

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

JEAN AMPOS FLESCH

**Uma Abordagem Agnóstica para Geração
de Texto Legal e aplicação em tarefas da
área de *LegalTech***

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em Ciência da
Computação

Orientador: Profa. Dra. Renata Galante

Porto Alegre
2023

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Flesch, Jean Ampos

Uma Abordagem Agnóstica para Geração de Texto Legal e aplicação em tarefas da área de *LegalTech* / Jean Ampos Flesch. – Porto Alegre: PPGC da UFRGS, 2023.

116 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2023. Orientador: Renata Galante.

1. Inteligência Artificial. 2. Aprendizado Supervisionado de Máquina. 3. LegalTech. 4. Geração de texto legal. 5. Classificação de documentos. I. Galante, Renata. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Prof^a. Patricia Pranke

Pró-Reitor de Pós-Graduação: Prof. Júlio Otávio Jardim Barcellos

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. Claudio Rosito Jung

Bibliotecário-chefe do Instituto de Informática: Alexander Borges Ribeiro

*“Enquanto se possa mexer, treine o corpo.
Enquanto não se possa mexer, treine a mente.”*

— BUSHIDO

*“Nós somos o que repetidamente fazemos.
Excelência, então, não é um ato, mas um hábito.”*

— WILL DURANT

AGRADECIMENTOS

Agradeço aos meus pais, Silvana Flesch e Sergio Flesch, pela educação recebida, pelos constantes ensinamentos e incentivos, pelos sacrifícios realizados, pelo apoio incondicional nos momentos de dificuldade, pela compreensão durante as súbitas alterações de humor.

Agradeço à UFRGS, ao Instituto de Informática, e ao PPGC pela qualificação e oportunidade de realizar os cursos de graduação e mestrado, pelo ambiente criativo, pelo aprendizado concedido e pela excelência no ambiente de ensino.

Agradeço a minha orientadora, Prof^{ra}. Dr^a. Renata Galante, por todos os anos de orientação, motivação, paciência, atenção e suporte em momentos difíceis ao longo da minha jornada acadêmica, tanta na graduação quanto no mestrado.

Agradeço a minha namorada, Daniela Heisler, pelo amor, pelo companheirismo e parceria, pela paciência em crises de ansiedade e estresse, pela compreensão em momentos de ausência.

Por último, mas não menos importante, agradeço a todos amigos e colegas de mestrado e trabalho que fiz ao longo dos últimos anos. Em especial, agradeço ao colega Régis Ebeling, pela amizade construída na graduação e mestrado, pelas assistências e conversas produtivas, pelos conselhos, e, claro, pelos inúmeros momentos de descontração que já contemplam três Copas do Mundo. Agradeço também ao colega Ruan Putka pela ajuda e pela disposição com suas dicas na área de ciência de dados.

RESUMO

A área de *LegalTech* é referida como a relação entre áreas da Ciência da Computação que operam em linguagem natural e áreas jurídicas. A proximidade entre as áreas da Linguagem e Direito possibilita que a *LegalTech* seja aplicada em diversas tarefas como geração de texto legal, predição de sentenças, recuperação de texto legal. Porém, apesar da flexibilidade em aplicações, muitas soluções apresentam lacunas em sua metodologia de implementação e/ou disponibilização das bases de dados extraídas e pré-processadas. Neste trabalho, é proposta uma Abordagem Agnóstica com o objetivo de providenciar recomendações para o desenvolvimento de soluções na área de *LegalTech* de forma coesa, estruturada, extensível, e robusta. Inicialmente, foi realizada a revisão sistemática da literatura para identificar e apresentar um panorama sobre diferentes soluções propostas na geração de texto legal. A partir da revisão sistemática, uma taxonomia para classificar os estudos em cinco categorias de acordo com o escopo principal de cada trabalho foi proposta a fim de embasar a Abordagem Agnóstica, que é composta por 4 etapas responsáveis pela coleta, processamento, modelagem e exibição de resultados a partir da utilização de técnicas de Inteligência Artificial e Processamento de Linguagem Natural em documentos legais. A abordagem é testada em três tarefas diferentes. A primeira tarefa testa a capacidade de construção de uma base de dados com documentos legais a partir da extração de informações dos acórdãos em páginas *Web* do Supremo Tribunal Brasileiro. Além da organização e disponibilização da base de dados, uma análise com visualização de informações sobre os dados é apresentada. A segunda tarefa realiza a classificação de palavras em documentos legais para geração de texto anotado, resultando no primeiro passo para a construção de uma solução capaz de facilitar a leitura de texto legal para pessoas leigas na área jurídica. Por fim, a terceira tarefa utiliza a base de dados construída para experimentos em classificação de documentos legais utilizando algoritmos de aprendizado de máquina supervisionado em modelos classificadores. Na realização dos experimentos, o uso dos modelos apresentou resultados promissores para predição das classes dos acórdãos. Em especial, o uso do classificador SVM resultou em um F1-score médio de 0,96.

Palavras-chave: Inteligência Artificial. Aprendizado Supervisionado de Máquina. *LegalTech*. Geração de texto legal. Classificação de documentos.

An Agnostic Approach to Legal Text Generation and Application to LegalTech Tasks

ABSTRACT

The area of LegalTech is referred as the relationship between fields of Computer Science that operate in natural language and fields of law. The proximity between the fields of Language and Law allows LegalTech to be applied in various tasks such as legal text generation, sentence prediction, legal text recovery. However, despite the flexibility in the range of applications, many solutions have gaps in their implementation methodology and/or availability of extracted and preprocessed databases. In this work, an Agnostic Approach is proposed with the objective of providing guidelines for the development of solutions in the area of LegalTech in a cohesive, structured, extensible, and robust way. Initially, a systematic literature review is performed to identify and present an overview of different solutions proposed in the generation of legal text. From the systematic review, a taxonomy to classify the studies into five categories according to the main scope of each work is proposed in order to support the Agnostic Approach, which is composed of 4 steps responsible for collecting, processing, modeling and displaying results from the use of Artificial Intelligence and Natural Language Processing techniques in legal documents. The approach is tested on three different tasks. The first task tests the ability to build a database with legal documents by extracting information from judgments on Web pages of the Brazilian Supreme Court. In addition to the organization and availability of the database, an analysis with visualization of information about the data is presented. The second task performs the classification of words in legal documents to generate annotated text, resulting in the first step towards the construction of a solution capable of facilitating the reading of legal texts for non practitioner people in the legal area. Finally, the third task uses the database built for experiments in legal document classification using supervised machine learning algorithms in classifier models. During the execution of the experiments, the use of models showed promising results for predicting the classes of judgments. In particular, the use of the SVM classifier resulted in an average F1-score of 0.96.

Keywords: Artificial Intelligence, Supervised Machine Learning, LegalTech, Legal text generation, Document classification.

LISTA DE ABREVIATURAS E SIGLAS

BOW	<i>Bag-of-words</i> - Saco-de-palavras
CNN	<i>Convolutional Neural Network</i> - Rede Neural Convolucional
IA	Inteligência Artificial
PLN	Processamento de Linguagem Natural
RC	Representação de Conhecimento
RI	Recuperação de informação
RF	<i>Random Forest</i> - Floresta aleatória
SVM	<i>Support Vector Machine</i> - Máquina de Vetores de Suporte
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i> - Frequência do Termo – Frequência Inversa dos Documentos

LISTA DE FIGURAS

Figura 2.1	Visão geral do papel dos campos de IA e PLN na área de <i>Legal Tech</i>	19
Figura 2.2	Representação de um hiperplano de SVM.	22
Figura 2.3	Visão geral sobre árvores de decisão em função da votação majoritária no algoritmo RF.	23
Figura 2.4	Exemplo de identificação e classificação de entidades em uma frase através de NER.	24
Figura 3.1	Processo de seleção de estudos e quantidade de artigos extraídos após cada etapa.....	28
Figura 3.2	Taxonomia de tarefas da LegalTech.	37
Figura 3.3	Representação da incidência de áreas da Ciência da Computação por categorias de taxonomia.....	47
Figura 3.4	Número de artigos publicados em cada um dos anos de 1963 a 2020 de acordo com seu objetivo principal.	51
Figura 3.5	A linha de tendência do total de artigos selecionados publicados dos anos de 1963 a 2020.....	52
Figura 3.6	O número de trabalhos publicados em cada país dos anos de 1963 a 2020. .	53
Figura 5.1	Arquitetura para execução de tarefas para geração de texto legal.	65
Figura 5.2	Etapa de Extração.	66
Figura 5.3	Etapa de Pré-processamento de dados da abordagem.	67
Figura 5.4	Etapa de Análise e Modelagem.	68
Figura 5.5	Etapa de Exibição.	69
Figura 6.1	Serviços que constituem o protótipo LEAF.....	71
Figura 6.2	Exemplo de informações exibidas de acórdãos em páginas do STF.	72
Figura 6.3	Tela inicial do DataToChart service exibindo todas as subpáginas disponíveis para consulta de dados a partir de diferentes critérios.....	75
Figura 6.4	Exemplo de página <i>Web</i> com texto jurídico anotado e marcações em palavras pertencentes a determinadas classes do modelo.	77
Figura 6.5	Quantidade de acórdãos por período. (a) De 2012 à agosto de 2022, (b) de 1934 à agosto de 2022.....	79
Figura 6.6	Total de acórdãos julgados por Ministro relator no período de janeiro de 2012 à agosto de 2022.....	80
Figura 6.7	Órgão julgador dos acórdãos no STF entre o período de janeiro de 2012 à agosto de 2022.....	81
Figura 6.8	Distribuição de acórdãos por classe processual entre o período de janeiro de 2012 à agosto de 2022.....	82
Figura 6.9	Distribuição de acórdãos por estado brasileiro entre o período de janeiro de 2012 à agosto de 2022.....	83
Figura 6.10	Metodologia utilizada nos experimentos.	84
Figura 6.11	Visão geral das etapas realizadas para configuração e calibragem de <i>stopwords</i> , <i>datasets</i> , parâmetros dos classificadores, e número de <i>features</i> , respectivamente.....	85
Figura 6.12	F1-score dos <i>datasets</i> de acordo com o número de <i>features</i> no modelo classificador NB.	91
Figura 6.13	F1-score dos <i>datasets</i> de acordo com o número de <i>features</i> no modelo classificador SVM.....	92

Figura 6.14 F1-score dos <i>datasets</i> de acordo com o número de <i>features</i> no modelo classificador RF.....	93
Figura 6.15 Tempo de execução do <i>Dataset A</i> de acordo com o número de <i>features</i> e modelos classificadores.....	95
Figura 6.16 Tempo de execução do <i>Dataset B</i> de acordo com o número de <i>features</i> e modelos classificadores.....	95
Figura 6.17 Tempo de execução do <i>Dataset C</i> de acordo com o número de <i>features</i> e modelos classificadores.....	96

LISTA DE TABELAS

Tabela 3.1 Tabela de extração de dados.	31
Tabela 3.2 Distribuição dos artigos por categoria dos 146 estudos selecionados para leitura completa.....	32
Tabela 3.3 Distribuição das categorias de taxonomia definidas, por objetivo principal, dos 146 estudos selecionados.....	46
Tabela 3.4 Distribuição dos artigos, em ordem crescente, das 10 áreas de ciência da computação mais recorrentes dos 146 estudos selecionados.	48
Tabela 3.5 Classificação em automática, semiautomática ou manual dos 132 artigos selecionados.....	50
Tabela 3.6 Distribuição dos artigos publicados por ano dos 146 estudos selecionados..	54
Tabela 4.1 Tabela comparativa dos <i>datasets</i> jurídicos públicos selecionados.	61
Tabela 4.2 Tabela comparativa de trabalhos brasileiros na área de LegalTech.	63
Tabela 6.1 Tabela com as informações coletadas e o tipo de dado atribuídos pelo STFScrap service.	72
Tabela 6.2 Tabela com as colunas da coleção de acórdãos importados ao MongoDB. ..	74
Tabela 6.3 Visão geral dos <i>datasets</i> utilizados nos experimentos para classificação de documentos.	86
Tabela 6.4 Visão geral das listas de <i>stopwords</i> criadas para calibragem das configurações iniciais dos experimentos.....	87
Tabela 6.5 Resultados obtidos durante a calibragem de <i>stopwords</i>	87
Tabela 6.6 Tabela comparativa do F1-score entre os modelos NB, SVM, RF em diferentes configurações de número de <i>features</i> e <i>datasets</i> construídos.....	90
Tabela 6.7 Tabela comparativa do tempo de execução nos algoritmos NB, SVM, e RF para diferentes número de <i>features</i> no <i>dataset 01</i> construído.	94

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Motivação	14
1.2 Objetivos e contribuições	16
1.3 Organização do texto	17
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 LegalTech	18
2.2 Inteligência Artificial	19
2.2.1 Naive Bayes (NB)	20
2.2.2 Support Vector Machine (SVM)	21
2.2.3 Random Forest (RF)	22
2.2.4 Named Entity Recognition (NER)	23
2.2.5 Bag-of-words (BOW) e Term Frequency - Inverse Document Frequency (TF-IDF)	24
2.3 Considerações finais	24
3 REVISÃO SISTEMÁTICA DA LITERATURA: GERAÇÃO DE TEXTO LEGAL	26
3.1 Metodologia	26
3.1.1 Questão de pesquisa (QP)	27
3.1.2 Processo de seleção dos estudos	27
3.1.3 Base de dados e estratégia de busca	28
3.1.4 Critérios de exclusão	29
3.1.5 Critérios de inclusão	29
3.1.6 Extração de dados	30
3.2 Classificação dos estudos	37
3.2.1 Argumentação	38
3.2.2 Extração	39
3.2.3 Geração	40
3.2.4 Iiação	41
3.2.5 Recuperação	43
3.3 Resultados e discussão	45
3.3.1 QP1 (principal)	46
3.3.2 QP1.1 - Com que finalidade a tecnologia é aplicada na área jurídica?	47
3.3.3 QP1.2 - Quais tecnologias são mais populares na área jurídica?	48
3.3.4 QP1.3 - A geração de texto é automática, semi-automática ou manual?	48
3.3.5 QP1.4 - Quão ativa é a pesquisa sobre geração de texto legal?	51
3.3.6 QP1.5 - Quais são as limitações da pesquisa atual?	56
3.4 Considerações finais	57
4 TRABALHOS RELACIONADOS	59
4.1 Datasets	59
4.2 Trabalhos brasileiros de <i>LegalTech</i>	62
4.3 Considerações finais	63
5 UMA ABORDAGEM AGNÓSTICA PARA GERAÇÃO DE TEXTOS LEGAIS	64
5.1 Visão geral	64
5.2 Extração	65
5.3 Pré-processamento	66
5.4 Análise e Modelagem	68
5.5 Exibição	69

6 EXPERIMENTOS E ANÁLISE DOS RESULTADOS	70
6.1 LEAF: um protótipo de framework agnóstico legal	70
6.1.1 Visão geral	70
6.1.2 Anotação de texto	76
6.1.3 Classificação de documentos	77
6.2 Análise Descritiva	78
6.3 Experimentos na classificação de documentos	83
6.3.1 Calibragem inicial e configuração escolhida	84
6.3.1.1 Datasets	85
6.3.1.2 <i>Stopwords</i>	86
6.3.1.3 Calibragem de <i>stopwords</i>	87
6.3.1.4 Calibragem dos algoritmos	88
6.3.1.5 Número de <i>features</i>	88
6.3.1.6 Considerações finais sobre as simulações	89
6.3.2 Análise F1-score	89
6.3.3 Análise e resultados sobre tempo de execução	93
6.3.4 Panorama geral dos resultados	96
6.3.5 Discussão geral dos resultados	97
7 CONCLUSÃO	99
7.1 Limitações	100
7.2 Direções futuras	101
REFERÊNCIAS	103

1 INTRODUÇÃO

Linguagem e direito sempre tiveram uma relação próxima, por isso, é esperado que áreas da Ciência da Computação que operam em linguagem natural possam ser aplicadas em determinadas áreas da advocacia. Esta relação é referida como *LegalTech*. A *LegalTech* tem uma vasta gama de áreas de aplicação para ajudar escritórios de advocacia e organizações com aspectos relacionados à desmaterialização de serviços jurídicos de texto e papel para formato digital. Nesse sentido, a *LegalTech* se beneficia dos serviços jurídicos aumentando a eficiência, produtividade e crescimento, reduzindo custos, melhorando resultados para clientes e organizações (MUNISAMI, 2020a).

Para melhorar o gerenciamento e a produtividade de escritórios de advocacia e profissionais da área jurídica, as áreas comuns de aplicação do *LegalTech* são gerenciamento de práticas, risco e conformidade, pesquisa/análise jurídica, análise/gerenciamento de contratos, propriedade intelectual, tecnologia de descoberta eletrônica, direito DIY (*Do It Yourself* - Faça Você Mesmo), documentos jurídicos como serviço, *marketplaces*, faturamento eletrônico e resolução de disputas *online* (HARPER; ZHANG, 2021). Para dar suporte às áreas de aplicação da *LegalTech*, é necessário a utilização de técnicas mais avançadas, entre as quais se destacam Inteligência Artificial (IA) e Processamento de Linguagem Natural (PLN) (SALMERÓN-MANZANO, 2021a). A IA é uma área da Ciência da Computação que trata da simulação de comportamento inteligente em computadores (MCCORDUCK, 2004). Já a PLN explora como as máquinas podem ser usadas para entender e controlar texto ou fala em linguagem natural para realizar tarefas desejadas (CHOWDHURY, 2005).

Os primeiros sistemas de busca de conteúdo jurídico *online* surgiram na década de 1960 e sistemas especialistas jurídicos foram um tema de discussão no final da década de 1970 (ENNALS, 1989). Como há um número crescente de repositórios de dados de texto jurídico legíveis por máquina digitalizados, com potencial para aumentar a eficácia dos serviços jurídicos e reduzir custos (CHOWDHURY, 2005), a *LegalTech* recebeu atenção considerável nos últimos anos (DALE, 2019).

A área jurídica no Brasil não é diferente e a Justiça segue a tendência de abertura para as novas tendências e possibilidades de inovação, conforme apontado pelo Poder Judiciário em 2006, quando foi instituída a Lei do Processo Eletrônico¹, abrindo caminho para a substituição dos processos físicos. O Sistema Judicial Brasileiro é atualmente o

¹<<https://justicadigital.com/lei-processo-eletronico>>

maior sistema judiciário do mundo e recebe um número extremamente elevado de processos judiciais todos os dias (SILVA; BRAZ; CAMPOS, 2018).

Conforme o relatório Justiça em Números 2022² (ano-base 2021), feito pelo Conselho Nacional de Justiça (CNJ), 97,2% dos novos processos ingressaram na Justiça em formato eletrônico em 2021, o que equivale a 27 milhões de casos novos ingressados por meio virtual, totalizando 182,7 milhões de processos ingressados na Justiça em formato eletrônico nos últimos 13 anos. A digitalização de processos também resulta em maior eficiência na resolução definitiva de ações judiciais em tramitação e na redução do tempo médio dos processos para quase um terço em relação aos processos físicos.

O compromisso de defender a ordem jurídica e a cidadania gera reflexos na rotina, principalmente no que diz respeito ao volume de processos. Naturalmente, com demandas cada vez maiores, o segmento jurídico necessita de meios eficazes para cumprir com suas obrigações e garantir um ambiente social seguro à população. Portanto, para a fidedignidade da *LegalTech*, é essencial desenvolver métodos que suportem linguagem e discurso jurídicos, explorem fontes jurídicas, possibilitem sistemas jurídicos transparentes e interoperáveis, e automatizem a gestão de processos (CHOWDHURY, 2005).

Em virtude da dependência de áreas como IA e PLN, alguns desafios e limitações da *LegalTech* estão vinculados a estas tecnologias (SALMERÓN-MANZANO, 2021a). Por exemplo, uma redução nas diferentes abordagens para resolver problemas recorrentes no campo de PLN (OMAR et al., 2022) pode dificultar a abordagem em tarefas de *LegalTech*. Além disso, por outro lado, as ferramentas de tecnologia jurídica também podem apresentar riscos, principalmente devido aos vieses perpetuados pelos algoritmos. A falta de parcialidade nas aplicações de IA foi repetidamente demonstrada em pesquisas anteriores, garantindo que as questões éticas em torno da tecnologia legal sejam totalmente consideradas (FEUERRIEGEL; DOLATA; SCHWABE, 2020). Além disso, outro desafio para a *LegalTech* é ajudar a projetar informações jurídicas de forma acessível a profissionais não jurídicos (NEWMAN; DOHERTY, 2008).

1.1 Motivação

A *LegalTech* exige técnicas modernas para uma ampla gama de problemas da área jurídica. A fim de analisar o estado-da-arte e identificar os avanços acadêmicos na li-

²<<https://www.cnj.jus.br/wp-content/uploads/2022/09/sumario-executivo-jn-v3-2022-2022-09-15.pdf>>

teratura, foram encontradas revisões sistemáticas de literatura sobre gerenciamento de documentos legais (MARKOVIĆ; GOSTOJIĆ, 2019), recuperação de informação legal (MOENS, 2004), *benchmark* de algoritmos (VU; NGUYEN, 2019; KANAPALA; PAL; PAMULA, 2019). Assim, o escopo dos trabalhos são bastante específicos, destinados a sub-áreas de *LegalTech* apenas. Para uma visão geral da área, dois trabalhos tiveram destaque: uma retrospectiva de vinte e cinco anos da *International Conference on AI and Law* (BENCH-CAPON et al., 2012) e uma *survey* sobre como PLN é utilizada em *LegalTech* (DALE, 2019). Em ambos os trabalhos, os autores se baseiam em uma análise de áreas da Ciência da Computação aplicadas em *LegalTech*. Entretanto, até onde sabemos, nenhuma revisão ou mapeamento sistemático foi feito para o caminho inverso, isto é, uma análise de quais áreas da Ciência da Computação são utilizadas ao para dar suporte a tarefas gerais de *LegalTech*.

