submetidos à validação da equipe do NIS/CINUGEP**, que analisou os agrupamentos gerados, os percentuais de similaridade identificados e a coerência das análises com os padrões empíricos observados nas auditorias manuais anteriores. Essa validação prática confere robustez aos dados apresentados e reforça a aplicabilidade da solução desenvolvida no contexto institucional do Poder Judiciário tocanтинense.

A seguir, são descritas as etapas executadas pela ferramenta e os respectivos tempos de execução.

Quadro 1 - Etapas executadas

Etapas	Início	Término	Duração	Descrição
Baixa de Petições Iniciais	20:35	20:45	10 min	Download das petições iniciais diretamente do sistema eProc via MNI.
Extração de Textos	20:45	20:51	6 min	Conversão dos arquivos PDF em texto. Utilização de OCR quando necessário.
Análise de Similaridade	20:51	02:23	5h 32min	Execução de 1.326 comparações cruzadas entre os textos das petições.

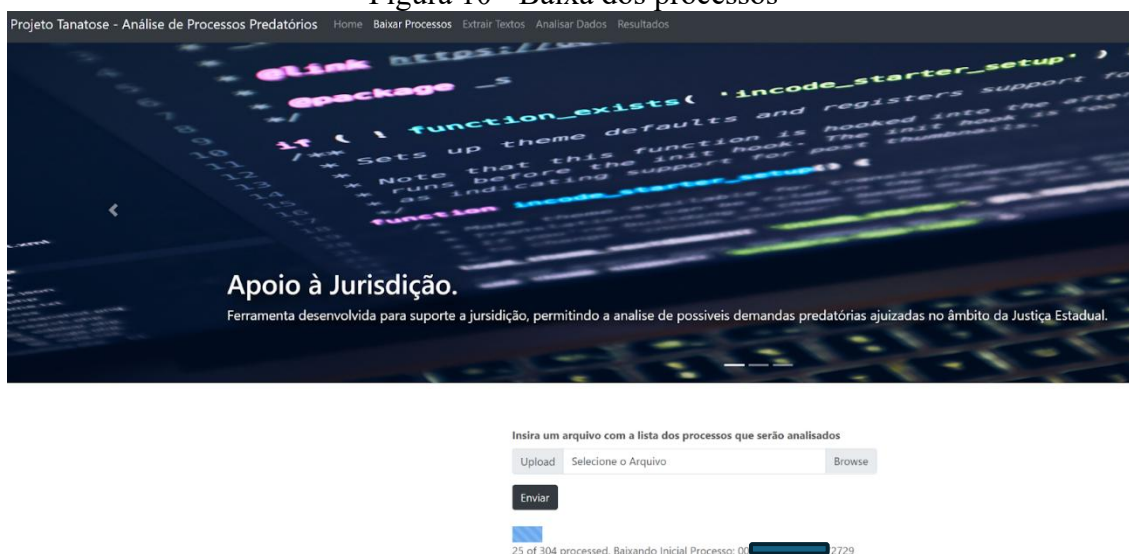
Fonte: Elaborado pelo autor (2025)

4.3.1 Baixa de Petições Iniciais

A etapa inicial, responsável pela recuperação das petições diretamente do sistema eProc por meio do Modelo Nacional de Interoperabilidade (MNI), revelou-se eficiente e estável. A execução foi concluída em aproximadamente 10 minutos para um lote de 304 processos, o que demonstra a eficácia da integração entre a ferramenta Tanatose e o sistema institucional. Esse desempenho evidencia que a solução é tecnicamente viável e suficientemente ágil para ambientes de produção, representando um avanço significativo em relação à prática anterior, que exigia o acesso manual ao sistema, com a abertura individualizada dos autos processuais.

A agilidade obtida nesta fase contribui diretamente para o aumento da produtividade da equipe da Central de Inteligência da Unidade de Gestão Estratégica de Processos (CINUGEP), ao eliminar o tempo ocioso com tarefas operacionais repetitivas. A automatização da baixa das petições iniciais constitui, portanto, um dos pilares do ganho de eficiência proporcionado pela ferramenta Tanatose.

Figura 10 - Baixa dos processos



Fonte: <https://tanatose.tjto.jus.br/download/> (2025)

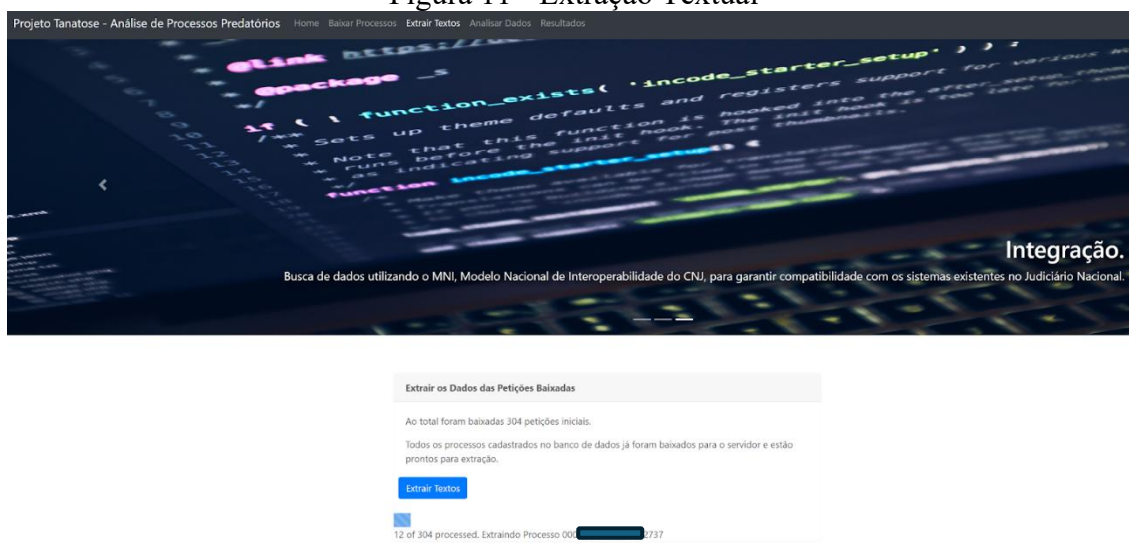
4.3.2 Extração de Textos

A extração do conteúdo textual das petições foi realizada de forma satisfatória, sendo finalizada em apenas 6 minutos para o mesmo conjunto de 304 processos, o que evidencia a eficiência do processamento em lote. Os métodos utilizados para conversão de arquivos PDF mostraram-se adequados, inclusive nos casos em que os documentos estavam digitalizados