A carência de uma revisão para este tópico foi o ponto de partida para delimitar o escopo de pesquisa desta dissertação. Posteriormente, foi adicionado ao contexto da revisão a condição de que o objetivo dos trabalhos fosse a geração de texto legal. A análise de 146 artigos após a realização de uma revisão sistemática possibilitou um entendimento tanto dos objetivos inerentes à tarefa de geração de texto legal para *LegalTech* quanto às áreas da Ciência da Computação envolvidas. Dessa forma, devido ao seu escopo mais amplo, a revisão realizada se diferencia das demais revisões mencionadas.

A análise dos trabalhos durante a revisão sistemática também identificou duas oportunidades interessantes referentes à *metodologia de desenvolvimento das soluções propostas (i)* e *coleções de informações de documentos legais organizadas em datasets (ii)*. Embora cada trabalho possua um objetivo em alguma etapa da geração de texto legal (técnicas para recuperar informações, criar argumentações, realizar predições), é extremamente comum analisar trabalhos sem demonstração clara de como a solução foi implementada ou de como os dados foram coletados e processados antes da aplicação de técnicas. A ausência dessas informações prejudica o avanço da pesquisa na área, uma vez que trabalhos futuros ficam limitados a informações escassas, retrabalho e ausência de padrões que podem ser aperfeiçoados através de melhoria constante com disciplinas como engenharia de *software* (DALE, 2019).

1.2 Objetivos e contribuições

O objetivo principal deste trabalho é a elaboração de uma abordagem com a capacidade de fornecer recomendações no desenvolvimento de soluções na área de *LegalTech* de forma coesa, estruturada, extensível, e robusta. Para tal, é proposta uma Abordagem Agnóstica composta por etapas de coleta, processamento, modelagem e exibição de resultados a partir da utilização de técnicas de IA em documentos legais. O conceito agnóstico é utilizado para definir uma abordagem genérica, mas com condições e potencial para ser adaptada ao escopo dos trabalhos de interesse, ganhando assim um propósito. Cada etapa possui um módulo de comunicação que viabiliza a abstração e independência de tarefas que compõem o trabalho de interesse.

Para avaliar a viabilidade da abordagem agnóstica, é construído um protótipo LEAF (Legal Agnostic Framework) que, através de diversos serviços, instancia duas soluções para *LegalTech*: geração de texto legal e classificação de documentos legais.

Para a geração de texto legal, o protótipo utiliza o modelo LeNER-Br (ARAÚJO et al., 2018) para classificação de entidades no texto e aplica anotação textual para expandir o trabalho relacionado.

Para a classificação de documentos, são realizadas diversas simulações para explorar diferentes calibrações e configurações de *stopwords*, *datasets*, *features*, e parâmetros de modelos classificadores. Os experimentos foram realizados em uma coleção de acórdãos extraídos do site do Supremo Tribunal Federal (STF), onde todas as informações das páginas *Web* e dos textos de arquivos PDFs são organizadas em um *dataset* estruturado construído para este trabalho e disponibilizado publicamente. Com base nas simulações, são destacados os desempenhos em métricas como F1-score e tempo de execução para os classificadores *Naive Bayes* (NB), *Support Vector Machines* (SVM), e *Random Forest* (RF). Em especial, destacam-se os classificadores SVM e RF com F1-score 96% e 95,83%, respectivamente, superando os valores de referência de trabalhos relacionados utilizados.

As contribuições deste trabalho são:

- Uma revisão sistemática da literatura para identificar o impacto da IA na área jurídica e, em especial, nos estudos sobre geração de texto legal;
- Uma taxonomia, dividida em 5 categorias, a partir da revisão sistemática, para tarefas durante a geração de texto legal na área de *LegalTech*;
- Um *dataset* disponibilizado publicamente com informações estruturadas a partir de

acórdãos de páginas do STF brasileiro desde 1892 à agosto de 2022;

- Uma coleção, também disponibilizada publicamente, com todos os acórdãos extraídos das páginas do STF brasileiro em formato PDF e TXT (após extração do texto do arquivo PDF);
- Uma Abordagem Agnóstica para realização de tarefas da área de *LegalTech*;
- Um protótipo modularizado, expansível e reusável com serviços para extração, pré-processamento, análise e modelagem, e visualização de informações para soluções da área de *LegalTech*;
- Uma demonstração de serviço de anotação de texto, em fase inicial, para apresentar oportunidades de avanço na tarefa de geração de texto legal;
- Três modelos classificadores, com diferentes propriedades, para a tarefa de classificação de documentos legais.

1.3 Organização do texto

Esta dissertação está estruturada como segue. O Capítulo 2 apresenta a fundamentação teórica necessária para este trabalho. O Capítulo 3 descreve a *survey* realizada sobre geração de texto legal. O Capítulo 4 descreve e compara os trabalhos relacionados ao escopo principal desta dissertação. O Capítulo 5 apresenta a Abordagem Agnóstica proposta. O Capítulo 6 descreve o protótipo implementado, os experimentos realizados com este, e os resultados obtidos. Finalmente, o Capítulo 7 contém as conclusões e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda conceitos e algoritmos fundamentais utilizados nesta dissertação. Inicialmente é apresentada a área de *LegalTech*, discorrendo sobre sua importância nos domínios jurídico e da Ciência da Computação. Em seguida, são discutidos conceitos de Inteligência Artificial (IA). Por fim, é realizado um resumo dos principais algoritmos de Aprendizado de Máquina e Processamento de Linguagem Natural (PLN) deste trabalho.

2.1 LegalTech

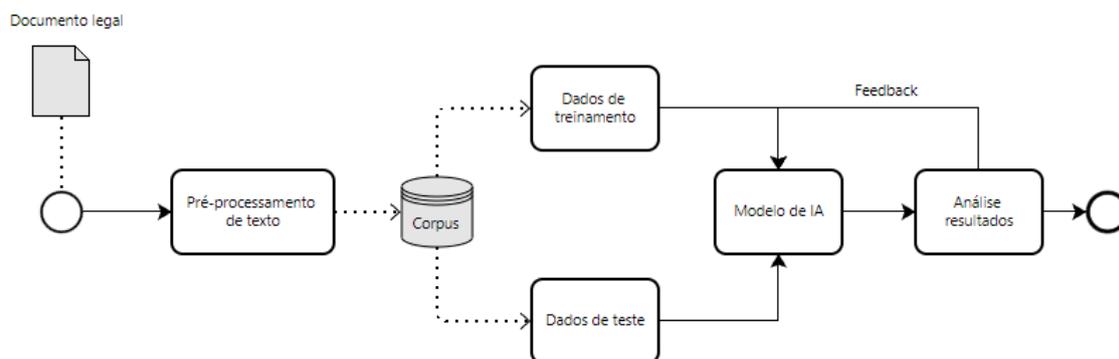
LegalTech é um acrônimo para *Legal Technology* (Tecnologia Legal), que se refere a tecnologias da Ciência da Computação que prestam serviços jurídicos (MUNISAMI, 2020b). Os serviços jurídicos podem se beneficiar de Inteligência Artificial, Linguagem Natural e dados estruturados para estimular a comunicação remota no setor jurídico, acelerar os procedimentos e a gestão das tarefas dos próprios advogados, e simplificar e modificar a forma de contato entre os profissionais jurídicos e as partes leigas interessadas (SALMERÓN-MANZANO, 2021b).

Como quase todas as leis são expressas em linguagem natural, a PLN desempenha um grande papel na compreensão e previsão da lei. Os avanços nos métodos de PLN criam oportunidades para a construção de novas abordagens para melhorar a eficiência, compreensão e consistência dos serviços jurídicos e sistemas jurídicos (FILTZ; KIRRANE; POLLERES, 2021). Quando combinada com técnicas de IA, a PLN fornece ferramentas para construir modelos preditivos que auxiliam em tarefas de geração de texto legal, previsão de sentenças, classificação de documentos legais, extração de informação e conhecimento em textos legais, construção de argumentação em processos, tradução de termos jurídicos para pessoas leigas (NAZARENKO; WYNER, 2017; CASANOVAS et al., 2016).

A Figura 2.1 mostra um resumo de como as abordagens de IA e PLN podem ser aplicadas juntas para realizar as tarefas de *LegalTech*. Uma vez digitalizados, os documentos legais são submetidos a abordagens de PLN para pré-processar o conteúdo de seus textos de acordo com os requisitos do algoritmo de IA. O algoritmo de IA, por sua vez, mapeia as entradas do modelo e as saídas paralelas dos dados disponíveis usando abordagens para o aprendizado. Em seguida, o algoritmo de IA fornece repetidamente os

exemplos de entradas e saídas do modelo e seleciona o modelo certo que atinge os resultados desejados, ajustando os recursos disponíveis e os dados de treinamento, construindo um modelo capaz de prever resultados (AISHATH et al., 2019).

Figura 2.1: Visão geral do papel dos campos de IA e PLN na área de *Legal Tech*.



Fonte: o autor

2.2 Inteligência Artificial

A IA é o campo da Ciência da Computação que lida com a simulação do comportamento inteligente em computadores (MCCORDUCK, 2004). A IA abrange uma enorme variedade de subcampos com características em comum, como a capacidade de raciocínio, aprendizagem, reconhecimento de padrões, e inferência.

Machine learning (Aprendizado de máquina) (ML) é uma subcampo da IA que estuda o reconhecimento de padrões através dos dados para deferir previsões mais eficazes (ÇELIK, 2018). Os algoritmos de ML podem ser divididos em 4 categorias:

- Aprendizado supervisionado;
- Aprendizado não supervisionado;
- Aprendizado semi-supervisionado;
- Aprendizado por reforço.

Aprendizado supervisionado ocorre quando o modelo aprende a partir de resultados predefinidos. Este aprendizado utiliza valores passados de referência para supervisionar e aprender quais devem ser seus resultados de saída, aperfeiçoando suas previsões a partir de erros anteriores. Os algoritmos com aprendizado supervisionado são divididos em duas subcategorias: *classificação* e *regressão*. Algoritmos da subcategoria classificação predizem o resultado da classe de saída de acordo com o conjunto de classes de

entrada. Por outro lado, os algoritmos de regressão identificam uma relação funcional entre os parâmetros de entrada e a saída.

Aprendizado não supervisionado é quando não há resultados predefinidos de referência para a aprendizagem do modelo. O modelo, então, encontra e agrupa sozinho os dados fornecidos de acordo com os padrões detectados. O método mais comum em aprendizados não supervisionados é o *Clustering* (Agrupamento) cujos dados não conhecidos, mas semelhantes entre si, são agrupados em diferentes *clusters*.

Aprendizado semi-supervisionado é um caso especial de algoritmo supervisionado utilizado quando existem dados rotulados para as classes de entrada; porém, existe um número maior de dados não rotulados. Dessa forma, o aprendizado supervisionado realiza previsões com dados inadequados em seu treinamento, uma vez que não há relação com a saída esperada.

Aprendizado por reforço: algoritmos desta categoria utilizam um sistema de recompensa. Com *checkpoints* de partida e fim, o objetivo é encontrar caminhos mais curtos e eficientes. Se o caminho percorrido estiver correto, o algoritmo recebe recompensas positivas, caso contrário, recompensas negativas.

Na elaboração e implementação dos experimentos deste trabalho, foram utilizados algoritmos de aprendizado supervisionado de classificação. O restante desta seção descreve os algoritmos utilizados.

2.2.1 Naive Bayes (NB)

Classificadores Naïve Bayes, também chamado de Bayesianos, pertencem a uma categoria de classificação estatística. Estes classificadores exigem uma pequena quantidade de dados de treinamento para estimar os parâmetros necessários, e podem ser extremamente rápidos em comparação a métodos mais sofisticados (CHEN et al., 2021). O classificador NB é baseado no teorema probabilístico de Bayes que calcula a probabilidade de um evento aleatório A ocorrer se um outro evento relacionado B ocorrer. Em geral, na classificação de NB, visa-se calcular os valores de probabilidade dos efeitos de cada critério no resultado com a suposição “ingênua” de independência condicional entre cada par de características dado o valor da variável de classe. Apesar da simplificação de suposições, os classificadores NB possuem desempenho satisfatório em tarefas como classificação de documentos e filtragem de spam (ZHANG; GAO, 2011).

Dado o vetor de variáveis de *features* $\langle F_i \rangle$, $i = 1, 2, \dots, n$, e o conjunto de classes

predefinidas $C = c_1, c_2, \dots, c_k$, para a variável de *feature* $F_j, j = 1, 2, \dots, n$, o teorema de Bayes estabelece a seguinte relação:

$$P(F_j|C) = \frac{P(F_j)P(C|F_j)}{P(C)}$$

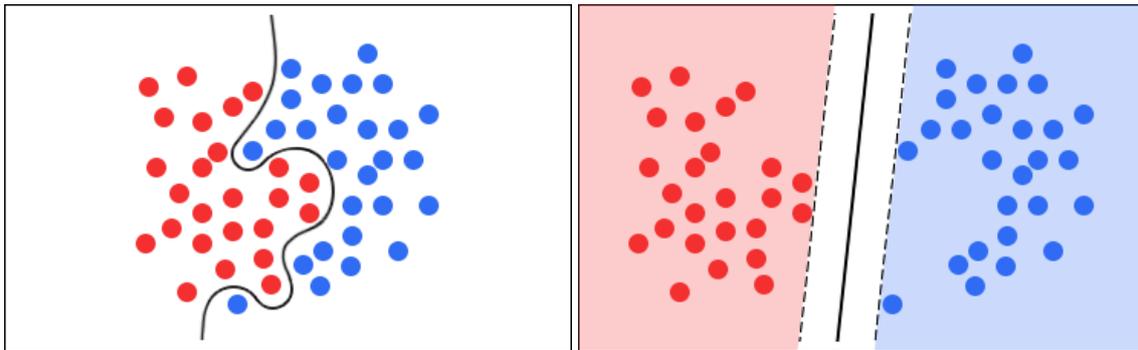
Existem diferentes configurações do algoritmo NB para determinados conjuntos de dados. Neste trabalho, foi utilizada a distribuição Multinomial. Esta distribuição é frequentemente utilizada em tarefas de classificação de texto, pois os dados são representados como contagens de vetores de palavras. A distribuição é parametrizada por vetores para cada classe C , onde n é o número de *features*, isto é, o tamanho do vocabulário, $P(F_j|C)$ é a probabilidade da *feature* j aparecer em uma amostra pertencente à classe C .

2.2.2 Support Vector Machine (SVM)

O algoritmo de *Support Vector Machines* (Máquina de Vetores de Suporte) é um classificador supervisionado não probabilístico binário. O classificador faz uso do conceito de hiperplanos de decisão que define os limites de decisão, separando grupos de objetos do conjunto de entrada em diferentes classes de dados. Desta forma, o hiperplano visa maximizar a margem entre os pontos mais próximos à cada classe de dados e, conseqüentemente, minimizar o erro em predições do conjunto de dados (EVGENIOU; PONTIL, 2001).

A Figura 2.2 mostra um hiperplano com um conjunto de entrada de duas classes (vermelha e azul), e uma linha definindo o limite. Os dados de entrada originais (Figura 2.2(a)) são mapeados ou rearranjados usando uma função matemática (também denominada *kernel*). Após esta transformação, os dados mapeados ficam linearmente separáveis, isto é, estruturas complexas com curvas para separação são evitadas (Figura 2.2(b)).

Figura 2.2: Representação de um hiperplano de SVM.



(a) Dados de entrada originais

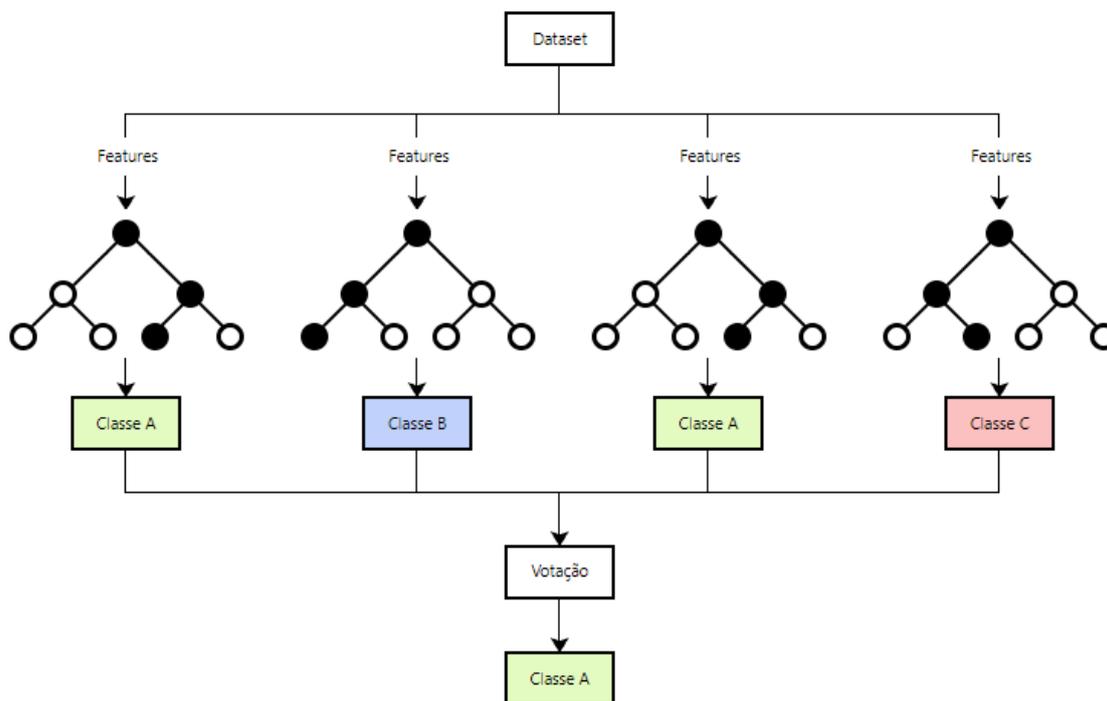
(b) Dados de entrada separados após mapeamento

Fonte: o autor, adaptado de (DEVIKA; SUNITHA; GANESH, 2016)

2.2.3 Random Forest (RF)

Random Forest (Floresta Aleatória) é uma classe de aprendizado supervisionado conjunto baseado em árvores de decisão. O algoritmo RF é composto por uma coleção de várias árvores sucessivas independentes criadas a partir do conjunto de dados inicial. Cada nó interno da árvore de decisão especifica uma condição em um atributo e a ramificação é feita com base no resultado. O nó folha da árvore possui um rótulo de classe obtido após o cálculo de todos os atributos anteriores. Ao final desta etapa, para cada rótulo de classe obtida em uma árvore de decisão, uma votação majoritária é realizada, isto é, o rótulo de classe mais frequente produzirá a classe prevista como resultado (CHEN et al., 2022; MASHAYEKHI; GRAS, 2015). A Figura 2.3 apresenta uma visão geral do algoritmo RF.

Figura 2.3: Visão geral sobre árvores de decisão em função da votação majoritária no algoritmo RF.



Fonte: o autor, adaptado de (IBM, 2022)

2.2.4 Named Entity Recognition (NER)

Named Entity Recognition (Reconhecimento de Entidade Nomeada) é uma área de PLN cuja tarefa principal consiste de duas etapas: (1) detectar entidades de um texto em linguagem natural, (2) classificar cada entidade em uma categoria (YADAV; BETHARD, 2019). Este objetivo é alcançado através da análise sintática do texto de acordo com a classe gramatical da palavra. As entidades identificadas podem ser classificadas de diferentes formas; por exemplo: Pessoa, Local, Organização, Tempo, Julgamento, Legislação. Soluções baseadas em NER são frequentemente utilizadas em sistemas de recuperação de informação, sistemas de perguntas e respostas, modelagem de tópicos, redes neurais, e aprendizado de máquina (LI, 2022). A Figura 2.4 exemplifica a identificação de entidades em uma frase extraída de um acórdão do STF: as palavras "Turma", "recurso ordinário", e "Luiz Fux" pertencem às entidades Organização, Jurisprudência, e Pessoa, respectivamente.

Figura 2.4: Exemplo de identificação e classificação de entidades em uma frase através de NER.

Por unanimidade, a Turma ^{Organização} negou provimento ao recurso ordinário ^{Jurisprudência}, nos termos do voto do Relator Senhor Luiz Fux ^{Pessoa}.

Fonte: o autor

2.2.5 Bag-of-words (BOW) e Term Frequency - Inverse Document Frequency (TF-IDF)

Embora não sejam algoritmos propriamente da área de IA, tanto BOW quanto TF-IDF desempenharam papel fundamental em diferentes etapas na aplicação dos algoritmos apresentados nas Seções anteriores deste capítulo.

O BOW é um modelo de representação simplificada utilizada nas áreas de PLN e recuperação de informações para classificação de objetos. Na classificação de texto, este modelo é expresso através de um conjunto de palavras, desconsiderando a gramática das palavras. O conjunto de palavras é obtido na coleção (uma frase ou documentos) onde, para cada palavra, uma tupla é associada a esta palavra e a sua respectiva frequência na coleção (K; JOSEPH, 2014).

O valor TF-IDF é uma medida estatística que indica a importância de uma palavra em um documento em relação à coleção de documentos. Este valor é diretamente proporcional ao número de ocorrências da palavras no documento e inversamente proporcional à frequência desta palavra na coleção inteira. Portanto, o valor TF-IDF é frequentemente utilizado como fator de ponderação na recuperação de informações e na mineração de dados, e também, mais especificamente, na recuperação de informações jurídicas (WEHNERT et al., 2021), uma vez que estas possuem coleções amplas de documentos com elevada quantidade de texto.

2.3 Considerações finais

Neste capítulo foi descrita a fundamentação teórica para a execução deste trabalho. Inicialmente, apresentou-se os conceitos da área multidisciplinar de *LegalTech* que abrange o domínio deste trabalho e é ponto inicial e fundamental da abordagem proposta. Posteriormente, direcionou-se os conceitos para o campo da IA, que está em ampla convergência com a Ciência da Computação e *LegalTech*. A partir da definição de conceitos

gerais de IA e aprendizado de máquina, foram discutidos os principais algoritmos de IA e PLN que compõem os experimentos aplicados nesta dissertação: modelos de classificação baseados em NB, SVM, e RF, a área de NER, e os algoritmos auxiliares de BOW e TF-IDF.

3 REVISÃO SISTEMÁTICA DA LITERATURA: GERAÇÃO DE TEXTO LEGAL

Este capítulo apresenta a revisão sistemática da literatura que foi realizada para identificar o impacto da IA na área jurídica, e, em especial, os estudos publicados sobre geração de texto legal durante o período entre janeiro de 1963 a dezembro de 2020. Nosso objetivo com esta revisão sistemática da literatura foi identificar e apresentar um panorama sobre diferentes soluções propostas na geração de texto legal, visando melhorar os avanços nas áreas de *LegalTech* e, amplamente, tirar proveito de técnicas de linguagem natural.

Esta revisão sistemática fornece um relatório sobre a relação entre diferentes áreas do campo da Ciência da Computação, como Inteligência Artificial (IA), Recuperação de Informações (RI), Processamento de Linguagem Natural (PLN), e Representação de Conhecimento (RC), aplicadas à *LegalTech* para alcançar objetivos variados em documentos legais. Para tal, realizamos estudos baseados na observação de semelhanças entre os artigos selecionados quanto ao seu escopo principal. Buscamos responder a perguntas como *quais tecnologias estão sendo usadas para gerar textos legais e qual a finalidade da tecnologia aplicada?* Até onde sabemos, nenhuma revisão sistemática da literatura nem mapeamento sistemático foi feito para investigar este tópico de pesquisa.

Este capítulo reproduz o artigo submetido (FLESCH; GALANTE, 2022). Na Seção 3.1, detalhamos a metodologia empregada para a realização da revisão sistemática da literatura. Na Seção 3.2, propomos uma taxonomia pra classificação dos estudos em cinco categorias de acordo com a análise do escopo principal. Na Seção 3.3, apresentamos a análise dos estudos. Por fim, na Seção 3.4, analisamos e discutimos os resultados.

3.1 Metodologia

Seguindo protocolos específicos de acordo com Kitchenham e Charters (BA; CHARTERS, 2007), esta revisão sistemática da literatura tem como objetivo sintetizar pesquisas sobre geração de textos jurídicos a fim de identificar, selecionar, interpretar e classificar as obras na área jurídica. A revisão sistemática apresenta uma metodologia robusta e destaca tanto as limitações dos artigos impostos quanto da própria revisão.

O início da nossa revisão sistemática da literatura foi precedido por uma etapa preliminar que nos forneceu uma ideia ampla da variedade de estudos convergentes entre as áreas de IA e jurídica. Nós direcionamos a pesquisa para investigar estudos que

objetivavam, de algum modo, gerar texto legal.

A nossa revisão sistemática está composta pelas seguintes etapas: seleção das bases de dados bibliográficos, a definição das estratégias de busca, do processo de seleção de artigos científicos, dos critérios de inclusão e exclusão, do processo de análise de cada artigo. Apresentamos detalhes do nosso protocolo de revisão nas subseções a seguir.

3.1.1 Questão de pesquisa (QP)

Definimos a questão de pesquisa principal da seguinte forma:

- **QP1 (primária):** Quais tecnologias estão sendo usadas para gerar textos legais?

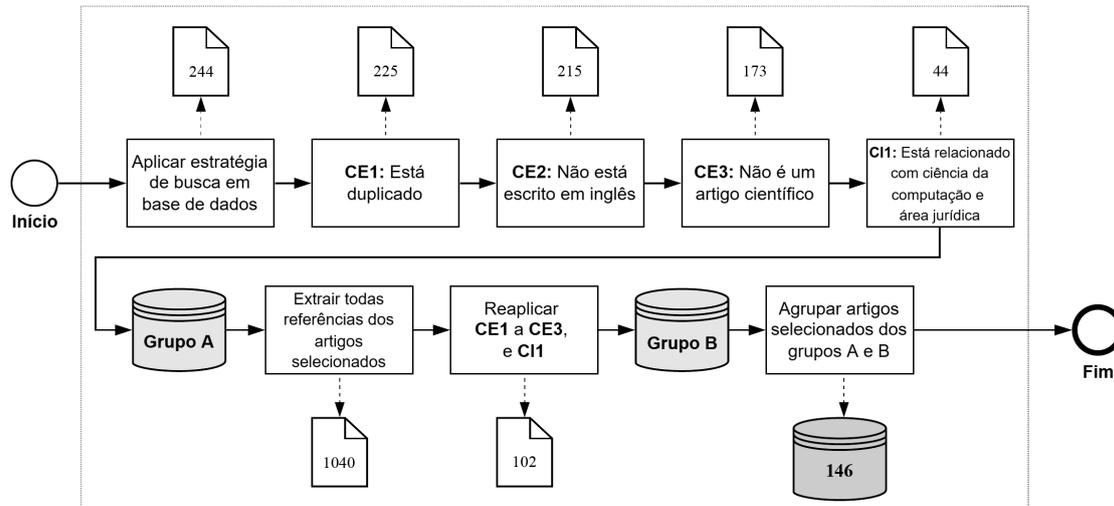
Para ajudar a obter dados e resumir diferentes aspectos do tema em estudo, bem como identificar lacunas na pesquisa atual, definimos quatro QPs secundárias. Essas QPs permitem identificar o que está sendo investigado mais detalhadamente, qual aspecto está faltando no conjunto atual de publicações, a frequência com que o tema tem sido abordado nas publicações e quem são os principais autores que publicam sobre o assunto. As QPs secundárias também orientaram a definição de alguns critérios de exclusão e inclusão como veremos nas próximas seções.

- **QP1.1:** Com que finalidade a tecnologia é aplicada na área jurídica?
- **QP1.2:** Quais tecnologias são mais populares na área jurídica?
- **QP1.3:** A geração de texto é automática, semi-automática ou manual?
- **QP1.4:** Quão ativa é a pesquisa sobre geração de texto legal?
- **QP1.5:** Quais são as limitações da pesquisa atual?

3.1.2 Processo de seleção dos estudos

Inicialmente, realizamos uma pesquisa preliminar, uma vez que a coleta de artigos devolvidos ajudou a definir a estratégia de pesquisa. Posteriormente, aplicamos a estratégia de pesquisa às fontes de pesquisa e, com os artigos selecionados, os critérios de exclusão e inclusão. A Figura 3.1 apresenta o conjunto de etapas aplicadas e o número de artigos selecionados respectivamente. A aplicação de cada etapa do processo de seleção de estudos será explicada com mais detalhes nas Seções 3.1.3–3.1.6.

Figura 3.1: Processo de seleção de estudos e quantidade de artigos extraídos após cada etapa.



Fonte: o autor

3.1.3 Base de dados e estratégia de busca

Cinco fontes relevantes foram identificadas para uso em nossa revisão de literatura: ACM Digital Library ¹, IEEE Xplore ², SpringerLink ³, Science Direct ⁴ e Scopus ⁵. Após escolher as fontes de busca, a partir dos resultados preliminares e QPs definidos, definimos os campos de busca *título*, *resumo* e *palavras-chave*, e a estratégia de busca: (*legal AND text AND generation*).

A estratégia de busca foi escolhida a partir da análise prévia da cobertura dos trabalhos resultantes. *Strings* com termos genéricos da área de linguagem natural ou com disjunção de termos envolvendo as duas áreas de interesse (computação e direito), retornaram um conjunto bastante abrangente de trabalhos com pouca relevância para pesquisa. Da mesma forma, termos estritamente legais ou com técnicas específicas de linguagem natural (ou aprendizado de máquina) obtiveram um conjunto de resultados muito limitado. Assim, a combinação de termos legais e um termo amplo de linguagem natural retornou um número considerável de trabalhos. Por fim, ainda testamos diferentes combinações com sinônimos de termos da área jurídica e de computação para avaliar a quantidade e relevância de artigos encontrados nas bases de dados. Como o número de trabalhos retornados era similar, nesta etapa final de calibragem, a estratégia escolhida foi a que obteve

¹<<https://dl.acm.org/>>

²<<https://ieeexplore.ieee.org/>>

³<<https://link.springer.com/>>

⁴<<https://www.sciencedirect.com/>>

⁵<<https://www.scopus.com/>>

o maior número de resultados para investigação.

3.1.4 Critérios de exclusão

Nesta revisão, foram adotados os seguintes critérios de exclusão (CE):

- **CE1:** está duplicado
- **CE2:** não em inglês
- **CE3:** não é um artigo

Do conjunto total de resultados obtidos, o critério CE1 foi aplicado duas vezes. A primeira foi realizada automaticamente utilizando a ferramenta de revisão sistemática St4rt⁶ para remover trabalhos com os mesmos títulos e reduzir o escopo de análise. A segunda, manualmente, garantiu que as mesmas obras, mas com pequenas alterações e ambiguidades no título, fossem corretamente excluídas.

Os critérios CE2 e CE3 também foram aplicados manualmente. Os critérios CE2 e CE3 foram escolhidos para manter uma estrutura padrão de análise para os trabalhos identificados na busca realizada. Em especial, o critério CE2 foi escolhido uma vez que a vasta maioria dos trabalhos, nas bases de dados utilizadas, estava no idioma inglês e notoriamente nenhum trabalho em português estava presente. Desta forma, ficamos confiantes com a manutenção do CE2, já que o risco de trabalhos relevantes no idioma português serem desconsiderados foi reduzido.

Os estudos resultantes foram preparados para posterior aplicação dos critérios de inclusão e extração de dados. Após a aplicação dos critérios de exclusão, 173 artigos foram obtidos, conforme é mostrado na Figura 3.1.

3.1.5 Critérios de inclusão

Os critérios de inclusão (CI) foram aplicados aos artigos de resultados para garantir que os trabalhos estivessem relacionados ao tema desta revisão sistemática da literatura. O CI foi:

- **CI1:** o artigo está relacionado às áreas de computação e jurídica

Para o conjunto de artigos resultantes da aplicação dos CEs, para garantir a inte-

⁶<http://lapes.dc.ufscar.br/tools/start_tool>

gridade da revisão sistemática, foram analisados os títulos, resumos e, em alguns casos, as conclusões e eventuais seções que se fizeram necessárias. Assim, foi garantida a inclusão do trabalho selecionado caso atendesse ao CI definido. Ao todo, 44 artigos foram selecionados após a aplicação dessas etapas, conforme mostrado na Figura 3.1.

No entanto, para diminuir as chances de um trabalho com contribuição significativa ser excluído desta revisão, todos os 44 artigos selecionados foram agrupados no *Grupo A*. Em seguida, extraímos todos os trabalhos relacionados dos artigos deste grupo e reaplicamos os CEs e CI definidos. Ao final desta etapa, 102 novos trabalhos foram selecionados e agrupados no *Grupo B*. Unindo os dois grupos, todo o processo de seleção resultou em um conjunto de 146 artigos para passar pela fase de extração de dados e compor o conjunto resultante a ser analisado para a produção do relatório do estado da arte. A Figura 3.1 apresenta todos os trabalhos selecionados em cada uma das etapas explicadas.

3.1.6 Extração de dados

Com base em Peterson (PETERSEN; VAKKALANKA; KUZNIARZ, 2015), criamos uma tabela para extração de dados do nosso conjunto final de artigos. A tabela consiste em três colunas (item de dados, valor, QP) e cada linha é um conteúdo de dados extraído. A Tabela 3.1 apresenta o formulário utilizado para extrair os dados de cada artigo. “Item de dados” são os dados a serem extraídos; “Valor” recebe o resultado da extração e “QP” representa a questão de pesquisa que justifica a extração do respectivo “Item de dados”.

Como pode ser visto na Tabela 3.1, o item de dados "Identificador" representa um número único para referenciar o artigo selecionado em revisões posteriores ao lado desta pesquisa. Os itens de dados "Título", "Ano de publicação" e "Palavras-chave" foram extraídos das informações básicas do artigo. Por fim, registramos o "Escopo principal" dos artigos selecionados, quais as técnicas de computação apresentadas, se continham avaliação e/ou validação de resultados, e se geravam soluções automáticas, semiautomáticas ou manuais para a proposta do respectivo artigo.

Após a leitura de cada estudo selecionado, construímos a Tabela 3.2 classificando os artigos quanto ao seu escopo principal e uma coleção das áreas de Ciência da Computação mais citadas. O escopo principal é um dos itens de dados extraídos de cada artigo, conforme visto na Tabela 3.1. Para classificar o item de dados Escopo Principal, primeiramente, lemos as propostas do estudo: título, palavras-chave, resumo e, em alguns casos,

Tabela 3.1: Tabela de extração de dados.

Data Item	Valor	RQ
Identificador	Inteiro	
Título	Texto	
Ano da publicação	Inteiro	QP4
Palavras-chave	Texto	QP1, QP2
Escopo principal	Texto	QP1
Técnicas abordadas	Lista de áreas da Ciência da computação	QP2, QP5
Avaliação e validação	Binário [Sim/Não]	QP2, QP5
Tipo de geração	Ternário [Automático/Semiautomático/Manual]	QP3, QP5

Fonte: o autor

o texto completo. Em seguida, agrupamos repetidamente os estudos em uma abordagem mais ampla até obtermos uma segmentação precisa das categorias. As demais colunas da tabela apresentavam um valor binário para cada Item de dados: se o artigo mencionava alguma área de computação ou não. Por exemplo, se o artigo selecionado mencionou alguma abordagem de "PLN", a célula correspondente recebeu uma marca de "x". Caso contrário, foi deixado em branco. Com essa categorização, nosso objetivo foi responder às questões de pesquisa QP1 e QP2.

Por fim, observamos a frequência de palavras-chave nos artigos, para identificar quais foram as mais comuns entre os artigos selecionados. Para realizar esta tarefa, extraímos todas as palavras-chave dos artigos selecionados e removemos manualmente as *stopwords* e palavras muito genéricas (por exemplo, *software*, sistema, XML) ou que apareceram apenas uma vez. Em seguida, analisamos 3 cenários: palavras-chave exclusivas, unigramas e bigramas. Nosso objetivo foi, também, responder às questões de pesquisa QP2 e QP3.

Palavras-chave exclusivas são as palavras-chave originais do autor. As palavras-chave "*case-based reasoning*" e "*machine learning*" foram as mais recorrentes, aparecendo 8 e 7 vezes, respectivamente. N-gramas são sequências contíguas de n itens de uma dada amostra de texto ou fala. Assim, o unigrama é uma sequência de 1 item das palavras-chave do autor (por exemplo, *machine*, *learning*) e o bigrama é uma sequência de 2 itens. Os unigramas mais recorrentes foram *legal*, *text* e *language*, aparecendo 65, 22 e 20 vezes, respectivamente. Além disso, os bigramas mais recorrentes foram "*information retrieval*", "*natural language*", e "*language processing*", aparecendo 80, 18 e 14 vezes, respectivamente.

Tabela 3.2: Distribuição dos artigos por categoria dos 146 estudos selecionados para leitura completa.

Artigo	Escopo principal	PLN	RI	IA	RC
(Sannier et al., 2017); (YE et al., 2018); (HOVY, 1993); (SANNIER et al., 2015); (FARZINDAR; LAPALME, 2009); (BOUAYAD-AGHA et al., 2009); (HUANG et al., 2020); (SPROWL, 2006); (SPROWL; STAUD†, 1981); (SAXON, 2006); (MILLE; WANNER, 2008); (SHEREMETYEVA, 2003); (SHINMORI; MARUKAWA, 2003); (BRANTING, 1998)	Geração	x			
(SINH; NGUYEN; SATOH, 2019); (GOLTZ; DONDOLI; CAMERON-HUFF, 2018); (SPROWL et al., 1984); (SINGH; SHARMA, 2018); (PERIC et al., 2020-12)	Geração	x		x	
(NAVAS-LORO; SATOH; RODRÍGUEZ-DONCEL, 2019)	Geração	x			x
(ALSCHNER; SKOUGAREVSKIY, 2017)	Geração			x	
(MOCHALES; MOENS, 2009a); (ASHLEY; BRÜNINGHAUS, 2006)	Argumentação	x	x		
(ALEVEN, 2003); (ATKINSON; BENCH-CAPON, 2005); (BRANTING, 1991)	Argumentação	x			x

Tabela 3.2 – continuação da página anterior

Artigo	Escopo principal	PLN	RI	IA	RC
(MOCHALES; MOENS, 2009b); (ZHANG; KOPPAKA, 2007); (ASHLEY; WALKER, 2013)	Argumentação		x	x	
(MOCHALES; MOENS, 2008); (CHORLEY; BENCH-CAPON, 2005)	Argumentação			x	
(BENCH-CAPON; SARTOR, 2003); (BERMAN; HAF- NER, 1993); (FETERIS, 2015); (GRABMAIR; ASHLEY, 2011)	Argumentação				x
(ROSSI; WIRTH; KANOULAS, 2019)	Extração	x	x		
(METSKER; TROFIMOV; GRE- CHISHCHEVA, 2019)	Extração	x	x	x	
(LEITNER; REHM; SCHNEI- DER, 2019)	Extração	x		x	
(CUNNINGHAM et al., 2004); (ALLEN, 1998); (LEHMANN, 2007); (MOULIN; ROUSSEAU, 1990b); (MOULIN; ROUSSEAU, 1990a); (LAGOS et al., 2010)	Extração	x			x
(AGNOLONI; BACCI; OPIJ- NEN, 2017); (BOLIOLI; MER- CATALI; ROMANO, 2004); (OPIJNEN; SANTOS, 2017); (DEEDMAN; GELBART; CO- LEMAN, 1992)	Extração		x		

Tabela 3.2 – continuação da página anterior

Artigo	Escopo principal	PLN	RI	IA	RC
(ANDERSSON et al., 2016); (TURTLE, 1995); (DOZIER et al., 2007); (ROEGEST; HU- DEK; MCNULTY, 2018)	Extração		x	x	
(FILTZ et al., 2020)	Extração		x	x	x
(MOENS; UYTTENDA- ELE; DUMORTIER, 1999); (CIFUENTES-SILVA; GAYO, 2019); (SHAHAB; KIENINGER; DENGEL, 2010)	Extração		x		x
(CONRAD et al., 2005)	Extração			x	
(ALLEN; TURY, 2007); (KONS- TANTINO; SYKES; YANNO- POULOS, 1993)	Extração				x
(NAGEL, 1963)	Ilação	x			
(MOLES; DAYAL, 1993)	Ilação	x	x		
(CHITTA; HUDEK, 2019)	Ilação	x	x	x	
(ALETRAS et al., 2016); (GUO et al., 2020); (YAN et al., 2019); (MEDVEDEVA; VOLS; WIE- LING, 2020)	Ilação	x		x	
(MCDONALD; PUSTEJOVSKY, 1986); (ASHLEY; BR, 2003)	Ilação	x			x
(ZHOU et al., 2019); (HU et al., 2018)	Ilação		x	x	
(FAWEI et al., 2019); (GRE- ENLEAF; MOWBRAY; DIJK, 1995); (WALKER et al., 2017)	Ilação		x		x

Tabela 3.2 – continuação da página anterior

Artigo	Escopo principal	PLN	RI	IA	RC
(LUO et al., 2017a); (VERHEIJ, 2016); (BRÜNINGHAUS; ASHLEY, 2005); (GRABMAIR, 2017); (HE et al., 2018); (XU et al., 2019); (CHEN et al., 2019); (ZHONG et al., 2018)	Ilação			x	
(VöLKER; LANGA; SURE, 2008); (BREUKER; PETKOV; WINKELS, 2000); (BRÜNINGHAUS; ASHLEY, 2003)	Ilação				x
(BRÜNINGHAUS; ASHLEY, 2001); (JIA et al., 2014)	Recuperação	x			
(WAGH; ANAND, 2020); (FARZINDAR; LAPALME, 2004); (Yamada; Teufel; Tokunaga, 2017); (FALAKMASIR; ASHLEY, 2017); (GROVER et al., 2003); (QUINTANA; KAMEL; LO, 1992)	Recuperação	x	x		
(KIM; XU; GOEBEL, 2017); (MAXWELL; OBERLANDER; LAVRENKO, 2009); (TUGGENNER et al., 2020a); (FILTZ et al., 2019); (KIM, 2014); (WANG et al., 2018)	Recuperação	x	x	x	
(LAU; LAW; WIEDERHOLD, 2003); (WYNER; PETERS, 2010)	Recuperação	x	x		x