como imagem, nos quais foi aplicada a técnica de Reconhecimento Óptico de Caracteres (OCR) por meio da biblioteca Tesseract OCR.

A correta implementação dessa etapa garantiu a integridade dos dados que alimentam a fase seguinte — a análise de similaridade textual — sendo, portanto, essencial para o sucesso da triagem automatizada. Não foram relatadas dificuldades quanto a perdas de informação ou falhas no processo de extração, o que reforça a robustez da abordagem adotada e a confiabilidade dos dados processados pela ferramenta Tanatose.

Figura 11 - Extração Textual

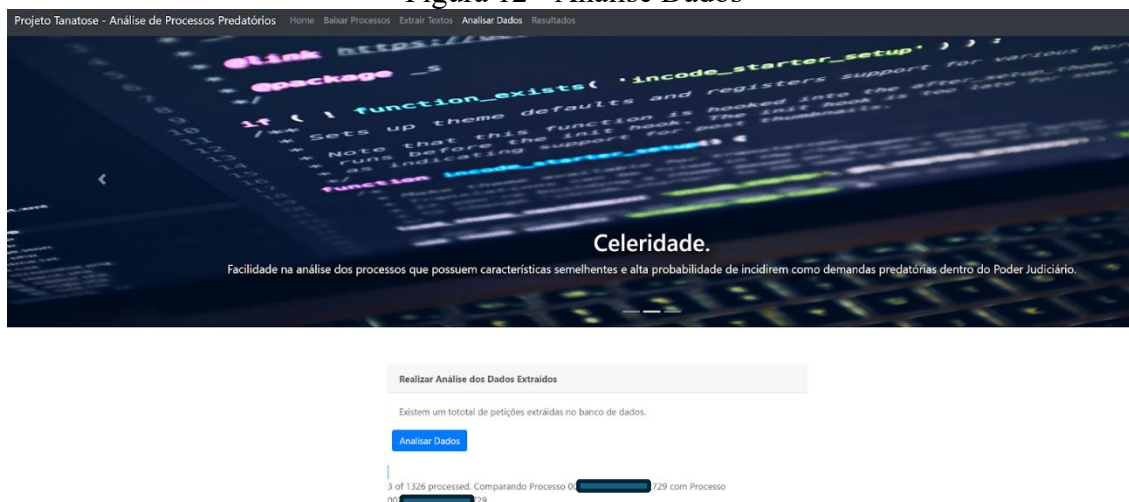


Fonte: <https://tanatose.tjto.jus.br/extrair/> (2025)

4.3.3 Análise de Similaridade Textual

A análise de dados, considerada a etapa mais complexa e intensiva computacionalmente, teve duração total de 5 horas e 32 minutos. Durante esse intervalo, a ferramenta Tanatose realizou 1.326 comparações cruzadas entre os textos das petições iniciais, com base no algoritmo de similaridade textual implementado, fundamentado na abordagem Bag of Words (BoW) e na métrica de Similaridade do Cosseno.

Figura 12 - Análise Dados



Fonte: <https://tanatose.tjto.jus.br/analise/> (2025)

Embora esta seja a fase mais longa do processamento, o tempo empregado é justificado pelo elevado volume de dados e pela complexidade da tarefa. Ainda assim, o desempenho foi considerado satisfatório. O experimento demonstrou que a ferramenta é capaz de processar, com rapidez e eficiência, grandes quantidades de dados judiciais, representando um avanço significativo em relação à análise manual até então realizada pelos servidores.

Do ponto de vista funcional, os resultados foram expressivos: aproximadamente 32% dos pares de processos apresentaram similaridade superior a 80%, o que constitui forte indício de litigância padronizada. Esse dado é especialmente relevante ao se considerar que, sem o apoio da ferramenta, padrões dessa natureza poderiam passar despercebidos ou, no máximo, serem detectados por meio de análises por amostragem.

A estrutura de agrupamento e comparação textual utilizada mostrou-se eficaz para evidenciar semelhanças relevantes entre petições iniciais, mesmo quando havia pequenas variações linguísticas ou estruturais nos textos. Essa capacidade reforça o potencial da Tanatose como instrumento estratégico no combate à litigância predatória.

4.3.4 Relatórios e Visualização dos Resultados

A apresentação final dos resultados por meio de relatórios estruturados e interativos foi um dos pontos mais valorizados pelos usuários da ferramenta Tanatose. Os painéis gerados proporcionam uma visualização clara e organizada dos dados obtidos por meio da análise de similaridade textual entre os processos, permitindo não apenas a compreensão objetiva dos

índices calculados, mas também a personalização da análise conforme critérios definidos pelos auditores.

No estudo de caso com o lote de 304 processos judiciais, foram realizadas 1.326 comparações cruzadas, resultando nos seguintes indicadores:

- **Maior índice de similaridade:** 99,8%, indicando a presença de petições iniciais com conteúdo praticamente idêntico — forte indício de padronização processual;
- **Menor índice de similaridade:** 34,6%, evidenciando documentos com baixa correlação textual;
- **Média geral de similaridade:** 64%, sugerindo um padrão moderado de semelhança entre os processos analisados.

Figura 13 - Resultado Ranking

A **maior** taxa de semelhança foi entre os processos 00 [REDACTED] 2729 e 00 [REDACTED] 2729: 0.998565309251478

A **menor** taxa de semelhança foi entre os processos 00 [REDACTED] 2729 e 00 [REDACTED] 2737: 0.345924690166523

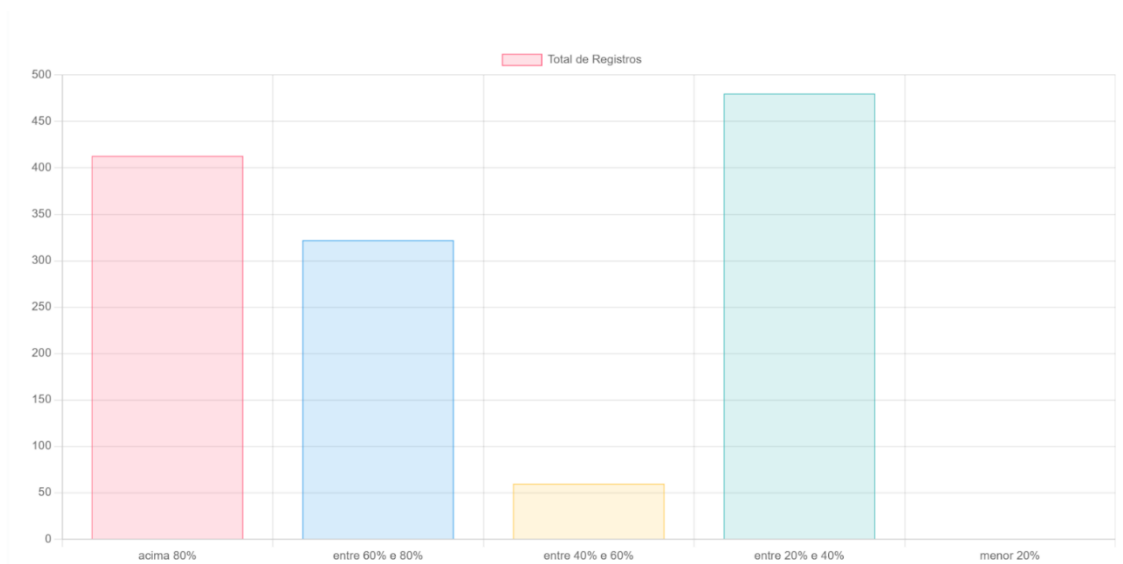
A **média** da taxa de semelhança foi entre os processos foi de: 0.6426172889328271

Fonte: <https://tanatose.tjto.jus.br/resultados/> (2025)

A partir desses dados, o sistema construiu automaticamente uma distribuição percentual dos pares de processos por faixas de similaridade. Essa categorização revelou que:

- 32% dos pares analisados apresentaram similaridade superior a 80%, sinalizando possível litigância repetitiva ou predatória;
- 25% estavam na faixa entre 60% e 80%, indicando padrões intermediários de repetição textual;
- 43% apresentaram similaridade inferior a 60%, o que demonstra uma maior diversidade entre os conteúdos processuais.

Gráfico 5 - Resultado Faixa Similaridade



Fonte: <https://tanatose.tjto.jus.br/resultados/> (2025)

Além da visão geral, os relatórios oferecem ao usuário a possibilidade de customização dos dados exibidos, com seleção de intervalo de similaridade, filtragem por pares mais semelhantes e ordenação dinâmica por número de processo ou índice percentual. No presente caso, foi configurado o filtro para exibir exclusivamente os pares com similaridade superior a 99%, permitindo a identificação de situações de elevada repetição textual com maior grau de confiabilidade.

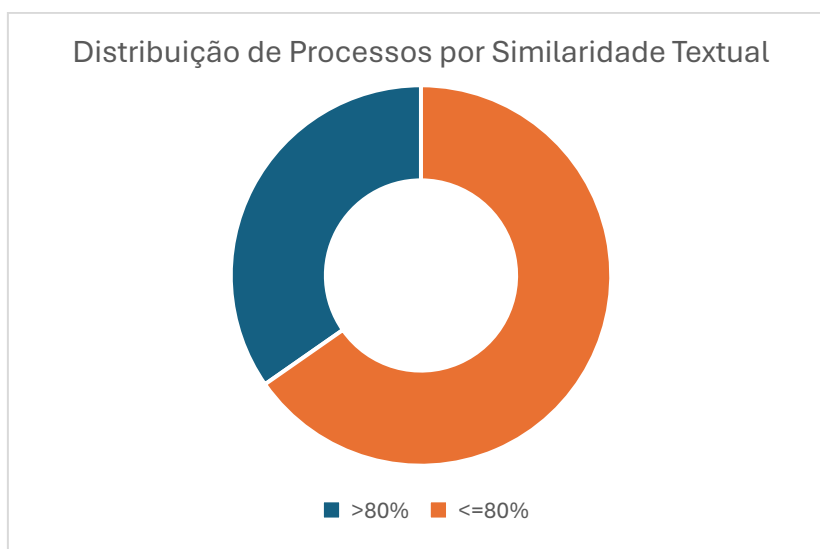
amplamente positiva, destacando-se diversos benefícios em comparação com o processo manual anteriormente adotado.

De acordo com relatos obtidos junto à equipe técnica da Central de Inteligência da Unidade de Gestão Estratégica de Processos (CINUGEP), a utilização da ferramenta proporcionou uma significativa redução no tempo médio necessário para a triagem de processos, além de favorecer maior clareza na priorização dos casos a serem auditados. O elevado volume de processos, aliado à limitação de recursos humanos, tornava inviável a realização de uma auditoria completa por meio da inspeção manual dos autos no sistema eProc. Nessas condições, a única alternativa viável era a análise por amostragem.