Tabela 3.2 – continuação da página anterior

Artigo	Escopo principal	PLN	RI	IA	RC
(HACHEY; GROVER, 2006); (WIDDISON; PRITCHARD; ROBINSON, 1992); (BRÜG- MANN et al., 2015)	Recuperação	x		x	
(GALGANI; COMPTON; HOFF- MANN, 2012a); (LEITH, 1986); (SHAHEEN; WOHLGENANNT; FILTZ, 2020)	Recuperação	x			x
(NAZARENKO; LEVY; WY- NER, 2018); (LOCKE; ZUC- CON; SCHELLS, 2017); (GAL- GANI; COMPTON; HOFF- MANN, 2012b); (TURTLE, 1994); (CHEN; LIU; HO, 2013); (LIU; CHEN; HO, 2015); (MOENS, 2006); (GELBART; SMITH, 1993); (KONIARIS; ANAGNOSTOPOULOS; VAS- SILIOU, 2017); (PEÑAS et al., 2010); (KUMAR et al., 2011); (XU; CROFT, 1998); (DANIELS; RISSLAND, 1997)	Recuperação		x		
(Merkl; Schweighofer, 1997); (NEJADGHOLI; TCHEMEUBE; WITHERSPOON, 2017); (BRÜ- NINGHAUS; ASHLEY, 1999); (DEANE, 2019); (GUO et al., 2019); (DO et al., 2017); (CHAL- KIDIS et al., 2019)	Recuperação		x	x	
(SARAVANAN; RAVINDRAN; RAMAN, 2006)	Recuperação		x		x

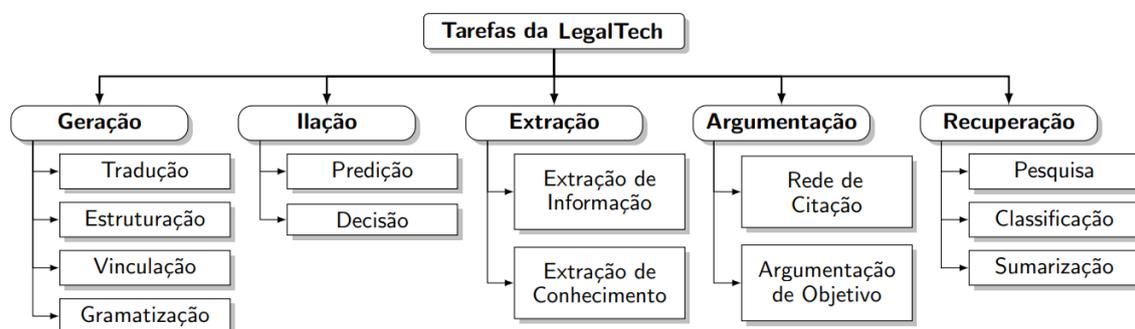
Tabela 3.2 – continuação da página anterior

Artigo	Escopo principal	PLN	RI	IA	RC
(HACHEY; GROVER, 2005); (GALGANI; HOFFMANN, 2011); (HAUSLADEN; SCHU- BERT; ASH, 2020)	Recuperação			x	
Total	39	65	61	53	34
%	-	44.52	41.78	36.3	23.29

3.2 Classificação dos estudos

Nós classificamos os 146 estudos selecionados em cinco categorias e desenvolvemos uma taxonomia, após observar semelhanças entre o escopo principal dos estudos quanto à geração de textos legais. A Figura 3.2 apresenta a taxonomia proposta para as *Tarefas de LegalTech*, que resume as principais tarefas durante a geração de texto legal na área jurídica de tecnologia: Argumentação (Seção 3.2.1), Extração (Seção 3.2.2), Geração (Seção 3.2.3), Ilação (Seção 3.2.4), e Recuperação (Seção 3.2.5). Em cada seção, discutimos as técnicas aplicadas para realizar a respectiva tarefa de acordo com os estudos selecionados.

Figura 3.2: Taxonomia de tarefas da LegalTech.



Fonte: o autor

3.2.1 Argumentação

Argumentação é o processo pelo qual os argumentos são construídos e tratados. Na geração de textos jurídicos, a tarefa de argumentação visa auxiliar os advogados na construção de argumentos jurídicos por meio de anotação, raciocínio e mineração de texto a partir de documentos jurídicos. Esta categoria inclui apenas 9,46% dos estudos selecionados.

A maioria dos estudos nesta categoria se concentra em diferentes abordagens para construir modelos de raciocínio jurídico, sendo o mais comum os de *Case-Based Reasoning* (Raciocínio Baseado em Casos) (ALEVEN, 2003; ASHLEY; BRÜNINGHAUS, 2006; ATKINSON; BENCH-CAPON, 2005; BENCH-CAPON; SARTOR, 2003; BERMAN; HAFNER, 1993; BRÜNINGHAUS; ASHLEY, 2003; FETERIS, 2015), seguido por *Semantic-Based Reasoning* (Raciocínio Baseado em Semânticas) (ZHANG; KOPAKA, 2007; BERMAN; HAFNER, 1993), e *Issue-Based Reasoning* (Raciocínio Baseado em Problemas) (ROEGEST; HUDEK; MCNULTY, 2018). Embora o raciocínio em si forneça uma ferramenta valiosa na tarefa de argumentação, é comum encontrar soluções com outras técnicas e algoritmos mesclados. Atkinson et al. (ASHLEY; BRÜNINGHAUS, 2006) aplicaram *Practical Reasoning* (Raciocínio Prático) para replicar as visões contrastantes envolvidas em decisões judiciais. Berman et al. (BENCH-CAPON; SARTOR, 2003) propuseram um modelo para representar os componentes teleológicos de decisões jurídicas para argumentação jurídica baseada em casos (*Case-based Legal Argument*). Alguns autores (ATKINSON; BENCH-CAPON, 2005; BRÜNINGHAUS; ASHLEY, 2003) utilizaram a construção da teoria para auxiliar no desenvolvimento e formalização do modelo de RC. Grabmair et al. (FETERIS, 2015) apresentaram um modelo de julgamento de valor que leva em conta a argumentação baseada em valor (*Value-Based Argumentation*).

A mineração de argumentação visa detectar os argumentos apresentados em um documento de texto, as relações entre eles e a estrutura interna de cada argumento individual. No estudo de Mochales e Moens (MOCHALES; MOENS, 2009b; MOCHALES; MOENS, 2009a), eles apresentaram diferentes técnicas para mineração de argumentação em aprendizado de máquina e gramáticas livres de contexto, e incentivaram mais pesquisas sobre gramáticas argumentativas. Mochales e Moens (MOCHALES; MOENS, 2008) também propuseram uma abordagem de argumentação em linguagem natural para detectar e estruturar argumentação em jurisprudência de linguagem natural.

Ashley e Walker (ASHLEY; WALKER, 2013) usaram anotações pressuposicionais e de texto para atribuir significado a informações relevantes à argumentação automaticamente a partir de um corpus de documentos de decisão legal. Em seguida, eles extraíram as informações e forneceram modelos de IA para construir novos argumentos.

3.2.2 Extração

Esta categoria corresponde a 16.21% dos trabalhos selecionados. A categoria de extração visa extrair informações de documentos legais. A informação extraída é conhecimento, para abordagens em RC, ou texto bruto para processamento posterior usualmente em sistemas de RI ou PLN.

Quando a extração de texto diz respeito ao texto bruto como saída de informação, alguns tópicos de RI são comuns em artigos selecionados. Por exemplo, Rossi, Wirth, Kanoulas, e Andersson et al. (ROSSI; WIRTH; KANOULAS, 2019; ANDERSSON et al., 2016) aplicaram a geração de consultas para auxiliar em questões de recuperação de patentes. Metsker et al. (METSKER; TROFIMOV; GRECHISHCHEVA, 2019) apresentaram a eficiência do fator de ponderação na recuperação de informações com TF-IDF ao analisar decisões judiciais russas com PLN. Conrado et. al (CONRAD et al., 2005) demonstraram como o *Document Clustering* (Agrupamento de Documentos) eficaz pode ser usado para extrair informações de coleções de escritórios de advocacia. Opijnen et. al (OPIJNEN; SANTOS, 2017) apresentaram percepções sobre a relevância da visão geral na extração de informações de documentos legais.

Outro grupo recorrente de abordagens em artigos selecionados é através de PLN. Bolioli et. al (BOLIOLI; MERCATALI; ROMANO, 2004) propuseram um modelo formal para uma gramática legislativa com marcação de texto. Dozier et. al (DOZIER et al., 2007) também apresentaram um estudo com marcação de texto, mas voltado para termos médicos em textos legais. Moens et. al (MOENS; UYTTENDAELE; DUMORTIER, 1999) exploraram o potencial da análise do discurso em documentos jurídicos. Alguns autores (METSKER; TROFIMOV; GRECHISHCHEVA, 2019; DEEDMAN; GELBART; COLEMAN, 1992) escolheram *Latent Semantic Analysis* (Análise Semântica Latente), enquanto outros (FILTZ et al., 2020; TUGGENER et al., 2020a) escolheram *Named Entity Recognition* (Reconhecimento de Entidade Nomeada). Roegiest et. al (ROEGIEST; HUDEK; MCNULTY, 2018) examinaram um conjunto de dados para *Due Diligence* expandindo *Named Entity Recognition* com uma técnica de *Conditional Random Fields*

(Campos Aleatórios Condicionais).

No contexto de extração de conhecimento sobre documentos legais, as abordagens mais frequentes empregam técnicas de IA e RC. Nos campos de IA, Turtle (TURTLE, 1995) aplicou uma rede de inferência bayesiana para recuperar estatisticamente informações relevantes em grandes coleções de documentos. Cunningham et. al (CUNNINGHAM et al., 2004), por outro lado, apresentaram um raciocínio textual baseado em casos para investigar grafos de conhecimento. Lagos et al. (LAGOS et al., 2010) propuseram uma ontologia de extração de eventos para construção e raciocínio jurídico. Cifuentes-Silva e Gayo (CIFUENTES-SILVA; GAYO, 2019); Shahab, Kieninger, e Dengel (SHAHAB; KIENINGER; DENGEL, 2010) usaram técnicas de *Web* semântica para extrair informações úteis de páginas da *Web* e tecnologias com conteúdo de documentos legislativos. Allen, Tury (ALLEN; TURY, 2007); Konstantinou, Sykes, Yannopoulos (KONSTANTINOOU; SYKES; YANNOPOULOS, 1993); Lehmann (LEHMANN, 2007) também utilizaram sistemas especialistas para auxiliar na tarefa de extração de conhecimento. No entanto, Allen (MOULIN; ROUSSEAU, 1990b; MOULIN; ROUSSEAU, 1990a); Moulin, Rousseau (MOULIN; ROUSSEAU, 1990b; MOULIN; ROUSSEAU, 1990a) auxiliaram a gestão da extração de conhecimento através de um modelo de base de conhecimento deôntico.

3.2.3 Geração

A categoria Geração agrega 14.86% dos estudos selecionados e se preocupa em aplicar diferentes técnicas para melhorar e processar a linguagem em documentos jurídicos para gerar um novo texto formatado. O texto recém-gerado é mais facilmente legível por humanos e também por máquinas. As máquinas podem utilizar a tarefa de geração e utilizar o resultado como entrada pré-processada entre diferentes etapas até atingir seu objetivo final.

Machine Translation (Tradução Automática) foi a abordagem mais frequente nos estudos selecionados. Farzindar et al. (FARZINDAR; LAPALME, 2009) escolheram um modelo estatístico para tradução de sentenças judiciais canadenses de Inglês para Francês e de Francês para Inglês. Ye et al. (YE et al., 2018) escolheram um modelo neural para interpretar as previsões de acusação para casos criminais. No entanto, Goltz et al. (GOLTZ; DONDOLI; CAMERON-HUFF, 2018) realizaram um estudo comparando abordagens neurais e estatísticas na área de *LegalTech*. Alguns autores (BOUAYAD-AGHA et al.,

2009; YAN et al., 2019; MILLE; WANNER, 2008; SHEREMETYEVA, 2003) aplicaram redes neurais no campo de reivindicações de patentes. Sannier et al. (SANNIER et al., 2015) desenvolveram um *framework* para detecção e resolução de referências cruzadas em textos legais.

Autores como Zelenikow, Hunter (ZELEZNIKOW; HUNTER, 1992), e Sprowl (SPROWL, 2006) implementaram estudos com geração de texto legal visando redigir documentos jurídicos. Hovy (HOVY, 1993) usou uma abordagem baseada no discurso para a redigir texto legal. Saxon (SAXON, 2006), por outro lado, aplicou uma abordagem baseada em *templates*. Branting (BRANTING, 1998) utilizou uma combinação de técnicas baseadas em discurso, baseadas em modelo e procedimentais para redação automatizada.

Um grupo de estudos baseia-se em métodos de aprendizado e raciocínio para melhorar a etapa de processamento da linguagem. Sihn, Nguyen, e Satoh (SINH; NGUYEN; SATOH, 2019) empregaram abordagem com *Deep Learning* (Aprendizado Profundo) e obtiveram resultados bem promissores. Huang et al. (HUANG et al., 2020), Peric et al. (PERIC et al., 2020-12), por outro lado, investiram em técnicas com modelagem de linguagem. Outras abordagens propostas por Alschner, Skougarevskiy (ALSCHNER; SKOUGAREVSKIY, 2017), e Singh, Sharma (SINGH; SHARMA, 2018) envolvem o uso de redes neurais. Os métodos baseados em raciocínio se apoiaram no uso de ontologia, conforme apresentado por Maxwell, Oberlander e Lavrenko (MAXWELL; OBERLANDER; LAVRENKO, 2009); além do uso de sistemas especialistas, como propuseram Sprowl et al. (SPROWL et al., 1984).

Outros estudos por Singh, Sharma (SINGH; SHARMA, 2018), Huang et al. (HUANG et al., 2020), Zeleznikow, Hunter (ZELEZNIKOW; HUNTER, 1992), e Sprowl (SPROWL, 2006) usaram técnicas como Sistemas de Perguntas e Respostas para melhorar a confiabilidade do texto processado. Sannier et al. (Sannier et al., 2017) escolheram uma abordagem de anotação de texto com marcação estrutural quando o texto legal é gerado.

3.2.4 Ilação

Esta categoria inclui 17.56% dos trabalhos selecionados, que propõem diferentes abordagens de aprendizagem e modelos de raciocínio em documentos legais para auxiliar advogados na previsão de acusações e sentenças de processos criminais. A tarefa de ilação na geração de texto jurídico usa amplamente não apenas abordagens de aprendizado e

raciocínio, como redes bayesianas, *Case-based Reasoning*, *Deep Learning*, Aprendizado de máquina (*Machine Learning*) e redes neurais, mas também ontologias, Sistemas de Suporte a Decisões (*Decision Support Systems*), e Sistemas de Perguntas e Respostas.

A aprendizagem é um tema frequente em diferentes estudos selecionados. Os campos mais comuns de aprendizado são *Support Vector Machine* (SVM) (Máquina de Vetores de Suporte) (ALETRAS et al., 2016; GUO et al., 2020; GUO et al., 2019; NAVAS-LORO; SATOH; RODRÍGUEZ-DONCEL, 2019) e *Deep Learning*, em particular CNN, (GUO et al., 2020; GUO et al., 2019; XIAO et al., 2018; XU et al., 2019; GRABMAIR; ASHLEY, 2011). Alguns autores trabalham com Aprendizado de Máquina (GUO et al., 2020; GRABMAIR, 2017; GUO et al., 2019). Outros autores exploram abordagens com redes neurais. Luo et. al (LUO et al., 2017b), por exemplo, utiliza RNN e, em outro estudo, um modelo baseado em atenção (LUO et al., 2017a). Chen et. al (CHEN et al., 2019), por sua vez, prefere aplicar LSTM (*Long Short-Term Memory*) (Memória de Curto Prazo Longa). Ainda assim, Verheij (VERHEIJ, 2016) aplicou Redes Bayesianas para melhorar os argumentos de raciocínio e probabilidades do modelo de aprendizagem proposto.

O raciocínio também desempenha um papel importante nas tarefas de ilação. *Case-based Reasoning* é a abordagem mais popular nos estudos selecionados. No entanto, apenas Grabmair (GRABMAIR, 2017) usou uma solução puramente com *Case-based Reasoning* para prever os resultados. Alguns autores combinaram *Case-based Reasoning* com outras abordagens para melhorar o processo de raciocínio: Brüninghaus et al. (BRÜNINGHAUS; ASHLEY, 2005) combinou com *Nearest Neighbor* (Vizinho mais Próximo) para classificar através de uma medida de similaridade; McDonald et al. (MCDONALD; PUSTEJOVSKY, 1986) combinou com geração de linguagem natural durante a etapa análise de discurso; Ashley, Brüninghaus (ASHLEY; BR, 2003), e Branting (BRANTING, 1991) apresentaram uma combinação de *Case-based Reasoning* com um método de predição baseado em problemas.

Em relação aos estudos que apresentam abordagens inteiramente de aprendizado e raciocínio, alguns autores focam na aplicação dessas abordagens em Sistemas de Suporte a Decisão (MOLES; DAYAL, 1993; WALKER et al., 2017; ZHOU et al., 2019) e Sistemas de Perguntas e Respostas (FAWEI et al., 2019; CHITTA; HUDEK, 2019) em previsões de acusações e sentenças na área jurídica. Moles et. al (MOLES; DAYAL, 1993) melhoraram os resultados de Sistemas de Perguntas e Respostas com manipulação semântica usando anotações de texto. Por fim, vários autores (FAWEI et al., 2019; VÖL-

KER; LANGA; SURE, 2008; BREUKER; PETKOV; WINKELS, 2000; GREENLEAF; MOWBRAY; DIJK, 1995; WALKER et al., 2017) aplicaram ontologias para modelar conhecimento e fazer inferências sobre decisões de domínio jurídico.

3.2.5 Recuperação

A categoria Recuperação é a tarefa de geração de textos jurídicos mais recorrente na área *LegalTech* e inclui 31,08% dos estudos selecionados. Esta categoria tem como objetivo lidar com armazenamento de documentos e recuperação automática de documentos legais. Os estudos selecionados propõem diferentes métodos para melhorar as tarefas relacionadas a busca, consulta, ponderação de termos, classificação, indexação e sumarização.

As abordagens de sumarização de texto são a pesquisa mais recorrentes entre os estudos selecionados (WAGH; ANAND, 2020; GALGANI; COMPTON; HOFFMANN, 2012b; HACHEY; GROVER, 2005; FARZINDAR; LAPALME, 2004; GALGANI; HOFFMANN, 2011; HACHEY; GROVER, 2006; GELBART; SMITH, 1993; GALGANI; COMPTON; HOFFMANN, 2012a; KONIARIS; ANAGNOSTOPOULOS; VASSILIOU, 2017; JIA et al., 2014; SARAVANAN; RAVINDRAN; RAMAN, 2006; GROVER et al., 2003; SHINMORI; MARUKAWA, 2003). A sumarização de texto jurídico é o processo de encurtar dados de documentos jurídicos para criar resumos e sumários que representem as informações mais relevantes dentro do conteúdo original.

Outra etapa importante durante a RI é a etapa de indexação. A indexação é a principal funcionalidade do processo de RI, pois auxilia na eficiência da recuperação de informações. Para melhorar a eficiência da etapa de indexação durante a recuperação do texto legal, Xu et al. (XU; CROFT, 1998) apresentaram uma derivação baseada em corpus usando a coocorrência de variantes de palavras que reduzem os documentos aos termos informativos neles contidos. Outros dois estudos de Brüninghaus e Ashley (BRÜNINGHAUS; ASHLEY, 2001; BRÜNINGHAUS; ASHLEY, 1999) propõem uma abordagem para indexar automaticamente textos de opinião jurídica para sistemas *Case-based Reasoning*.

Em relação aos métodos de consulta e busca, alguns autores (NAZARENKO; LEVY; WYNER, 2018; NEJADGHOLI; TCHEMEUBE; WITHERSPOON, 2017; Yamada; Teufel; Tokunaga, 2017; NAVAS-LORO; SANTOS, 2019; SHAHEEN; WOHLGENANT; FILTZ, 2020) propõem uma busca semântica onde uma consulta visa não

apenas encontrar palavras-chave, mas determinar a intenção e o significado contextual da consulta de entrada e recuperar informações relevantes de documentos legais. Em outros dois estudos realizados por Nazarenko, Levy, e Wyner (NAZARENKO; LEVY; WYNER, 2018), e Yamada, Teufel, Tokunaga, (Yamada; Teufel; Tokunaga, 2017), um esquema de anotação de texto foi utilizado para desenvolver uma linguagem de anotação que auxilia a busca semântica em corpora jurídicos. Locke et al. (LOCKE; ZUCCON; SCHELLS, 2017) apresenta uma abordagem de geração automática de consultas para recuperação de jurisprudência por meio de estratégia de expansão de consultas. Em quatro outros estudos (MOENS, 2006; PEÑAS et al., 2010; HAUSLADEN; SCHUBERT; ASH, 2020; DO et al., 2017), um Sistema de Perguntas e Respostas foi escolhido para aprimorar a consulta de entrada original e recuperar informações mais relevantes do processo de RI. Em outros dois estudos de Chen, Liu, e Ho (CHEN; LIU; HO, 2013; LIU; CHEN; HO, 2015), algoritmos de mineração de texto foram aplicados para aumentar a relevância das informações recuperadas em documentos legais em relação a uma determinada consulta de entrada.