Com a adoção da ferramenta Tanatose, tornou-se possível direcionar a análise exclusivamente para os processos identificados como altamente semelhantes, eliminando a necessidade de amostragem e aumentando a assertividade das informações apuradas. A automação do processo de análise textual contribuiu de forma decisiva para a qualificação do trabalho executado pela equipe, conferindo maior racionalidade ao uso do tempo e dos recursos disponíveis.

Considerando que um servidor leva, em média, uma hora para baixar e ler manualmente cada petição inicial, o tempo total estimado para análise do lote de 304 processos seria de aproximadamente 304 horas — o equivalente a cerca de 50 dias úteis, considerando uma jornada diária de 6 horas. Contudo, a ferramenta identificou que 34% dos pares de processos apresentavam similaridade superior a 80%, reduzindo o universo de interesse para 97 processos.

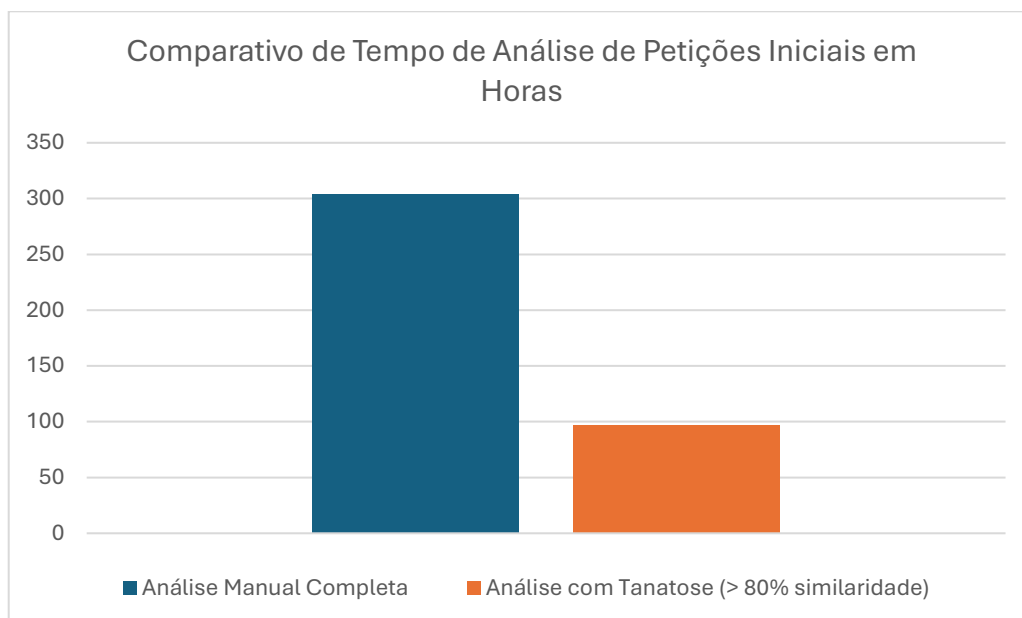
Gráfico 6 - Distribuição de Processos por Similaridade Textual



Fonte: Elaborado pelo autor (2025)

Ao focar apenas nesses casos, o tempo necessário para triagem foi reduzido para 97 horas, ou cerca de 16 dias úteis, resultando em uma economia de 207 horas — o equivalente a 26 dias úteis.

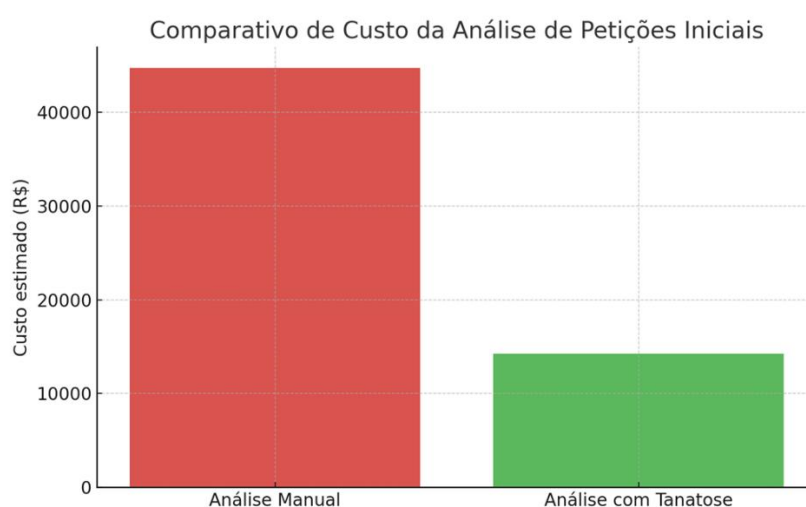
Gráfico 7 - Comparativo de Tempo de Análise de Petições Iniciais em Horas



Fonte: o autor Fonte: Elaborado pelo autor (2025)

Além do ganho temporal, houve impacto financeiro mensurável. Considerando o salário inicial de um Analista Judiciário no TJTO (R\$ 17.662,79 para 120 horas mensais), a hora trabalhada equivale a R\$ 147,18. Assim, a economia estimada gerada pela ferramenta neste único lote de análise foi de aproximadamente R\$ 30.466,26.

Gráfico 8 - Comparativo de Custo da Análise de Petições Iniciais



Fonte: o autor Fonte: Elaborado pelo autor (2025)

Entre os principais ganhos observados, destacam-se:

- Redução expressiva do tempo médio de análise por processo, com eliminação de etapas repetitivas e manuais;
- Diminuição do retrabalho, em virtude da maior precisão na identificação de processos correlatos;
- Melhoria na uniformização dos agrupamentos analisados, garantindo critérios objetivos e padronizados;
- Aprimoramento na qualidade dos relatórios, tornando-os mais completos, fundamentados e úteis para a tomada de decisão.

Em síntese, os resultados demonstram que a ferramenta Tanatose atingiu com sucesso os objetivos propostos em sua concepção. Cada etapa do processo contribuiu de forma significativa para a eficiência da auditoria, possibilitando ganhos operacionais, técnicos e estratégicos. A experiência prática com os servidores validou a aplicabilidade da solução em ambiente real, destacando seu potencial como instrumento de apoio à gestão do acervo judicial no combate à litigância padronizada.

Entretanto, a etapa de análise de similaridade ainda representa um ponto de atenção para futuras otimizações, especialmente no que se refere ao tempo de processamento. Da mesma forma, a ausência de um mecanismo de armazenamento histórico das análises pode ser considerada uma limitação a ser superada em versões futuras da ferramenta, ampliando sua capacidade de fornecer inteligência histórica ao processo decisório.

5 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as principais conclusões da pesquisa, destacando a contribuição prática e científica da solução proposta, suas limitações, o cumprimento dos objetivos iniciais e as perspectivas para futuras investigações.

5.1 Problema e Solução Proposta

Este trabalho partiu da constatação de uma fragilidade concreta no Poder Judiciário do Estado do Tocantins: a dificuldade de detectar rapidamente litigâncias predatórias no ambiente eletrônico do eProc. Embora a digitalização tenha ampliado o acesso à Justiça, também facilitou o ajuizamento em massa de ações com petições quase idênticas, criando um novo desafio institucional.

Para enfrentar esse problema, foi concebida a ferramenta Tanatose, baseada em técnicas de inteligência computacional. A solução automatiza a análise textual de petições iniciais, identificando agrupamentos de processos com alto grau de repetição. Com isso, oferece suporte estratégico à auditoria processual realizada pela Central de Inteligência da Unidade de Gestão Estratégica de Processos (CINUGEP), proporcionando maior eficiência, agilidade e assertividade na triagem de casos suspeitos de litigância predatória.

5.2 Contribuições da Pesquisa

Esta pesquisa trouxe contribuições práticas e científicas relevantes. No âmbito institucional, a ferramenta Tanatose demonstrou eficácia ao automatizar a análise de grandes volumes de processos, reduzindo o tempo de triagem, eliminando a necessidade de amostragem e elevando a assertividade das auditorias realizadas pela equipe do TJTO.

Ao permitir que os servidores concentrem esforços em casos prioritários, a solução contribuiu diretamente para otimizar recursos humanos e melhorar a qualidade das informações apuradas.

Cientificamente, o trabalho reforçou a viabilidade da aplicação de técnicas de mineração de dados e processamento de linguagem natural no ambiente jurídico, valorizando a interdisciplinaridade entre Computação e Direito. A metodologia privilegiou a explicabilidade

e a rastreabilidade dos resultados, pilares fundamentais para a adoção de tecnologias no setor público.

5.3 Limitações do Estudo

A principal limitação da pesquisa reside no escopo pontual da ferramenta Tanatose. A solução foi desenvolvida para apoiar auditorias específicas, sem manter armazenamento permanente dos dados analisados, em respeito ao princípio da minimização de dados e para simplificar a operação do sistema.

Embora essa escolha atenda aos objetivos imediatos da pesquisa, ela restringe a possibilidade de análises históricas sobre a evolução das práticas predatórias ao longo do tempo. Essa limitação abre espaço para aprimoramentos em versões futuras da ferramenta, buscando ampliar seu potencial estratégico.

5.4 Aspectos Éticos

Todos os processos utilizados para validação da ferramenta Tanatose foram públicos, sem qualquer grau de sigilo, e a análise respeitou integralmente os princípios legais e éticos previstos no ordenamento jurídico, em especial a Lei Geral de Proteção de Dados (LGPD). O sistema foi projetado para atuar exclusivamente sobre o texto da petição inicial, preservando a confidencialidade das demais informações processuais.

5.5 Propostas de Trabalhos Futuros

O desenvolvimento da ferramenta Tanatose demonstrou o potencial da inteligência computacional no apoio à identificação de litigâncias predatórias no Tribunal de Justiça do Estado do Tocantins. No entanto, diversas oportunidades de aprimoramento e aprofundamento da pesquisa foram identificadas durante a realização deste estudo.

As propostas de trabalhos futuros incluem:

- **Armazenamento histórico das auditorias realizadas:** Implementar um módulo de persistência de dados que permita armazenar os resultados das análises realizadas pela ferramenta. Essa funcionalidade possibilitará a

construção de séries históricas, o acompanhamento da reincidência de práticas predatórias e a geração de relatórios estatísticos para subsidiar a gestão judicial.

- **Adoção de técnicas de inteligência computacional mais avançadas:** Explorar o uso de modelos de aprendizado profundo (deep learning) e representações semânticas de textos baseadas em embeddings, como Word2Vec ou BERT. A utilização dessas técnicas poderá aumentar a acurácia da detecção de similaridades entre petições iniciais e permitir uma análise textual mais sofisticada.
- **Realização de estudos comparativos formais:** Conduzir estudos de caso controlados que comparem auditorias realizadas manualmente com auditorias assistidas pela ferramenta Tanatose. A comparação de indicadores como tempo de análise, taxa de identificação de litigâncias predatórias e nível de precisão poderá evidenciar os ganhos proporcionados pela automação.

Então, a implementação dessas propostas poderá expandir significativamente o potencial da ferramenta, tornando-a um instrumento ainda mais eficaz no combate à litigância predatória e na modernização da gestão processual no Poder Judiciário.

REFERÊNCIAS

AGGARWAL, Charu C. Machine learning for text. Springer, 2018.

ANGELLO, Tiago; VITAL, Danilo. Em São Paulo, litigância predatória responde por 337 mil processos por ano. Consultor Jurídico, São Paulo, 9 out. 2023. Disponível em: <https://www.conjur.com.br/2023-out-09/sp-litigancia-predatoria-responsavel-337-mil-processos-ano/>. Acesso em: 25 maio 2025.