A relevância desempenha um papel central no processo de RI. Um documento relevante implica que o documento ou conjunto de documentos recuperados atende à necessidade de informação requisitada pelo usuário. Assim, também é um tema recorrente na recuperação de informações jurídicas. Alguns autores (CHEN; LIU; HO, 2013; FALAKMASIR; ASHLEY, 2017; DEANE, 2019) aplicam a similaridade de cosseno para medir a similaridade entre documentos legais e classificá-los por relevância. Guo Z. (GUO et al., 2019) combina ainda *Latent Dirichlet Allocation* (Alocação de Dirichlet Latente) (LDA) com similaridade de cosseno para melhorar o processo de classificação. Em outros dois estudos propostos por Merkl, Schweighofer, (Merkl; Schweighofer, 1997) e Falakmasir, Ashley (FALAKMASIR; ASHLEY, 2017), a abordagem selecionada foi um modelo de espaço vetorial para representar termos de índice e melhorar as etapas de indexação e classificação de relevância. Kumar et al. (KUMAR et al., 2011), e Lau, Law, e Wiederhold (LAU; LAW; WIEDERHOLD, 2003) comparam diferentes técnicas de análise de similaridade e apresentam uma visão geral dos resultados de suas aplicações na recuperação de informações jurídicas. No entanto, Turtle (TURTLE, 1994) realiza uma comparação de desempenho entre linguagem natural e abordagens de consulta booleana na tarefa de recuperação em documentos legais.

Por fim, a categoria Recuperação contém abordagens de aprendizado e raciocínio dos artigos selecionados. As abordagens usam principalmente técnicas de IA e RC para melhorar processos relacionados a RI, como indexação, consulta, classificação e pesquisa.

Alguns autores (SHAHEEN; WOHLGENANNT; FILTZ, 2020; MEDVEDEVA; VOLS; WIELING, 2020) propõem uma classe de algoritmo *Deep Learning* para o processo de aprendizado. Em uma abordagem mais específica, em (HAUSLADEN; SCHUBERT; ASH, 2020; DO et al., 2017; CHALKIDIS et al., 2019; FILTZ et al., 2019), CNN é o modelo de rede neural utilizado e alguns benefícios na recuperação de texto legal são apresentados. Filtz et al. (SHAHEEN; WOHLGENANNT; FILTZ, 2020) apresentaram uma solução de incorporação de palavras que aplica técnicas de PLN para refinar a consulta inicial e depois mesclar o aprendizado com uma abordagem de aprendizado profundo. Em outros dois estudos (BRÜGMANN et al., 2015; HAUSLADEN; SCHUBERT; ASH, 2020), o processo de aprendizagem foi desenvolvido através de técnicas com SVM. Farzindar et al. (CHALKIDIS et al., 2019) aplicou *Large-Scale Multi-Label Text Classification* (LMTC) (Classificação de Texto em Grande Escala com Vários Rótulos) com vários classificadores neurais no domínio legal e também divulgaram o conjunto de dados de documentos legislativos usados, além da avaliação de performance com outros métodos atuais de última geração. Wang et al. (KIM, 2014) apresenta uma modelagem com *Dynamic Pairwise Attention* (Atenção Dinâmica de Pares) para classificação de crimes sobre artigos legais. Sistemas especialistas também são uma solução comum para tarefas de raciocínio sobre a categoria de Recuperação em alguns estudos (LEITH, 1986; SHAHEEN; WOHLGENANNT; FILTZ, 2020). Certos autores (BRÜNINGHAUS; ASHLEY, 1999; GROVER et al., 2003; SHAHEEN; WOHLGENANNT; FILTZ, 2020) recorrem a técnicas de *Case-based Reasoning* para aprimorar o raciocínio com o conhecimento adquirido durante as etapas de RI. Galgani (GALGANI; HOFFMANN, 2011) apresenta uma abordagem para classificação de citações legais usando aquisição incremental de conhecimento. Deane (SAXON, 2006) propõe um raciocínio baseado em regras com paradigma de lógica padrão para recuperação de fatos jurídicos. Quintana et al. (ZELEZNIKOW; HUNTER, 1992) apresenta uma combinação de aquisição de conhecimento e recuperação baseada em grafos em sistemas de hipertexto e como essa abordagem pode ser aplicada na área *LegalTech*.

3.3 Resultados e discussão

Nesta seção, discutimos nossos achados respondendo às questões de pesquisa que abordamos em nossa revisão sistemática da literatura.

3.3.1 QP1 (principal)

A principal questão de pesquisa, QP1, visa investigar, de forma ampla, quais soluções estão sendo aplicadas na área de tecnologia para gerar textos jurídicos em documentos legais. A Tabela 3.3 apresenta a distribuição dos estudos entre as categorias definidas na Taxonomia proposta na Seção 3.2: Argumentação, Extração, Geração, Ilação, e Recuperação.

Tabela 3.3: Distribuição das categorias de taxonomia definidas, por objetivo principal, dos 146 estudos selecionados.

Categoria	Total (%)
Recuperação	46 (31.5%)
Ilação	25 (17.12%)
Extração	24 (16.43%)
Geração	21 (14.38%)
Argumentação	14 (9.59%)
<i>Survey</i>	14 (9.59%)
<i>Dataset</i>	2 (1.37%)

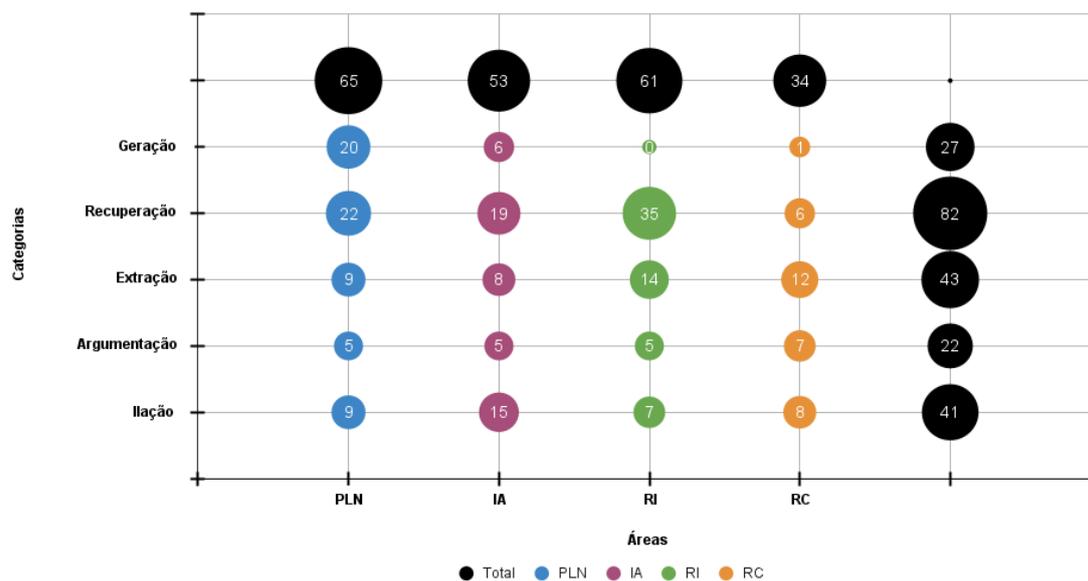
Fonte: o autor

Além disso, detectamos áreas comuns da Ciência da Computação que aplicaram algoritmos e/ou formalismos dos dados extraídos para atingir o objetivo do estudo selecionado. Os mais recorrentes foram PLN, IA, RI e RC. Em seguida, analisamos quais áreas foram escolhidas em relação ao objetivo principal do artigo selecionado. Como pode ser visto na Figura 3.3, PLN foi a área de Ciência da Computação mais frequente entre os objetivos dos estudos selecionados (66). Alternativamente, RC foi o menos frequente, aparecendo 34 vezes. A maioria dos estudos selecionados é da categoria Recuperação. Portanto, a categoria Recuperação possui um número maior de áreas envolvidas. A recíproca é verdadeira: a categoria Argumentação teve a minoria de estudos selecionados e áreas envolvidas.

Apesar da taxonomia ser dividida em cinco categorias principais, alguns trabalhos foram agrupados em outras duas categorias auxiliares: *survey* e *dataset*. Esta decisão foi adotada porque o escopo destes trabalhos era mais amplo, uma vez que objetivava apresentar coleções de dados e/ou revisar outros trabalhos da literatura na área de *LegalTech*, o que dificultava agrupá-los em uma das categorias principais, uma vez que estas possuem um escopo mais fechado.

Também é perceptível que cada categoria tem uma área de Ciência da Computação

Figura 3.3: Representação da incidência de áreas da Ciência da Computação por categorias de taxonomia.



Fonte: o autor

predominante. A categoria de Geração corresponde com mais frequência às soluções de PLN (75%). A categoria de Recuperação geralmente adota algoritmos de RI (42%). A categoria de Extração também usa RI de forma mais recorrente (33,33%), mas RC também compartilha um papel importante (28,57%). A categoria Argumentação possui as áreas aplicadas mais bem distribuídas, embora RC seja a mais comum entre elas (31,81%). Por fim, a categoria Ilação utiliza a IA como área predominante (37,02%).

3.3.2 QP1.1 - Com que finalidade a tecnologia é aplicada na área jurídica?

Para responder à QP1.1, classificamos os estudos selecionados quanto ao seu objetivo principal, conforme mostrado na Seção 3.2. Para identificar quais objetivos principais visavam cada estudo selecionado, focamos no título e palavras-chave extraídas dos itens de dados e seus resumos. Caso o objetivo principal não tenha sido apresentado com clareza, eram lidas a seção de introdução dos estudos e, sempre que necessário, a seção de conclusões e, por fim, as demais seções.

No entanto, é importante notar que alguns estudos apresentaram múltiplos objetivos. Então, novamente, lemos todas as seções necessárias para classificar o artigo em uma categoria que se encaixa no objetivo predominante do estudo.

3.3.3 QP1.2 - Quais tecnologias são mais populares na área jurídica?

A questão de pesquisa QP1.2 diz respeito às soluções mais recorrentes em computação aplicadas à área de tecnologia jurídica. Durante a etapa de extração de dados da pesquisa, extraímos o título e as palavras-chave de cada um dos 146 artigos selecionados. Como havia muitos termos, removemos manualmente *stopwords*, palavras muito genéricas, ou que apareciam apenas uma vez. PLN foi a solução mais recorrente, aparecendo em 21 artigos. A Tabela 3.4 apresenta todos os resultados obtidos.

Tabela 3.4: Distribuição dos artigos, em ordem crescente, das 10 áreas de ciência da computação mais recorrentes dos 146 estudos selecionados.

Tópico	Total (%)
Abstract Meaning Representation	4 (2,7%)
Machine translation	4 (2,7%)
Suppor Vector Machine	6 (4,05%)
Redes Neurais	6 (4,05%)
Ontologia	7 (4,73%)
Deep Learning	7 (4,73%)
Case-Based Reasoning	11 (7,43%)
Information Retrieval	17 (11,48%)
Machine Learning	18 (12,16%)
Processamento de Linguagem Natural	21 (14,19%)

Fonte: o autor

3.3.4 QP1.3 - A geração de texto é automática, semi-automática ou manual?

A questão de pesquisa QP1.3 objetiva responder se o estudo selecionado apresenta soluções automáticas, semiautomáticas ou manuais para técnicas de geração de texto. Primeiramente, para responder a essa pergunta, definimos o que é uma solução automática, semiautomática e manual. Uma solução automática é uma solução que fornece uma saída de dados sem ação do usuário. Uma solução semiautomática é uma solução que fornece uma saída de dados com uma ou mais ações do usuário durante a execução da técnica e, pelo menos, uma ação automática. Por fim, uma solução manual é uma solução em que todo o processo entre a entrada de dados e a saída de dados tem ação do usuário. Além disso, consideramos trabalhos de solução manual que apresentavam abordagens puramente teóricas e/ou *benchmarks* entre diferentes técnicas. A Tabela 3.5 apresenta a distribuição dos estudos entre os métodos de geração de texto e os classifica em automá-

tica, semiautomática ou manual. Ao todo, foram selecionados 132 artigos de acordo com suas estratégias para atingir seu objetivo principal. A quantidade de artigos é um subconjunto do total de 146 artigos selecionados porque não consideramos artigos de pesquisa nesta análise.

A maioria dos estudos, 59 (44,03%), apresentou soluções automáticas para geração de texto legal (direta ou indiretamente). As técnicas mais populares relacionadas à área de processamento de linguagem natural são o *aprendizado não supervisionado com métodos probabilísticos* e *redes neurais*. As soluções manuais correspondem a 42 artigos (31,34%) e estudos em grupo que apresentaram avaliação de desempenho entre técnicas, abordagens teóricas e conjuntos de dados para métodos de aprendizagem e treinamento. Além disso, 33 artigos (24,62%) apresentaram uma solução semiautomática que consiste principalmente em uma abordagem de *aprendizado supervisionado* onde os usuários podem fornecer feedback ao algoritmo para melhorar a saída esperada do software.

Tabela 3.5: Classificação em automática, semiautomática ou manual dos 132 artigos selecionados.

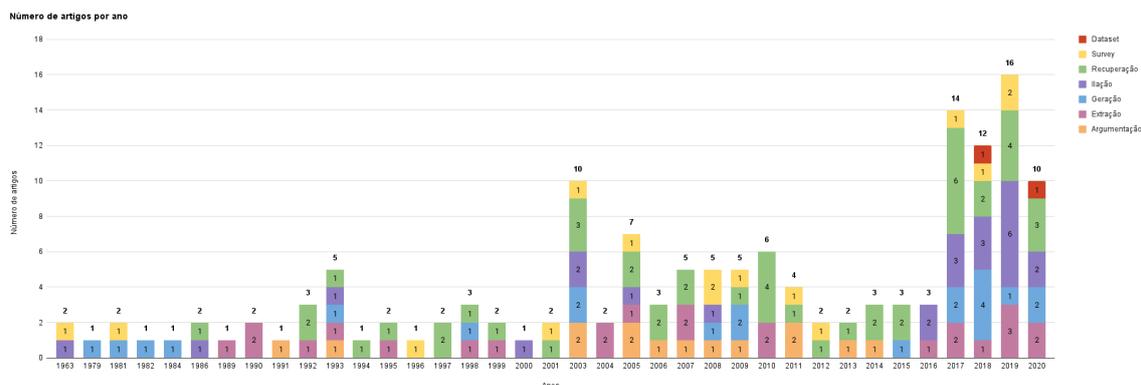
Categoria	Artigos	Total (%)
Automático	(ALSCHNER; SKOUGAREVSKIY, 2017)(Sannier et al., 2017)(Merkl; Schweighofer, 1997)(MOENS; UYTENDAELE; DUMORTIER, 1999)(WAGH; ANAND, 2020)(METS-KER; TROFIMOV; GRECHISHCHEVA, 2019)(CIFUENTES-SILVA; GAYO, 2019)(MOCHALES; MOENS, 2009b)(YE et al., 2018)(LOCKE; ZUCCON; SCELLS, 2017)(AGNOLONI; BACCI; OPIJNEN, 2017)(ANDERSSON et al., 2016)(GALGANI; COMPTON; HOFFMANN, 2012b)(SHAHAB; KIENINGER; DENGEL, 2010)(HACHEY; GROVER, 2005)(HOVY, 1993)(MOCHALES; MOENS, 2008)(CHEN; LIU; HO, 2013)(LIU; CHEN; HO, 2015)(LUO et al., 2017a)(SANNIER et al., 2015)(ALETRAS et al., 2016)(FARZINDAR; LAPALME, 2004)(FARZINDAR; LAPALME, 2009) (GALGANI; HOFFMANN, 2011)(HACHEY; GROVER, 2006)(DEEDMAN; GELBART; COLEMAN, 1992)(GELBART; SMITH, 1993)(GALGANI; COMPTON; HOFFMANN, 2012a)(Yamada; Teufel; Tokunaga, 2017)(JIA et al., 2014)(FALAKMASIR; ASHLEY, 2017)(SARAVANAN; RAVINDRAN; RAMAN, 2006)(BRÜNINGHAUS; ASHLEY, 2005)(DOZIER et al., 2007)(CHITTA; HUDEK, 2019)(ASHLEY; WALKER, 2013)(BOUAYAD-AGHA et al., 2009)(SINGH; SHARMA, 2018)(HUANG et al., 2020)(PERIC et al., 2020-12)(GROVER et al., 2003)(QUINTANA; KAMEL; LO, 1992)(SPROWL, 2006)(HE et al., 2018)(XU et al., 2019)(YAN et al., 2019) (SHINMORI; MARUKAWA, 2003)(DO et al., 2017)(BRANTING, 1991)(BRÜNINGHAUS; ASHLEY, 2003)(CHORLEY; BENCH-CAPON, 2005)(BRANTING, 1998)(CHEN et al., 2019)(HU et al., 2018)(MEDVEDEVA; VOLTS; WIELING, 2020)(WANG et al., 2018)	57 (39.04%)
Manual	(SINH; NGUYEN; SATOH, 2019)(VÖLKER; LANGA; SURE, 2008)(CONRAD et al., 2005)(BOLIOLI; MERCATALI; ROMANO, 2004)(GREENLEAF; MOWBRAY; DIJK, 1995)(TURTLE, 1994)(OPIJNEN; SANTOS, 2017)(TURTLE, 1995)(MOCHALES; MOENS, 2009a)(MOLES; DAYAL, 1993)(KONIARIS; ANAGNOSTOPOULOS; VASSILOU, 2017)(PEÑAS et al., 2010)(WALKER et al., 2017)(KUMAR et al., 2011)(VERHEIJ, 2016)(GRABMAIR, 2017)(FILTZ et al., 2020)(ASHLEY; BR, 2003)(CUNNINGHAM et al., 2004)(ALLEN, 1998)(LEHMANN, 2007)(MOULIN; ROUSSEAU, 1990a)(DEANE, 2019) (GUO et al., 2019)(XIAO et al., 2018)(SHEREMETYEVA, 2003) (HAUSLADEN; SCHUBERT; ASH, 2020)(ATKINSON; BENCH-CAPON, 2005)(BENCH-CAPON; SARTOR, 2003)(BERMAN; HAFNER, 1993)(FETERIS, 2015)(GRABMAIR; ASHLEY, 2011)(ZHONG et al., 2018)(ZHONG et al., 2019a)(MAXWELL; OBERLANDER; LAVRENKO, 2009)(NAVAS-LORO; SATOH; RODRÍGUEZ-DONCEL, 2019)(CHALKIDIS et al., 2019)(SHAHEEN; WOHLGENANNT; FILTZ, 2020)(FILTZ et al., 2019)(LEITNER; REHM; SCHNEIDER, 2019)(KIM, 2014)(NAGEL, 1963)	42 (28.76%)
Semiautomático	(FAWEI et al., 2019)(GOLTZ; DONDOLI; CAMERON-HUFF, 2018)(NAZARENKO; LEVY; WYNER, 2018)(ROSSI; WIRTH; KANOULAS, 2019)(BREUKER; PETKOV; WINKELS, 2000)(BRÜNINGHAUS; ASHLEY, 2001)(ZHANG; KOPPAKA, 2007)(MOENS, 2006)(NEJADGHOLI; TCHEMEUBE; WITHERSPOON, 2017)(MCDONALD; PUSTEJOVSKY, 1986)(ALLEN; TURY, 2007)(KONSTANTINOUS; SYKES; YANNOPOULOS, 1993)(ZHOU et al., 2019)(XU; CROFT, 1998)(LAU; LAW; WIEDERHOLD, 2003)(SPROWL et al., 1984)(GUO et al., 2020)(ALEVEN, 2003)(BRÜNINGHAUS; ASHLEY, 1999)(DANIELS; RISSLAND, 1997)(LEITH, 1986)(MOULIN; ROUSSEAU, 1990b) (WIDDISON; PRITCHARD; ROBINSON, 1992)(SPROWL; STAUD†, 1981)(SAXON, 2006) (MILLE; WANNER, 2008)(BRÜGMANN et al., 2015)(KIM; XU; GOEBEL, 2017)(ROEGIEST; HUDEK; MCNULTY, 2018)(ASHLEY; BRÜNINGHAUS, 2006)(LAGOS et al., 2010)(WYNER; PETERS, 2010)(TUGGENER et al., 2020a)	33 (22.6%)

Fonte: o autor

3.3.5 QP1.4 - Quão ativa é a pesquisa sobre geração de texto legal?

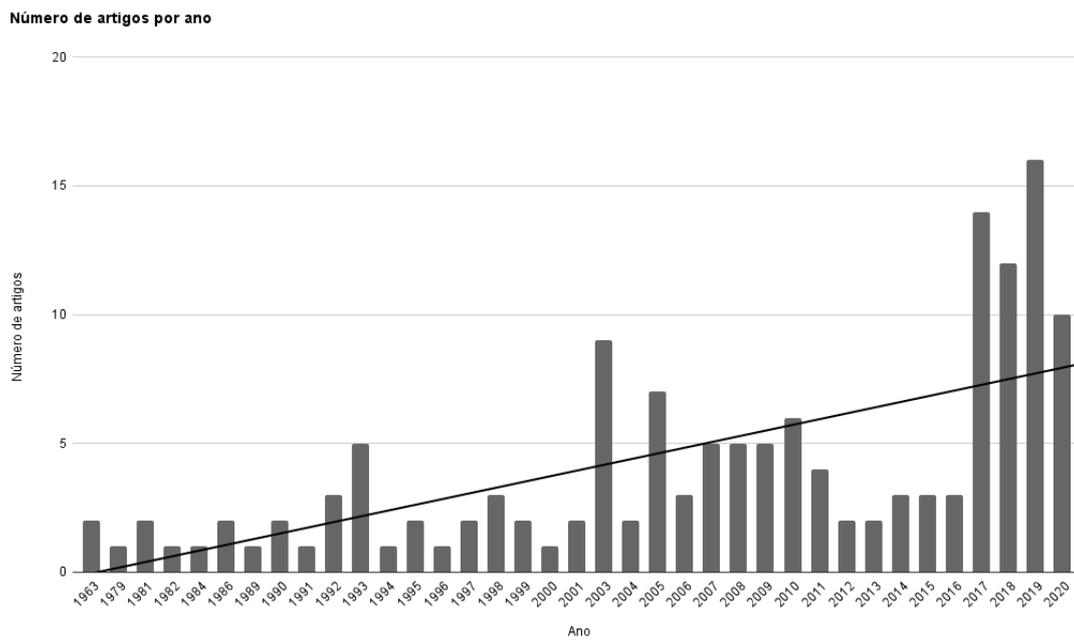
A questão de pesquisa QP1.4 está relacionada ao ano de publicação de cada artigo. Agrupamos os artigos selecionados por ano, com o objetivo principal de investigar se a literatura nesta área tende a crescer ou diminuir nos próximos anos. A Figura 3.4 mostra a distribuição dos artigos selecionados por ano, por objetivo principal. Adicionalmente foram incluídos os trabalhos de pesquisas da área de tecnologia jurídica e trabalhos apresentando conjuntos de dados jurídicos disponíveis. Além disso, a Figura 3.5 apresenta uma linha de tendência para mostrar que há um aumento no número de artigos na área de tecnologia jurídica por ano. A Figura 3.6 apresenta a distribuição dos trabalhos selecionados de acordo com o país de origem da instituição vinculada aos autores dos trabalhos.

Figura 3.4: Número de artigos publicados em cada um dos anos de 1963 a 2020 de acordo com seu objetivo principal.



Fonte: o autor

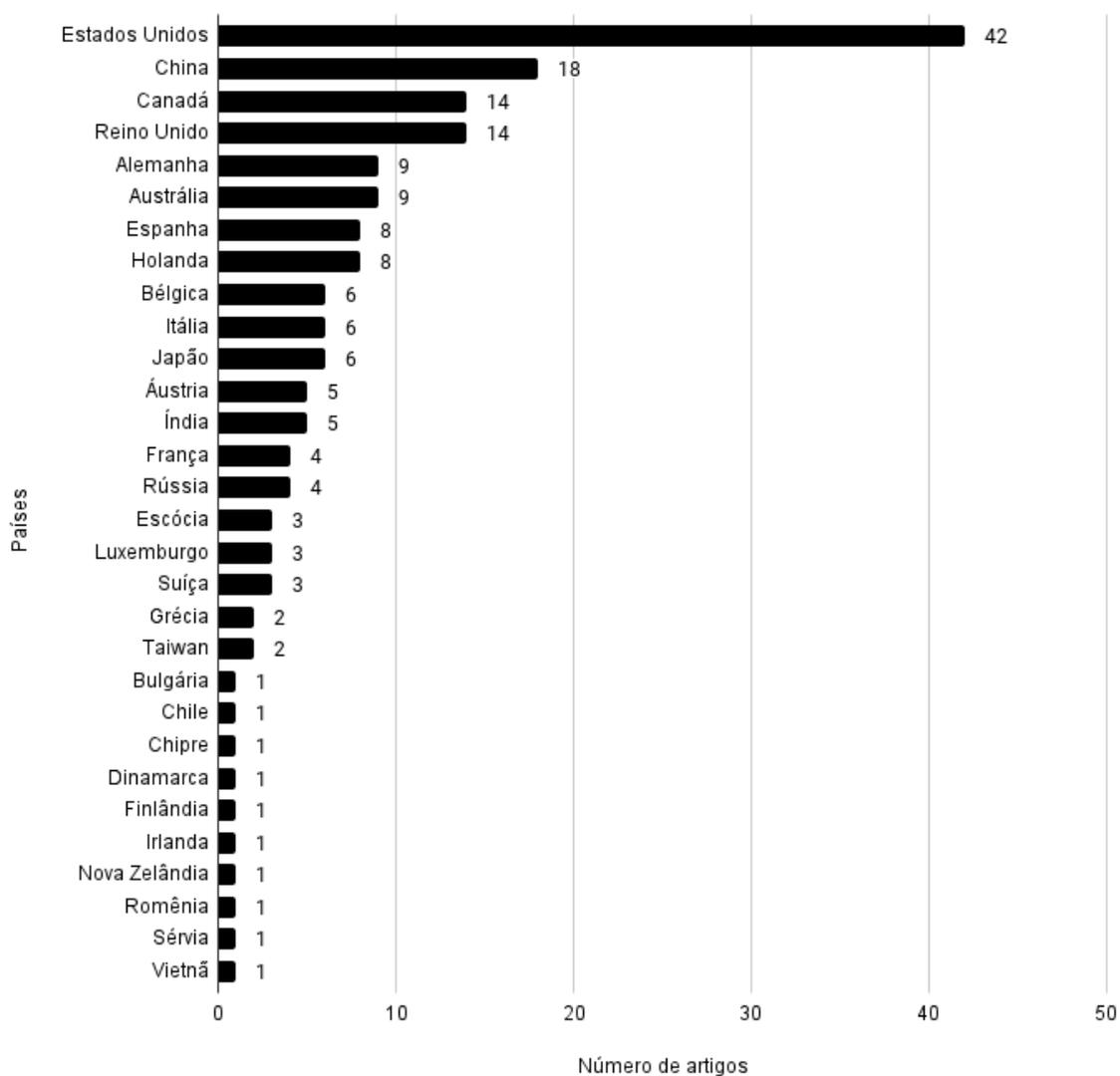
Figura 3.5: A linha de tendência do total de artigos selecionados publicados dos anos de 1963 a 2020.



Fonte: o autor

Figura 3.6: O número de trabalhos publicados em cada país dos anos de 1963 a 2020.

Número de artigos por país



Fonte: o autor

Por fim, a Tabela 3.6 fornece a distribuição de artigos por ano. É possível notar que os últimos quatro anos foram responsáveis por 36,49% dos trabalhos publicados. O ano mais ativo foi 2019 com 16 artigos publicados.

Tabela 3.6: Distribuição dos artigos publicados por ano dos 146 estudos selecionados.

Ano	Artigos	Total (%)
2020	(METSKER; TROFIMOV; GRECHISHCHEVA, 2019)(GUO et al., 2020)(FILTZ et al., 2020)(HUANG et al., 2020)(PERIC et al., 2020-12)(HAUSLADEN; SCHUBERT; ASH, 2020) (ZHONG et al., 2019a)(MEDVEDEVA; VOLS; WIELING, 2020)(SHAHEEN; WOHLGENANT; FILTZ, 2020)(TUGGENER et al., 2020a)	10 (6.85%)
2019	(WAGH; ANAND, 2020)(SINH; NGUYEN; SATOH, 2019)(FAWEI et al., 2019)(CIFUENTES-SILVA; GAYO, 2019)(ROSSI; WIRTH; KANOULAS, 2019)(VU; NGUYEN, 2019) (ZHOU et al., 2019)(CHITTA; HUDEK, 2019)(GUO et al., 2019)(XU et al., 2019)(YAN et al., 2019) (CHEN et al., 2019)(MARKOVIĆ; GOSTOJIĆ, 2019)(CHALKIDIS et al., 2019)(FILTZ et al., 2019)(LEITNER; REHM; SCHNEIDER, 2019)	16 (10.96%)
2018	(GOLTZ; DONDOLI; CAMERON-HUFF, 2018)(NAZARENKO; LEVY; WYNER, 2018)(YE et al., 2018)(SINGH; SHARMA, 2018)(HE et al., 2018)(XIAO et al., 2018) (ROEGEST; HUDEK; MCNULTY, 2018)(HU et al., 2018)(ZHONG et al., 2018)(NAVAS-LORO; SANTOS, 2019)(NAVAS-LORO; SATOH; RODRÍGUEZ-DONCEL, 2019)(WANG et al., 2018)	12 (8.22%)
2017	(ALSCHNER; SKOUGAREVSKIY, 2017)(Sannier et al., 2017)(LOCKE; ZUCCON; SCELLS, 2017)(AGNOLONI; BACCI; OPIJNEN, 2017)(OPIJNEN; SANTOS, 2017)(LUO et al., 2017a) (PEÑAS et al., 2010)(NEJADGHOLI; TCHEMEUBE; WITHERSPOON, 2017)(Yamada; Teufel; Tokunaga, 2017)(WALKER et al., 2017)(FALAKMASIR; ASHLEY, 2017)(KANAPALA; PAL; PAMULA, 2019)(GRABMAIR, 2017) (DO et al., 2017)	14 (9.59%)
2016	(ANDERSSON et al., 2016)(ALETRAS et al., 2016)(VERHEIJ, 2016)	3 (2.05%)
2015	(LIU; CHEN; HO, 2015)(SANNIER et al., 2015)(KIM; XU; GOEBEL, 2017)	3 (2.05%)
2014	(JIA et al., 2014)(FETERIS, 2015)(KIM, 2014)	3 (2.05%)
2013	(CHEN; LIU; HO, 2013)(ASHLEY; WALKER, 2013)	2 (1.37%)

Tabela 3.6 – continuação da página anterior

Ano	Artigos	Total (%)
2012	(GALGANI; COMPTON; HOFFMANN, 2012b)(BENCH-CAPON et al., 2012)	2 (1.37%)
2011	(MOCHALES; MOENS, 2009b)(KONIARIS; ANAGNOSTOPOULOS; VASSILIOU, 2017)(OPIJNEN, 2012)(GRABMAIR; ASHLEY, 2011)	4 (2.74%)
2010	(SHAHAB; KIENINGER; DENGEL, 2010)(GALGANI; HOFFMANN, 2011)(KUMAR et al., 2011)(GRABMAIR; ASHLEY, 2010)(LAGOS et al., 2010)(WYNER; PETERS, 2010)	6 (4.11%)
2009	(MOZGOVOY; KAKKONEN; COSMA, 2010)(FARZINDAR; LAPALME, 2009)(MOCHALES; MOENS, 2009a)(BOUAYAD-AGHA et al., 2009)(MAXWELL; OBERLANDER; LAVRENKO, 2009)	5 (3.42%)
2008	(VöLKER; LANGA; SURE, 2008)(MOCHALES; MOENS, 2008)(WALTER, 2008)(MILLE; WANNER, 2008)(KIYAVITSKAYA et al., 2008)	5 (3.42%)
2007	(ZHANG; KOPPAKA, 2007)(ALLEN; TURY, 2007)(DOZIER et al., 2007)(DEANE, 2019)(BRÜGMANN et al., 2015)	5 (3.42%)
2006	(HACHEY; GROVER, 2006)(MOENS, 2006)(ASHLEY; BRÜNINGHAUS, 2006)	3 (2.05%)
2005	(CONRAD et al., 2005)(HACHEY; GROVER, 2005)(SARAVANAN; RAVINDRAN; RAMAN, 2006)(BRÜNINGHAUS; ASHLEY, 2005)(ATKINSON; BENCH-CAPON, 2005)(CHORLEY; BENCH-CAPON, 2005)(BOLIOLI; MERCATALI; ROMANO, 2004)	7 (4.79%)
2004	(BOLIOLI; MERCATALI; ROMANO, 2004)(CUNNINGHAM et al., 2004)	2 (1.37%)
2003	(FARZINDAR; LAPALME, 2004)(LAU; LAW; WIEDERHOLD, 2003)(ASHLEY; BR, 2003)(GROVER et al., 2003)(SHEREMETYEVA, 2003)(SHINMORI; MARUKAWA, 2003)(BENCH-CAPON; SARTOR, 2003)(BRÜNINGHAUS; ASHLEY, 2003)	9 (6.16%)
2001	(BRÜNINGHAUS; ASHLEY, 2001)(MOENS, 2004)	2 (1.37%)
2000	(BREUKER; PETKOV; WINKELS, 2000)	1 (0.68%)

Tabela 3.6 – continuação da página anterior

Ano	Artigos	Total (%)
1999	(MOENS; UYTTENDAELE; DUMORTIER, 1999)(BRÜNINGHAUS; ASHLEY, 1999)	2 (1.37%)
1998	(XU; CROFT, 1998)(ALLEN, 1998)(BRANTING, 1998)	3 (2.05%)
1997	(Merkl; Schweighofer, 1997)(DANIELS; RISSLAND, 1997)	2 (1.37%)
1996	(ZELEZNIKOW; HUNTER, 1992)	1 (0.68%)
1995	(GREENLEAF; MOWBRAY; DIJK, 1995)(TURTLE, 1995)	2 (1.37%)
1994	(TURTLE, 1994)	1 (0.68%)
1993	(HOVY, 1993)(MOLES; DAYAL, 1993)(GELBART; SMITH, 1993)(KONSTANTINOU; SYKES; YANNOPOULOS, 1993)(BERMAN; HAFNER, 1993)	5 (3.42%)
1992	(DEEDMAN; GELBART; COLEMAN, 1992)(QUINTANA; KAMEL; LO, 1992)(WIDDISON; PRITCHARD; ROBINSON, 1992)	3 (2.05%)
1991	(BRANTING, 1991)	1 (0.68%)
1990	(LEHMANN, 2007)(MOULIN; ROUSSEAU, 1990a)	2 (1.37%)
1989	(MOULIN; ROUSSEAU, 1990b)	1 (0.68%)
1986	(MCDONALD; PUSTEJOVSKY, 1986)(HUANG et al., 2020)	2 (1.37%)
1984	(SPROWL et al., 1984)	1 (0.68%)
1982	(SAXON, 2006)	1 (0.68%)
1981	(SPROWL; STAUD†, 1981)(COOK et al., 1981)	2 (1.37%)
1979	(SPROWL, 2006)	1 (0.68%)
1963	(LAWLOR, 1963)(NAGEL, 1963)	2 (1.37%)

3.3.6 QP1.5 - Quais são as limitações da pesquisa atual?

Nossa pesquisa tem algumas ameaças à validade. Embora tenhamos aplicado protocolos específicos de acordo com (BA; CHARTERS, 2007), a metodologia não garante que cobrimos todos os trabalhos relacionados. Além disso, nosso conjunto inicial teve apenas 44 artigos publicados. Esse pequeno conjunto de documentos pode levar a um grupo restrito de tarefas de *LegalTech* e a um panorama enganoso de geração de texto legal. No entanto, também extraímos

as referências do conjunto inicial e reaplicamos as etapas da metodologia conforme mostrado na Seção 3.1. A quantidade total de trabalhos selecionados foi de 146. Assim, acreditamos que esse conjunto mais amplo de trabalhos evita resultados que não se concentrem em um pequeno nicho da área de *LegalTech*.

Como consideramos apenas trabalhos revisados por pares (por exemplo, artigos de conferências), deixamos de fora trabalhos relacionados não publicados. Por exemplo, não incluímos fairseq⁷, nem CRFsuite⁸ em nosso grupo de trabalhos selecionados, uma vez que fornecem abordagens genéricas de *software* para IA, não necessariamente aplicadas na área de tecnologia jurídica, e não possuem artigos publicados. Fairseq é um *kit* de ferramentas de modelagem de sequência que permite que pesquisadores e desenvolvedores treinem modelos personalizados para tradução, resumo, modelagem de linguagem e outras tarefas de geração de texto. CRFsuite é uma implementação de Campos Aleatórios Condicionais para rotular dados sequenciais por meio de uma única anotação de importação que substitui métodos internos.

Analizamos apenas artigos com acesso ao seu conteúdo, atendendo aos nossos critérios de inclusão e exclusão. Dos 150 estudos, não conseguimos acessar quatro artigos. Esses quatro trabalhos não foram incluídos no conjunto final de trabalhos selecionados por serem considerados antigos (ambos anteriores a 1980) para analisar o cenário atual da área de *LegalTech* e geração de textos legal, e não serem referenciados por outros trabalhos.

Por fim, outra ameaça é a dificuldade de classificação dos artigos. Excluímos os artigos lendo apenas seus resumos e títulos. Embora alguns artigos possam ocultar o escopo de seu objetivo principal no meio do texto, acreditamos que minimizamos a ameaça de classificação ao acompanhar e ler cada seção do artigo quando necessário.

3.4 Considerações finais

Neste Capítulo, descrevemos uma revisão sistemática da literatura sobre geração de textos legais e propusemos uma taxonomia para descrever abordagens de geração de texto legal em relação ao seu objetivo principal. Nossa revisão sistemática da literatura considerou artigos publicados há quase 60 anos (de janeiro de 1963 a dezembro de 2020). Após aplicar um processo iterativo de critérios de exclusão e inclusão a um conjunto de 244 artigos e 1040 artigos de referência, respectivamente, 146 estudos foram selecionados para serem lidos na íntegra e aprovados através do processo de extração de dados.

Com base nos dados extraídos dos 146 artigos selecionados, analisamos os estudos e identificamos quais objetivos e campos da Ciência da Computação estavam relacionados à geração de

⁷<<https://github.com/pytorch/fairseq>>

⁸<<https://www.chokkan.org/software/crfsuite/>>

texto legal. Criamos uma taxonomia e as classificamos em cinco categorias de acordo com seu objetivo principal: Argumentação, Extração, Geração, Ilação e Recuperação. Apesar do texto dos documentos legais como dado de entrada, cada categoria tem propriedades diferentes para manipular a informação e retorná-la como saída. A saída pode ser um resultado final, ou seja, uma previsão de sentença, frase, texto traduzido; ou um valor parcial que é novamente manipulado iterativamente por outro trabalho, ou seja, uma lista de argumentos, texto anotado.

Com base nos resultados obtidos, apresentamos as áreas de pesquisa com maior potencial de crescimento, e mais tradicionais e constantes. Extração, Geração, Ilação foram as categorias com maior aumento de interesse durante os estudos sobre geração de textos legais. A categoria Recuperação foi a mais frequente. Além disso, também identificamos que os estudos selecionados carecem de uma diversidade de conjuntos de dados legais facilmente disponíveis para ajudar a treinar modelos de IA e RC e/ou melhorar outros trabalhos amplamente propostos.

Como limitação, entendemos que a metodologia não garante que cobrimos todos os trabalhos relacionados. Além disso, nosso conjunto inicial teve apenas 44 artigos publicados. Isso poderia levar a um panorama enganoso de geração de texto legal. Assim, mesmo extraindo referências do conjunto inicial e reaplicando-as na metodologia, o conjunto final de artigos selecionados ficou um pouco restrito. Outra limitação foi a dificuldade de classificação dos artigos. Foi difícil encontrar definições comuns entre campos de Ciência da Computação e algoritmos entre diferentes autores e objetivos de estudo. Assim, procuramos minimizar as inconsistências na taxonomia agrupando e eliminando ambiguidades e contradições em relação às áreas de estudo e objetivos principais.

Além de identificar e apresentar a pesquisa atual sobre geração de texto legal e propor uma taxonomia sobre as tarefas de geração de textos legais, nossa motivação com esta revisão também foi abordar e inspirar pesquisadores para novos trabalhos visando melhorar os avanços nas áreas de IA, RC, PLN, e RI para gerar texto legal e aproveitar técnicas com texto natural como dado de entrada dentro da área *LegalTech*.

4 TRABALHOS RELACIONADOS

Este capítulo apresenta e compara os principais trabalhos relacionados ao contexto desta dissertação. Inicialmente, são descritos trabalhos que utilizaram (ou construíram a partir de uma base de dados jurídica) e disponibilizaram *datasets* para tarefas na área de *LegalTech*, considerando um contexto mais amplo de publicações. Em seguida, em um contexto mais fechado de publicações, são descritos e comparados trabalhos brasileiros na área de *LegalTech* em relação ao protótipo proposto nesta dissertação. Por fim, uma tabela comparativa sobre os trabalhos é apresentada, destacando as principais características que os diferem uns dos outros e apontando tópicos em aberto contemplados nesta dissertação.

4.1 Datasets

Durante a consulta à literatura referente a trabalhos que usaram e/ou disponibilizaram *datasets* construídos a partir de bases de dados legais, 17 trabalhos foram considerados relevantes para esta dissertação. Entre os trabalhos selecionados, foi possível classificá-los de acordo com seu escopo principal em 6 categorias: Agnóstico, Anotação de texto, Classificação, Predição de sentença, Pergunta e Resposta, e Sumarização.

A categoria de *datasets* Agnósticos corresponde aos conjuntos de dados usados para pré-treinar modelos maiores, e possibilitam também a construção de soluções de propostas generalistas, permitindo o auxílio em tipos de tarefas variados. Entre os trabalhos encontrados, destacam-se o Iudicium Textum Dataset (ITD) (SOUSA; FABRO, 2019), que é uma base de textos jurídicos em Língua Portuguesa composta por documentos dos acórdãos do Supremo Tribunal Federal brasileiro (STF), tratada e disponibilizada abertamente; e Legifrance (BONNIOL et al., 2020): uma editora jurídica francesa que fornece acesso a códigos legais e decisões jurídicas.