BIRD, Steven; LOPER, Edward. NLTK: The natural language toolkit. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2004. p. 214–217.

BRASIL. Constituição da República Federativa do Brasil de 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 27 abr. 2025.

BRASIL. Lei nº 11.419, de 19 de dezembro de 2006. Dispõe sobre a informatização do processo judicial e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, 20 dez. 2006. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/11419.htm. Acesso em: 27 abr. 2025.

BRASIL. Superior Tribunal de Justiça. Tema Repetitivo 1198 – Possibilidade de imposição de multa por litigância de má-fé e honorários advocatícios a escritórios de advocacia em demandas padronizadas. Brasília: STJ, 2023. Disponível em: <https://www.stj.jus.br/sites/portalp/Temas-Repetitivos/tema-1198>. Acesso em: 25 maio 2025.

CNJ – CONSELHO NACIONAL DE JUSTIÇA. Litigância predatória. Disponível em: <https://www.cnj.jus.br/programas-e-acoes/litigancia-predatoria/>. Acesso em: 2 abr. 2025.

CONJUR. Impactos e desafios da litigância predatória no sistema judicial. 2024. Disponível em: <https://www.conjur.com.br/2024-ago-09/impactos-e-desafios-da-litigancia-predatoria-no-sistema-judicial/>. Acesso em: 2 abr. 2025.

CONSELHO NACIONAL DE JUSTIÇA. Modelo Nacional de Interoperabilidade – MNI. Brasília: CNJ, 2023. Disponível em: <https://www.cnj.jus.br/sistemas/modelo-nacional-de-interoperabilidade-mni/>. Acesso em: 13 abr. 2025.

CONSELHO NACIONAL DE JUSTIÇA. Programa Justiça 4.0: Inovação e efetividade na realização da justiça. Brasília: CNJ, 2020. Disponível em: <https://www.cnj.jus.br/justica-4-0/>. Acesso em: 27 abr. 2025.

CONSELHO NACIONAL DE JUSTIÇA. Relatório Justiça em Números 2023. Brasília: CNJ, 2023. Disponível em: <https://www.cnj.jus.br>. Acesso em: 24 abr. 2025.

FUX, Luiz. O novo processo civil brasileiro: uma construção constitucional. 2. ed. São Paulo: Forense, 2021.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining: concepts and techniques. 3. ed. San Francisco: Morgan Kaufmann, 2012.

HONNIBAL, Matthew; MONTANI, Ines; VAN LANDGEGHEM, Sofie; BOITELLE, Adriane. spaCy: industrial-strength natural language processing in Python. 2020. Disponível em: <https://spacy.io/>. Acesso em: 27 abr. 2025.

JOTA. Litigância predatória: a estranha história de uma microempresa que tem 303 ações trabalhistas. JOTA, São Paulo, 2024. Disponível em: <https://www.jota.info/tributos-e-empresas/trabalho/litigancia-predatoria-a-estranha-historia-de-uma-microempresa-que-tem-303-acoes-trabalhistas>. Acesso em: 25 maio 2025.

JURAFSKY, Daniel; MARTIN, James H. Speech and language processing. 3. ed. draft. Stanford: Stanford University, 2020. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 24 abr. 2025.

KALID, Guilherme. Tanatose. InfoEscola. Disponível em: <https://www.infoescola.com/biologia/tanatose/>. Acesso em: 20 abr. 2025.

KRUSE, Rudolf; BORGELT, Christian; STEINBRECHER, Martin; KLAR, Rainer. Computational intelligence: a methodological introduction. 2. ed. London: Springer, 2013.

MACHADO, Daniel. Direito e inteligência artificial: a necessidade de adaptação dos sistemas jurídicos. Revista de Estudos Jurídicos da UNESP, v. 25, n. 2, 2021.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. Introduction to information retrieval. Cambridge: Cambridge University Press, 2008.

MANNING, Christopher D.; SCHÜTZE, Hinrich. Foundations of statistical natural language processing. Cambridge: MIT Press, 1999.

MITCHELL, Tom M. Machine learning. New York: McGraw-Hill, 1997.

PODER JUDICIÁRIO DO ESTADO DO TOCANTINS. Instrução Normativa nº 5, de 24 de outubro de 2011. Disponível em: <https://www.tjto.jus.br/>. Acesso em: 27 abr. 2025.

POOLE, David; MACKWORTH, Alan; GOEBEL, Randy. Computational intelligence: a logical approach. New York: Oxford University Press, 1998.

PRESSMAN, Roger S. Engenharia de Software. 8. ed. Porto Alegre: AMGH, 2016.

RANGANATHAN, S. R. Elements of information retrieval. 2. ed. New Delhi: Ess Ess Publications, 2008.

RODOLFO JUNIOR, Renato Tadeu. Adoção do sistema e-Proc pelos tribunais brasileiros. In: AJUFESC. Congresso de Magistrados Federais, 2023. Disponível em: <https://ajufesc.org.br/wp-content/uploads/2023/09/Renato-Tadeu-Rodolfo-Junior.pdf>. Acesso em: 24 abr. 2025.

SANTOS, Fabiano Hartmann dos. Aplicação da inteligência artificial no direito brasileiro: desafios e limitações. Revista de Direito, Tecnologia e Sociedade, v. 1, n. 1, 2020.

SILVA, João; PEREIRA, Marcos. Inteligência artificial aplicada à justiça: perspectivas e desafios. Revista Brasileira de Inovação Tecnológica, v. 10, n. 2, 2021.

SILVA, Rafael A.; PEREIRA, Luana M. Tecnologias aplicadas ao cuidado paliativo: uma revisão integrativa. Revista Brasileira de Cuidados Paliativos, v. 13, n. 2, p. 123–130, 2021. Disponível em: <https://www.rbc.org.br/details/tecnologias-aplicadas-ao-cuidado-paliativo-uma-revisao-integrativa>. Acesso em: 24 abr. 2025.