Anotação de texto é a categoria que agrupa o maior número de trabalhos. O objetivo desta categoria é treinar um modelo para anotar cláusulas, seções, e/ou sentenças em um documento legal de acordo com critérios variados como injustiça, estrutura argumentativa etc. A Suprema Corte dos Estados Unidos, *US Supreme Court* (SCOTUS) (SPAETH et al., 2020) é o mais alto tribunal federal dos Estados Unidos da América e geralmente julga apenas os casos mais controversos ou complexos que não foram suficientemente bem resolvidos por tribunais inferiores. A legislação da União Europeia (UE) é publicada no portal EUR-Lex (CHALKIDIS; FERGADIOTIS; ANDROUSOPOULOS, 2021). Todas as leis da UE são anotadas pelo Serviço das Publicações da UE e mantidas com vários conceitos de um tesouro multilíngue intitulado EuroVoc. UNFAIR-ToS (LIPPI et al., 2019) contém 50 Termos de Serviço (ToS) de plataformas online (YouTube, Ebay, Facebook etc) cujo termos potencialmente violam os direitos do usuário de acordo com a lei euro-

peia do consumidor. German Civil Code (BGB) (GLASER; SCEPANKOVA; MATTHES, 2018) é um *dataset* alemão composto por mais de 900 frases de contratos de alugueis com anotação semântica. O *dataset* ICAIL2017 (CHALKIDIS; ANDROUTSOPOULOS; MICHOS, 2017) é um conjunto de 3.500 contratos anotados manualmente com 11 elementos de contrato diferentes.

Treinar um modelo para classificar documentos legais (tipicamente longos) é a principal função da categoria Classificação. O conjunto de dados LEDGAR (TUGGENER et al., 2020b) visa a classificação da provisão de contratos. As cláusulas contratuais vêm de contratos obtidos dos arquivos da Comissão de Valores Mobiliários dos EUA (SEC), que estão disponíveis publicamente na base de dados EDGAR. O *dataset* Overruling (ZHENG et al., 2021a) foi treinado para determinar quando uma sentença anula uma decisão anterior, através de uma tarefa de classificação binária, onde exemplos positivos são sentenças anulatórias e exemplos negativos são sentenças não anuladas extraídas de pareceres jurídicos.

A categoria Predição visa treinar um modelo para prever o resultado de um caso a partir de vários recursos específicos deste caso. O *dataset* ECHR (CHALKIDIS; ANDROUTSOPOULOS; ALETRAS, 2019) é composto por alegações de Estados que violaram as disposições de direitos humanos da Convenção Europeia dos Direitos Humanos (ECHR), ouvidas pelo *European Court of Human Right* (ECtHR) (Tribunal Europeu dos Direitos Humanos).

A categoria Pergunta e Resposta treina um modelo para responder a perguntas ou identificar passagens de um documento alvo que sejam relevantes para uma consulta específica. CaseHOLD (*Case Holdings on Legal Decisions*) (ZHENG et al., 2021b) inclui perguntas de múltipla escolha sobre acervos de processos judiciais dos EUA do corpus de jurisprudência da *Harvard Law Library*, com participações sendo breves resumos de decisões judiciais acompanhadas por decisões referenciadas e relevantes para cada caso. PrivacyQA (RAVICHANDER et al., 2019) é composto por 1.750 perguntas sobre políticas de privacidade de aplicativos móveis e 3.500 anotações relevantes de especialistas. O conjunto de dados Cornell LII vinculam 36.000 parágrafos de opiniões da Suprema Corte a 41.000 links para Constituição dos EUA. *Statutory Reasoning Assessment* (Avaliação de Raciocínio Estatutário)(SARA) (HOLZENBERGER; BLAIR-STANEK; DURME, 2020) é um *dataset* com coleção de regras extraídas do Código da Receita Federal dos EUA e questões de linguagem natural que exigem a aplicação destas regras.

Por fim, a categoria Sumarização é composta por *datasets* com modelos treinados para resumir jargões contratuais complexos ou análises jurídicas. Entre os trabalhos encontrados, Zhong2019 et. al (ZHONG et al., 2019b) apresenta um modelo para sumarização treinado a partir da base de dados *Board of Veterans' Appeals Case* (Casos de Apelação do Conselho de Veteranos) (BVA) dos Estados Unidos. Keymanesh et. al (KEYMANESH; ELSNER; SARTHASARATHY, 2020) treinou o modelo para explorar a sumarização de políticas de *cookies* de 151 empresas, com

seções anotadas a partir da base de dados Terms of Service; Didn't Read (TOS;DR)¹. De modo similar, Manor e Li (MANOR; LI, 2019), em seu trabalho, resumiram 446 contratos em inglês simples com resumos paralelos em nível de seção de texto, utilizando a base de dados TOS;DR e, também, tldrLegal².

A Tabela 4.1 apresenta uma comparação de *datasets* de 17 trabalhos relevantes para esta dissertação, considerando critérios como escopo, tamanho, idioma, e formato. A categoria mais frequente foi Anotação de texto que corresponde a 29.41% (5) dos trabalhos, enquanto a menos frequente foi Predição 5.88% (1). O trabalho que apresentou o *dataset* com maior tamanho, isto é, maior corpus, foi o LEDGAR com 70.000 contratos. O menor *dataset* foi apresentado no trabalho de Manor e Li (MANOR; LI, 2019) com 446 contratos. O idioma mais comum entre as bases de dados é o inglês, que foi utilizado em 82.35% (14) dos trabalhos. Por sua vez, alemão, francês, e português foram os menos utilizados, aparecendo uma vez (5.88%). O formato mais recorrente de divulgação dos *datasets* é CSV, presente em 41.17% (7) dos trabalhos. Os formatos RTF e TXT foram os menos frequentes, aparecendo uma vez apenas (5.88%).

Tabela 4.1: Tabela comparativa dos *datasets* jurídicos públicos selecionados.

Dataset	Escopo	Tamanho	Idioma	Formato
ITD	Agnóstico	41.353	Português	JSON
Légifrance	Agnóstico	40.000	Francês	RTF
BGB	Anotação	913	Alemão	CSV
EUR-LEX	Anotação	60.000	Inglês	JSONL
ICAIL2017	Anotação	12.000	Inglês	CSV
SCOTUS	Anotação	6.400	Inglês	CSV
UNFAIR-TOS	Anotação	12.000	Inglês	JSON
LEDGAR	Classificação	70.000	Inglês	JSON
Overruling	Classificação	37.018	Inglês	CSV
ECHR	Predição	11.500	Inglês	JSONL
CaseHOLD	QA	53.000	Inglês	CSV
Cornell LII	QA	36.000	Inglês	JSON
PrivacyQA	QA	1.750	Inglês	CSV
SARA	QA	1.402	Inglês	PL, TXT
BVA	Sumarização	27.903	Inglês	JSON
TOSDR	Sumarização	5.752	Inglês	CSV
TOSDR & TldrLegal	Sumarização	446	Inglês	JSON

Fonte: o autor

¹<<https://tosdr.org/>>

²<<https://tldrlegal.com/>>

4.2 Trabalhos brasileiros de *LegalTech*

Ao pesquisar na literatura trabalhos brasileiros relacionados à área de *LegalTech* que se aplicavam ao contexto desta dissertação, 3 estudos foram selecionados. Diferentemente da seção anterior, onde Sousa e Fabro (SOUSA; FABRO, 2019) disponibilizaram o *dataset* jurídico Iucidium Textum Dataset, nesta seção os trabalhos aplicam técnicas de IA e PLN e provêm resultados sobre seus experimentos.

Todos os 3 estudos selecionados estão relacionados ao projeto VICTOR³, e são produtos de uma parceria entre o STF e a Universidade de Brasília (UnB). O projeto é um importante marco no Judiciário brasileiro e referência no cenário internacional, por seu pioneirismo na aplicação de IA em desafios para maior eficiência processual. Iniciado no final de 2017, na gestão da Ministra Cármen Lúcia na presidência da Corte, VICTOR foi idealizado para auxiliar o STF na análise dos recursos extraordinários recebidos de todo o país, especialmente quanto a sua classificação em temas de repercussão geral de maior incidência.

Araujo et. al, em seu trabalho LeNER-Br (ARAUJO et al., 2018), propõem um *dataset* para NER composto inteiramente de documentos legais manualmente anotados. O *dataset* contém categorias para pessoas, localizações, entidades temporais, organizações, legislações e casos legais que possibilitam extrair conhecimento dos textos em linguagem natural. Todo o trabalho foi realizado através de 70 documentos legais obtidos do STF, Superior Tribunal de Justiça, Tribunal de Justiça de Minas Gerais, Tribunal de Contas da União, e Lei Maria da Penha.

Os trabalhos de Braz, Silva, Campos (SILVA; BRAZ; CAMPOS, 2018) e, posteriormente, Braz et al. (BRAZ et al., 2018) utilizaram um *dataset* de 6814 peças extraídos exclusivamente do STF. Ambos os trabalhos tinham como objetivo analisar técnicas de IA para a tarefa de classificação de documentos que eram tratados pelo STF. Os documentos eram classificados em 6 possíveis categorias: Acórdão, Recurso Extraordinário, Agravo de Recurso Extraordinário, Despacho, Sentença, ou Outros. O primeiro trabalho propôs uma solução através de uma arquitetura simples de CNN para auxiliar em uma ferramenta de anotação de texto manualmente utilizado por advogados, com score F1 de 91%. O segundo trabalho utilizou um algoritmo de *Long Short-Term Memory* bidirecional com o mesmo objetivo; porém obteve um score F1 de 84%.

A Tabela 4.2 apresenta a comparação entre os trabalhos brasileiros citados e o protótipo desenvolvido na abordagem desta dissertação. A tabela mostra os trabalhos considerando as bases jurídicas consultadas para a construção do *dataset*, se o *dataset* construído é disponibilizado no trabalho, quais algoritmos e soluções foram utilizados, e o número de documentos legais que compõem o *dataset* utilizado para treinamento, testes e validação dos algoritmos.

³<<https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=471331&ori=1>>

Tabela 4.2: Tabela comparativa de trabalhos brasileiros na área de LegalTech.

Autor	Bases	Dataset	Técnicas	Tamanho
(ARAUJO et al., 2018)	STF, STJ, TJMG, TCU, LMP	Não	NER	70
(BRAZ et al., 2018)	STF	Não	Bi-LSTM	6814
(SILVA; BRAZ; CAMPOS, 2018)	STF	Não	CNN	6814
(SOUSA; FABRO, 2019)	STF	Sim	-	41353
Nosso trabalho	STF	Sim	NB, SVM, RF, NER	73798

Fonte: o autor

4.3 Considerações finais

Este capítulo apresentou dois contextos de trabalhos relacionados: *datasets* construídos e disponibilizados a partir de bases de dados legais, e trabalhos brasileiros com técnicas para auxílio de geração de texto legal em alguma instância. Inicialmente, foram apresentados 17 trabalhos que disponibilizaram *datasets* em suas pesquisas. Todos os *datasets* foram descritos, comparados e classificados de acordo com seu escopo original. Por fim, o capítulo também fez uma análise e comparação entre os 4 trabalhos brasileiros de *LegalTech* e o protótipo desenvolvido a partir da abordagem proposta nesta dissertação.

Desta forma, este trabalho avança na literatura com a disponibilização de uma base de dados brasileira legal e de um protótipo que aplica técnicas de IA e PLN para tarefas de Geração, através de anotação de texto, e Ilação para classificação de documentos legais.

5 UMA ABORDAGEM AGNÓSTICA PARA GERAÇÃO DE TEXTOS LEGAIS

Neste capítulo, é apresentada a contribuição principal desta dissertação: uma Abordagem Agnóstica para a Geração de textos legais em linguagem natural. A Abordagem Agnóstica é interoperável entre os sistemas e não há restrições quanto ao uso de uma tecnologia, modelo, metodologia, ou dados específicos. Para possibilitar a comunicação entre diferentes sistemas, um módulo para entrada e saída de dados é estabelecido. O módulo é independente para cada etapa da Abordagem Agnóstica, e é de sua responsabilidade a interpretação das informações consumidas.

A Abordagem Agnóstica, embora tenha caráter generalista, pois fornece uma estrutura para a interoperabilidade de soluções, é desenvolvida com o objetivo de gerar valor a um domínio específico. No contexto desta dissertação, a Abordagem Agnóstica deve facilitar o desenvolvimento de soluções para tarefas de *LegalTech*. Ao analisar a literatura, em nenhum dos trabalhos foi apresentada, até onde sabemos, uma Abordagem Agnóstica para alcançar determinada solução na área de *LegalTech*. Esta dissertação visa preencher esta lacuna e guiar o desenvolvimento de soluções em *LegalTech* através de recomendações e estratégias que possibilitem maior facilidade e manutenibilidade ao desenvolver sistemas robustos para o referido domínio.

As seções a seguir descrevem a Abordagem Agnóstica a partir de uma visão geral e, posteriormente, uma visão detalhada de cada uma das etapas que a constitui.

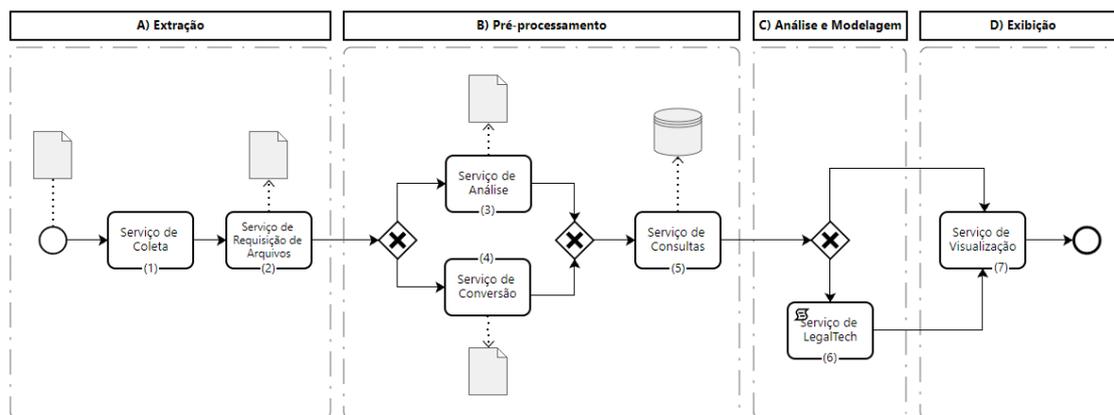
5.1 Visão geral

A Figura 5.1 ilustra a arquitetura da abordagem proposta para a execução de tarefas para geração de textos legais a partir de textos em linguagem natural. A Abordagem Agnóstica consiste em 4 etapas principais: (A) Extração, (B) Pré-processamento, (C) Análise e Manipulação, e (D) Exibição. Para que todas as etapas sejam satisfeitas, a abordagem é composta por 7 serviços que contribuem para a realização de cada etapa. Os serviços são:

1. *Serviço de Coleta*
2. *Serviço de Requisição de Arquivos*
3. *Serviço de Análise*
4. *Serviço de Conversão*
5. *Serviço de Consultas*
6. *Serviço de LegalTech*
7. *Serviço de Visualização*

Nas seções seguintes, é descrita cada uma das etapas da abordagem, bem como todos os serviços que as compõem.

Figura 5.1: Arquitetura para execução de tarefas para geração de texto legal.



Fonte: o autor

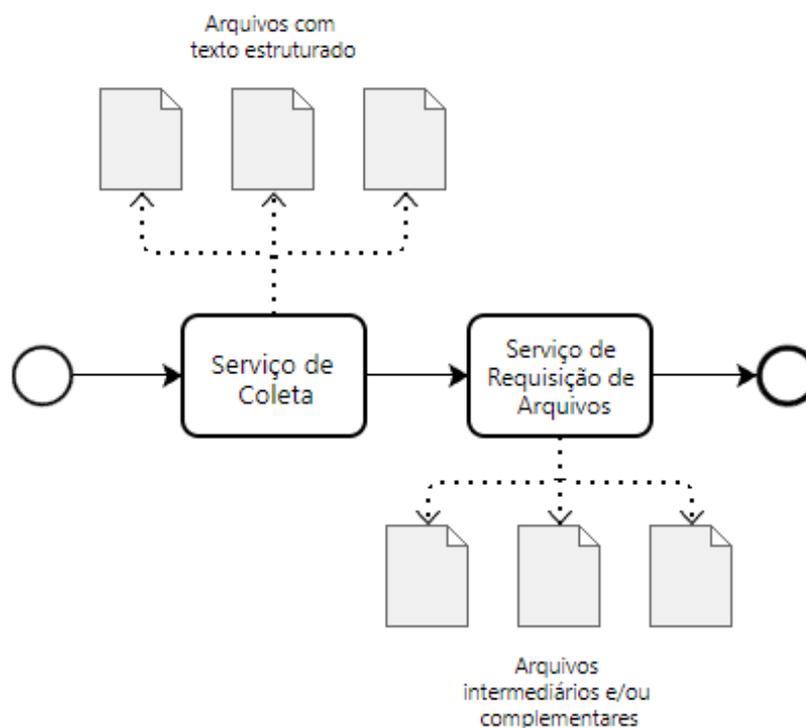
5.2 Extração

A Extração consiste na obtenção de informações a partir de textos em linguagem natural de diversas fontes (páginas *Web*, e-mails, documentos textuais etc). Os dados extraídos podem ser diretamente em formato de texto, ou, ainda, em um formato de arquivo intermediário que posteriormente deve ser transformado em texto. A etapa de Extração é composta por dois serviços: *Serviço de Coleta* e *Serviço de Requisição de Arquivos*. A Figura 5.2 apresenta uma visão detalhada da etapa de Extração.

Serviço de Coleta é o primeiro serviço e é responsável por extrair informações de uma base de dados. O parâmetro de entrada do serviço é uma base de dados que pode ser composta por páginas e/ou endereços na *Web*, arquivos de texto, imagens, ou qualquer documento digitalizado onde informações textuais podem ser obtidas. É de responsabilidade do serviço também identificar e filtrar quais informações são elegíveis e de interesse para as etapas posteriores. Ao final da execução do serviço, a saída esperada é um arquivo (ou um conjunto) de texto estruturado, facilitando a leitura de máquina. Arquivos no formato XML, HTML, JSON, YAML, e semelhantes são exemplos de saída apropriados.

Serviço de Requisição de Arquivos é um serviço opcional e complementar ao *Serviço de Coleta*. Seu objetivo é ajudar em requisições de *download* para extração de arquivos intermediários e/ou complementares. Algumas bases de dados apresentam apenas um texto parcial da informação de interesse, mas anexos com arquivos contendo o texto completo. Assim, para não sobrecarregar um serviço com múltiplas tarefas, idealmente devem existir dois serviços: um para extração de dados e outro para extração de arquivos. É válido ressaltar que, em caso de extração de arquivos, é necessária uma etapa de processamento dos dados destes arquivos de modo que os dados estejam padronizados, independente da fonte de origem.

Figura 5.2: Etapa de Extração.



Fonte: o autor

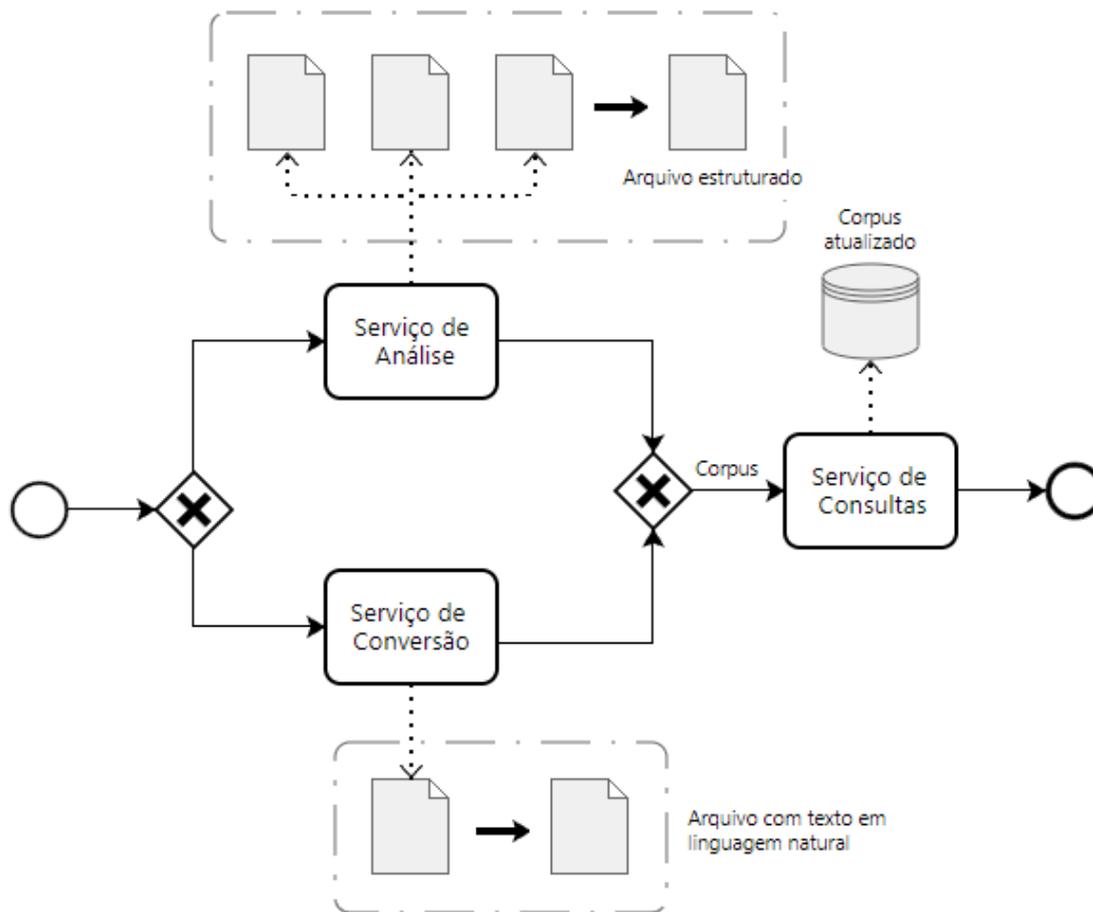
5.3 Pré-processamento

O Pré-processamento objetiva organizar as informações extraídas na etapa anterior e estruturar os dados orientados à leitura de máquina, criando, assim, um corpus¹. Adicionalmente, se algum arquivo intermediário foi extraído, ele deve ser convertido para texto e pré-processado também. Ao final dessa etapa, o corpus deve ser organizado em algum banco de dados ou sistema de arquivos que possibilite análises e modelagens dos dados. Esta etapa de Pré-processamento possui três serviços, sendo eles: *Serviço de Análise*, *Serviço de Conversão*, *Serviço de Consultas*. A Figura 5.3 apresenta uma visão detalhada da etapa de Pré-processamento da abordagem.