SURESH, Harini; GUTTAG, John V. A framework for understanding unintended consequences of machine learning. Communications of the ACM, v. 64, n. 11, p. 62–71, 2021.

SURDEN, Harry. Machine learning and law. Washington Law Review, v. 89, p. 87–115, 2014.

TJDFT – TRIBUNAL DE JUSTIÇA DO DISTRITO FEDERAL E DOS TERRITÓRIOS. Litigância predatória compromete garantia constitucional. 2022. Disponível em:

<https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2022/litigancia-predatoria-compromete-garantia-constitucional>. Acesso em: 27 abr. 2025.

TJTO – TRIBUNAL DE JUSTIÇA DO ESTADO DO TOCANTINS. Eproc - Tribunal de Justiça do Tocantins. Palmas, TO: TJTO, 2025. Disponível em: <https://www.tjto.jus.br/eproc>. Acesso em: 24 abr. 2025.

TRF4 – TRIBUNAL REGIONAL FEDERAL DA 4ª REGIÃO. TRF4 cede sistema processual e-Proc para Tribunal de Justiça do Tocantins. Porto Alegre: TRF4, 9 jul. 2019. Disponível em: https://www.trf4.jus.br/trf4/controlador.php?acao=noticia_visualizar&id_noticia=14507. Acesso em: 24 abr. 2025.

TRIBUNAL DE JUSTIÇA DO ESTADO DO TOCANTINS. Gestão de Pessoas – Anexo III-A – Vencimentos. Disponível em: https://gestaodepessoas.tjto.jus.br/site/transparencia/anexo_iii_a. Acesso em: 25 maio 2025.

WANG, Fei-Yue; LIU, Derong. Advances in computational intelligence: theory and applications. London: World Scientific, 2006.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. Data mining: practical machine learning tools and techniques. 3. ed. Burlington: Morgan Kaufmann, 2011.

6 APÊNDICE

6.1 Especificação de Casos de uso

Caso se Uso 1 – Importar Lista de Processos

Item	Descrição
Casos de Uso	Importar Lista de Processos
Ação	Importar
Descrição	Permite ao servidor importar uma lista de números de processos a serem analisados, preenchendo a base de dados inicial do sistema.
Atores	Servidor do TJTO
Precondições	O usuário deve estar na rede interna do TJTO e acessado o sistema.
Pós-condições	A lista de processos será armazenada na tabela de análise.
Restrições	Impedir importação de arquivos em formato inválido.
Cenário de insucesso	O sistema exibe uma mensagem de erro se o arquivo de entrada estiver com problemas de formato ou dados inválidos.

Caso de Uso 2 - Baixar Petições Iniciais

Item	Descrição
Casos de Uso	Baixar Petições Iniciais
Ação	Baixar
Descrição	Permite ao sistema automatizar a busca e o download das petições iniciais dos processos indicados.
Atores	Sistema Tanatose
Precondições	Lista de processos deve estar cadastrada.
Pós-condições	Os arquivos das petições serão armazenados localmente.
Restrições	Acesso somente a processos públicos; petições protegidas não serão baixadas.
Cenário de insucesso	Mensagem de erro exibida se a petição não for localizada ou o acesso for negado.

Caso de Uso 3 - Extrair Texto das Petições

Item	Descrição
Casos de Uso	Extrair Texto das Petições
Ação	Extrair
Descrição	Realiza a leitura e extração automática do conteúdo textual das petições iniciais baixadas.
Atores	Sistema Tanatose
Precondições	As petições devem estar baixadas corretamente.
Pós-condições	O texto extraído será salvo para posterior análise.
Restrições	Arquivos corrompidos ou não lidos serão descartados.
Cenário de insucesso	O sistema registra erro e não extrai conteúdo se o arquivo estiver ilegível.

Caso de Uso 4 - Analisar Similaridade entre Petições

Item	Descrição
Casos de Uso	Analisar Similaridade entre Petições
Ação	Analisar
Descrição	Compara o texto das petições utilizando técnicas de similaridade textual para identificar possíveis repetições predatórias.
Atores	Sistema Tanatose
Precondições	Os textos das petições devem estar extraídos.
Pós-condições	Gera um índice de similaridade para cada par de processos analisados.
Restrições	Análise restrita a petições com conteúdo disponível.
Cenário de insucesso	Se a comparação falhar, o sistema registra o par como "não analisado".

Caso de Uso 5 - Gerar Relatório de Similaridade

Item	Descrição
Casos de Uso	Gerar Relatório de Similaridade
Ação	Gerar
Descrição	Gera relatórios com os resultados das análises de similaridade entre processos, apresentando índices e agrupamentos.
Atores	Servidor do TJTO
Precondições	Análises de similaridade devem estar concluídas.
Pós-condições	Relatórios salvos ou exportados para consulta.
Restrições	Permissão de acesso somente rede interna do TJTO.
Cenário de insucesso	Mensagem de nenhum dado se não houver dados para gerar o relatório.

6.2 Dicionário de Dados

Tabela: analiseprocessospredatorios_processosanalise

Campo	Tipo	Tamanho	Descrição	Observações
id	bigint	-	Identificador único do processo para análise.	Chave primária. Gerado automaticamente.
num_processo	varchar	20	Número do processo judicial.	Obrigatório.
sin_baixado	varchar	1	Indicador se o arquivo da petição foi baixado. (S/N)	Padrão: "N". Atualizado pelo sistema.
sin_extraido	varchar	1	Indicador se o texto da petição foi extraído. (S/N)	Padrão: "N". Atualizado pelo sistema.
sin_analizado	varchar	1	Indicador se o processo já foi analisado quanto à similaridade. (S/N)	Padrão: "N". Atualizado pelo sistema.

Tabela: analiseprocessospredatorios_analisedados

Campo	Tipo	Tamanho	Descrição	Observações
id	bigint	-	Identificador único da análise de similaridade.	Chave primária. Gerado automaticamente.
num_processo1_id	bigint	-	Identificador do primeiro processo comparado.	Chave estrangeira para "processosanalyse".
num_processo2_id	bigint	-	Identificador do segundo processo comparado.	Chave estrangeira para "processosanalyse".
indice_simelhanca	double	-	Valor de similaridade entre os dois processos (0 a 1).	Calculado via Similaridade do Cosseno.

Tabela: analiseprocessospredatorios_processotexto

Campo	Tipo	Tamanho	Descrição	Observações
id	bigint	-	Identificador único do texto da petição.	Chave primária. Gerado automaticamente.

num_processo_id	bigint	-	Identificador do processo associado ao texto.	Chave estrangeira para "processosanalyse".
conteudo_peticao	longtext	-	Texto completo da petição inicial extraído do arquivo.	Pode ser grande (longtext).

6.3 Comparação de Dados de Técnicas de Similaridade

Os Dados abaixo foram extraídos das mesmas Petições iniciais, e os campos num_processo1_id e num_processo2_id são representações numéricas dos processos dentro do banco de dados próprio e não representam os números reais dos processos.

Coefficiente de Jaccard

linha	num_processo1_id	num_processo2_id	porcentagem
1	37729	37730	25,75%
2	37729	37731	96,51%
3	37729	37732	96,37%
4	37729	37733	96,79%
5	37729	37734	96,37%
6	37729	37735	96,37%
7	37729	37736	96,64%
8	37729	37737	96,24%
9	37729	37738	96,51%
10	37730	37731	25,92%
11	37730	37732	25,75%
12	37730	37733	25,89%
13	37730	37734	25,78%
14	37730	37735	25,78%
15	37730	37736	25,86%
16	37730	37737	25,73%
17	37730	37738	25,84%
18	37731	37732	95,96%
19	37731	37733	96,10%
20	37731	37734	96,50%
21	37731	37735	96,50%
22	37731	37736	96,78%
23	37731	37737	96,37%
24	37731	37738	96,37%
25	37732	37733	96,51%
26	37732	37734	96,92%
27	37732	37735	96,92%
28	37732	37736	96,37%
29	37732	37737	97,06%
30	37732	37738	97,06%
31	37733	37734	96,51%
32	37733	37735	96,51%
33	37733	37736	96,51%
34	37733	37737	96,65%

35	37733	37738	96,37%
36	37734	37735	98,03%
37	37734	37736	97,47%
38	37734	37737	96,23%
39	37734	37738	97,89%
40	37735	37736	97,75%
41	37735	37737	96,23%
42	37735	37738	98,17%
43	37736	37737	96,51%
44	37736	37738	97,61%
45	37737	37738	96,37%

Distância Euclidiana

linha	num_processo1_id	num_processo2_id	porcentagem
1	37729	37730	0,57%
2	37729	37731	14,12%
3	37729	37732	12,39%
4	37729	37733	14,29%
5	37729	37734	12,18%
6	37729	37735	12,18%
7	37729	37736	12,50%
8	37729	37737	12,28%
9	37729	37738	12,39%
10	37730	37731	0,57%
11	37730	37732	0,57%
12	37730	37733	0,57%
13	37730	37734	0,57%
14	37730	37735	0,57%
15	37730	37736	0,57%
16	37730	37737	0,57%
17	37730	37738	0,57%
18	37731	37732	12,08%
19	37731	37733	12,50%
20	37731	37734	12,28%
21	37731	37735	12,28%
22	37731	37736	12,61%
23	37731	37737	12,39%
24	37731	37738	12,28%
25	37732	37733	12,61%
26	37732	37734	13,96%
27	37732	37735	13,10%
28	37732	37736	12,50%
29	37732	37737	13,23%

30	37732	37738	14,64%
31	37733	37734	12,39%
32	37733	37735	12,39%
33	37733	37736	12,50%
34	37733	37737	12,73%
35	37733	37738	12,39%
36	37734	37735	13,96%
37	37734	37736	12,97%
38	37734	37737	12,08%
39	37734	37738	15,44%
40	37735	37736	13,51%
41	37735	37737	12,28%
42	37735	37738	14,64%
43	37736	37737	12,61%
44	37736	37738	13,51%
45	37737	37738	13,23%

Similaridade do Cosseno

linha	num_processo1_id	num_processo2_id	porcentagem
1	37729	37730	66,18%
2	37729	37731	99,75%
3	37729	37732	99,67%
4	37729	37733	99,76%
5	37729	37734	99,65%
6	37729	37735	99,65%
7	37729	37736	99,67%
8	37729	37737	99,66%
9	37729	37738	99,67%
10	37730	37731	66,22%
11	37730	37732	66,16%
12	37730	37733	66,18%
13	37730	37734	66,17%
14	37730	37735	66,12%
15	37730	37736	66,15%
16	37730	37737	66,06%
17	37730	37738	66,20%
18	37731	37732	99,65%
19	37731	37733	99,67%
20	37731	37734	99,66%
21	37731	37735	99,66%
22	37731	37736	99,68%
23	37731	37737	99,67%
24	37731	37738	99,66%

25	37732	37733	99,68%
26	37732	37734	99,75%
27	37732	37735	99,71%
28	37732	37736	99,67%
29	37732	37737	99,71%
30	37732	37738	99,77%
31	37733	37734	99,67%
32	37733	37735	99,67%
33	37733	37736	99,67%
34	37733	37737	99,69%
35	37733	37738	99,67%
36	37734	37735	99,75%
37	37734	37736	99,70%
38	37734	37737	99,65%
39	37734	37738	99,80%
40	37735	37736	99,73%
41	37735	37737	99,66%
42	37735	37738	99,77%
43	37736	37737	99,68%
44	37736	37738	99,73%
45	37737	37738	99,71%