Serviço de Análise é o serviço principal para a etapa de Pré-processamento de dados. O propósito deste serviço é organizar todos os arquivos estruturados da etapa de extração e aplicar os ajustes finais para geração do corpus. Se a saída do Serviço de Coleta tiver gerado múltiplos arquivos estruturados, o Serviço de Análise pode unificar todos os dados em um arquivo único, facilitando as tarefas posteriores desta etapa. Adicionalmente, correções pontuais de texto em linguagem natural no arquivo são realizados neste serviço bem como a criação de vínculos com textos extraídos de arquivos intermediários e/ou complementares. Ao final da execução do serviço, a abordagem garantidamente terá um módulo apto para o serviço de criação do corpus.

¹Corpus é uma coletânea ou conjunto de documentos sobre determinado tema.

Figura 5.3: Etapa de Pré-processamento de dados da abordagem.



Fonte: o autor

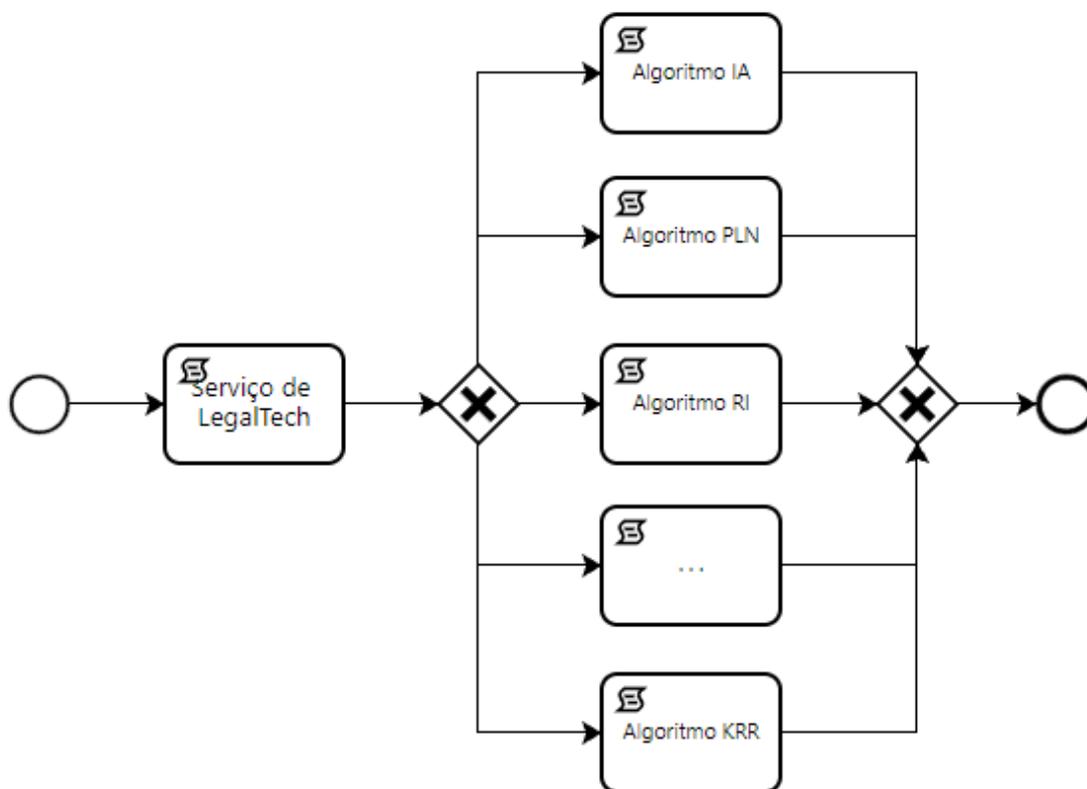
Serviço de Conversão é um serviço auxiliar para transformação de um arquivo de formato genérico para formato textual. O serviço deve garantir que anexos com informações complementares para um dado registro também tenham informações processadas e, dessa forma, produza dados estruturados e orientados à leitura de máquina, aumentando o volume de texto para processamento.

Serviço de Consultas é o último serviço executado nesta etapa. Ele deve organizar os dados processados do Serviço de Análise e Serviço de Conversão para construir um corpus orientado à etapa de Análise e Modelagem. O corpus pode ser armazenado em um sistema de gerenciamento de banco de dados, caso a implementação priorize flexibilidade, gerência, e robustez dos dados; ou sistema de arquivos, em caso de performance. É de responsabilidade do Serviço de Consultas garantir a consistência dos dados, bem como conexões com banco de dados e configurações necessárias para a interface dos dados entre as plataformas. Ao final desta etapa, a abordagem deve possuir um corpus para analisar e modelar dados.

5.4 Análise e Modelagem

A etapa de Análise e Modelagem tem a responsabilidade de aplicar técnicas de PLN, IA, KRR, entre outras, que constituem a área de *LegalTech*. Esta etapa é fundamental para tarefas orientadas à utilização de alguma inteligência sobre os dados, em especial argumentação e ilação, e no auxílio do aperfeiçoamento de outras tarefas como recuperação de informações. Esta etapa possui um único serviço denominado *Serviço de LegalTech*. A Figura 5.4 apresenta uma visão detalhada da etapa de Análise e Modelagem.

Figura 5.4: Etapa de Análise e Modelagem.



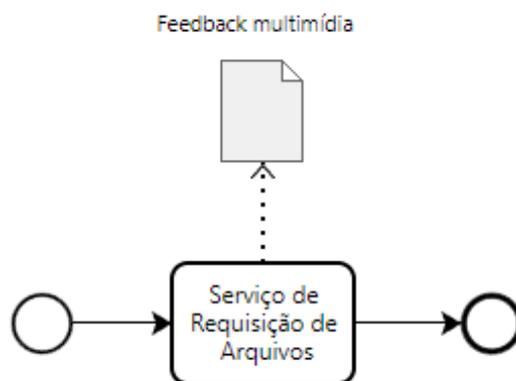
Fonte: o autor

O Serviço de *LegalTech* é um serviço abstrato em alto nível cujo objetivo é ser decomposto e instanciado por um ou mais serviços auxiliares de acordo com o contexto desejado. Assim, é possível aumentar o nível de abstração e reduzir a complexidade de toda abordagem. Os serviços podem incorporar técnicas de uma ou mais áreas de IA, PLN, RI, KRR, entre outras de *LegalTech*. Contudo, o módulo de comunicação deve se basear no corpus como valor de entrada e a saída deve ser armazenada em um arquivo estruturado que possibilite a interpretação na etapa de Exibição.

5.5 Exibição

Por fim, a última etapa, *Exibição*, é responsável pela apresentação dos dados analisados e/ou modelos da etapa de Análise e Modelagem. Através da exibição dos dados, pelo *Serviço de Visualização*, o usuário final pode ter um *feedback* visual do resultado obtido ao longo de todo o processo de geração de texto legal. A Figura 5.5 apresenta uma visão detalhada da etapa de Exibição.

Figura 5.5: Etapa de Exibição.



Fonte: o autor

O Serviço de Visualização é o último serviço e sua função é apresentar os resultados obtidos na etapa de Análise e Modelagem. A apresentação é através de um *feedback* multimídia, isto é, pode ser um arquivo de áudio, imagem, ou um hiperdocumento.

O critério ou a forma de apresentação dos dados é de responsabilidade da instância do serviço. Entretanto, idealmente, análises estatísticas através de gráficos, ou a construção de uma camada de visão (pensando em um conceito *Model-View-Controller*), por exemplo, são estratégias interessantes para o usuário final receber um *feedback* visual do resultado de toda a abordagem realizada.

6 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

Este capítulo inicialmente descreve o protótipo LEAF (LEgal Agnostic Framework) desenvolvido como alternativa de implementação da Abordagem Agnóstica proposta nesta dissertação, para as tarefas de geração de texto legal e classificação de documentos. Em seguida, é apresentada a Análise Descritiva do *dataset* construído a partir das informações coletadas em páginas do Supremo Tribunal Federal (STF). Por fim, através do protótipo LEAF, são avaliados os experimentos para as tarefas de classificação de documentos legais, onde são consideradas métricas como F1-score e tempo de execução.

6.1 LEAF: um protótipo de framework agnóstico legal

Esta Seção descreve o protótipo para a Abordagem Agnóstica. Inicialmente, é apresentada uma visão geral do protótipo além das interfaces de comunicação entre cada uma das etapas da Abordagem Agnóstica e, posteriormente, as soluções escolhidas para trabalhar com documentos legais utilizando algoritmos de IA e PLN. Os algoritmos de IA e PLN foram escolhidos devido à grande quantidade de soluções que aplicam estas técnicas tanto na literatura quanto na indústria. Diversos trabalhos relacionados, por exemplo, utilizaram algoritmos destes campos da Ciência da Computação.

6.1.1 Visão geral

A fim de demonstrar a viabilidade da realização das etapas propostas pela Abordagem Agnóstica, foi construído o protótipo LEAF. O protótipo objetiva auxiliar em tarefas de geração de texto legal e foi desenvolvido utilizando as linguagens de programação TypeScript¹ e Python². Ambas as linguagens foram escolhidas devido às bibliotecas auxiliares que trivializam a complexidade de tarefas em que as linguagens se propõe a solucionar, facilitando a implementação dos serviços relacionados. Para o corpus do experimento, foi construído um *dataset* a partir da extração de acórdãos do Supremo Tribunal Federal (STF) brasileiro³.

A escolha para extração de dados em páginas STF se deve ao trabalho relacionado de (SOUSA; FABRO, 2019) que construiu o Iudicium Textum Dataset. Assim, o protótipo LEAF estende a quantidade de registros do *dataset* (de 41353 para 73798) além de incluir acórdãos mais recentes. Embora não tenham sido incluídos no corpus utilizado para os serviços de *LegalTech*,

¹<https://www.typescriptlang.org/>

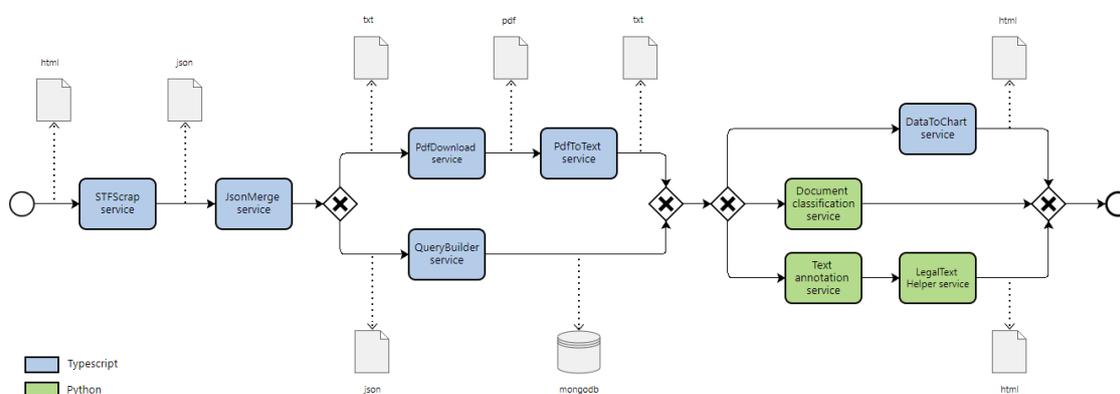
²<https://www.python.org/>

³<https://portal.stf.jus.br/>

foram extraídos todos os acórdãos presentes no site do STF, do período de janeiro de 1892 à agosto de 2022, totalizando 274245 registros.

Todo o código desenvolvido para o protótipo LEAF, além do dataset construído, está disponível no seguinte repositório GitHub: <<https://github.com/jaflesch/legal-agnostic-framework>>. Adicionalmente, no mesmo repositório, todos os dados coletados e pré-processados, e as coleções do banco de dados estão disponíveis para *download* mediante realização de solicitação do material, de acordo com as instruções na página, de forma a seguir as recomendações da LGPD⁴. A Figura 6.1 apresenta uma visão geral dos serviços do protótipo, com destaque para os formatos de arquivos de entrada e saída de dados.

Figura 6.1: Serviços que constituem o protótipo LEAF.



Fonte: o autor

A extração de dados foi realizada através do *STFScrap service*, utilizando a biblioteca Puppeteer⁵ que fornece uma API de alto nível para controlar navegadores baseados em Chromium, por meio do protocolo DevTools⁶. Ao recorrer a funcionalidades da biblioteca -como navegar entre páginas, clicar em botões e abrir novas janelas-, foi possível extrair todo o conteúdo HTML de elementos de interesse das páginas do STF. Assim, o *STFScrap service* do protótipo consegue percorrer todos os resultados de acórdãos mais recentes dos últimos 10 anos (08 de agosto de 2012 a 8 de agosto de 2022). A Figura 6.10 exibe a estrutura de um acórdão no site do STF cujas informações foram extraídas pelo scraper.

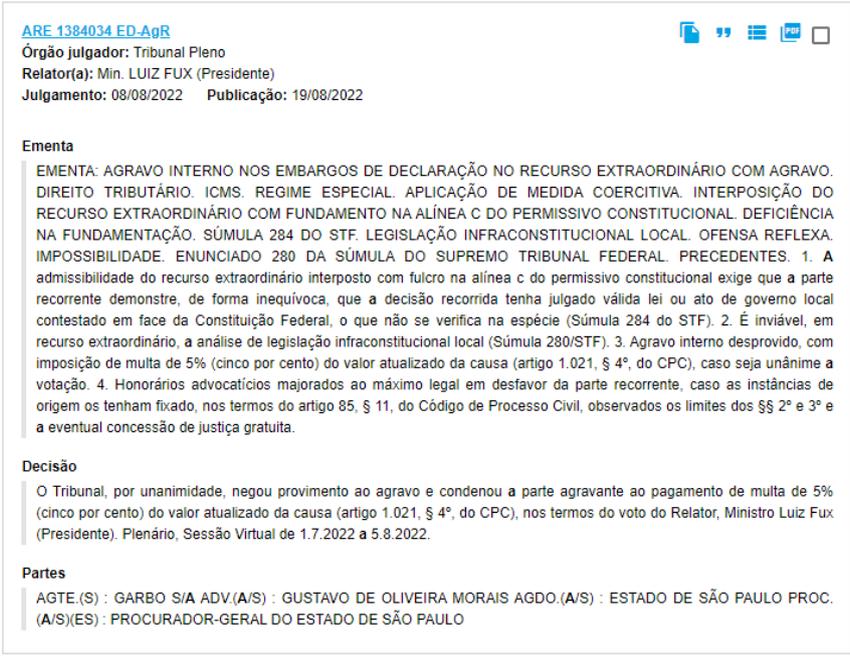
Ao final da execução do serviço, para cada página percorrida, o protótipo gera um arquivo JSON, com diversas informações coletadas a partir do HTML, mas estruturadas em formato "chave-valor", para facilitar a leitura de dados pelos serviços posteriores. A Tabela 6.1 apresenta um resumo do tipo das informações coletadas.

⁴<https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm>

⁵<<https://github.com/puppeteer/puppeteer>>

⁶<<https://chromedevtools.github.io/devtools-protocol/>>

Figura 6.2: Exemplo de informações exibidas de acórdãos em páginas do STF.



[ARE 1384034 ED-AgR](#)

Órgão julgador: Tribunal Pleno

Relator(a): Min. LUIZ FUX (Presidente)

Julgamento: 08/08/2022 Publicação: 19/08/2022

Ementa

EMENTA: AGRAVO INTERNO NOS EMBARGOS DE DECLARAÇÃO NO RECURSO EXTRAORDINÁRIO COM AGRAVO. DIREITO TRIBUTÁRIO. ICMS. REGIME ESPECIAL. APLICAÇÃO DE MEDIDA COERCITIVA. INTERPOSIÇÃO DO RECURSO EXTRAORDINÁRIO COM FUNDAMENTO NA ALÍNEA C DO PERMISSIVO CONSTITUCIONAL. DEFICIÊNCIA NA FUNDAMENTAÇÃO. SÚMULA 284 DO STF. LEGISLAÇÃO INFRACONSTITUCIONAL LOCAL. OFENSA REFLEXA. IMPOSSIBILIDADE. ENUNCIADO 280 DA SÚMULA DO SUPREMO TRIBUNAL FEDERAL. PRECEDENTES. 1. A admissibilidade do recurso extraordinário interposto com fulcro na alínea c do permissivo constitucional exige que a parte recorrente demonstre, de forma inequívoca, que a decisão recorrida tenha julgado válida lei ou ato de governo local contestado em face da Constituição Federal, o que não se verifica na espécie (Súmula 284 do STF). 2. É inviável, em recurso extraordinário, a análise de legislação infraconstitucional local (Súmula 280/STF). 3. Agravo interno desprovido, com imposição de multa de 5% (cinco por cento) do valor atualizado da causa (artigo 1.021, § 4º, do CPC), caso seja unânime a votação. 4. Honorários advocatícios majorados ao máximo legal em desfavor da parte recorrente, caso as instâncias de origem os tenham fixado, nos termos do artigo 85, § 11, do Código de Processo Civil, observados os limites dos §§ 2º e 3º e a eventual concessão de justiça gratuita.

Decisão

O Tribunal, por unanimidade, negou provimento ao agravo e condenou a parte agravante ao pagamento de multa de 5% (cinco por cento) do valor atualizado da causa (artigo 1.021, § 4º, do CPC), nos termos do voto do Relator, Ministro Luiz Fux (Presidente). Plenário, Sessão Virtual de 1.7.2022 a 5.8.2022.

Partes

AGTE(S) : GARBO S/A ADV.(A/S) : GUSTAVO DE OLIVEIRA MORAIS AGDO.(A/S) : ESTADO DE SÃO PAULO PROC.(A/S)(ES) : PROCURADOR-GERAL DO ESTADO DE SÃO PAULO

Fonte: o autor, adaptado de <https://portal.stf.jus.br/>

Tabela 6.1: Tabela com as informações coletadas e o tipo de dado atribuídos pelo STFS-crap service.

Nome	Descrição	Tipo de dado
id	Chave primária do acórdão	String
docId	Chave primária do documento PDF do acórdão	Número
pdfFileUrl	Endereço do do documento PDF do acórdão	String
innerPageUrl	Endereço para página interna do acórdão	String
innerHTML	Código HTML do acórdão na página geral de resultados	String
innerPageHTML	Código HTML da página interna do acórdão	String

Fonte: o autor

Ainda durante a extração, um serviço auxiliar, *JsonMerge service*, é responsável por unificar todos os JSONs paginados. O serviço recebe uma lista dos JSONs criados durante o serviço anterior e como saída gera um JSON unificado. Essa estratégia foi necessária devido a instabilidades dos servidores do STF, e uma melhor gerência do volume de dados enquanto o *scrap* era realizado. Desta forma, com um único arquivo para iterar sobre os dados, a manipulação para as etapas anteriores é trivializada.

A partir do arquivo unificado, são realizadas duas tarefas em paralelo: extração de PDFs, e importação dos dados para um banco de dados. O serviço *PdfDownload service* realiza o *down-*

load de todos os arquivos PDFs a partir da informação de URL salva no JSON. Logo, ao final da execução do serviço, a etapa de extração é finalizada no protótipo.

A importação dos dados é realizada através do *QueryBuilder service*. Como os dados são salvos em uma tabela única, sem relações, foi escolhido o banco de dados NoSQL MongoDB⁷ para armazenar a coleção de acórdãos. Após a execução deste serviço, o MongoDB criará uma coleção `judgements` com as colunas apresentadas na Tabela 6.2. Além da importação para o MongoDB, a etapa de préprocessamento é responsável por continuar o tratamento de dados dos PDFs coletados. Através do *PdfToText service*, todo o conteúdo textual dos arquivos PDFs é convertido e salvo em arquivos TXT, facilitando a utilização das informações dos acórdãos.

A análise e modelagem dos dados no protótipo objetiva demonstrar a viabilidade do *dataset* construído para o uso de técnicas de IA e PLN. Para tal, o protótipo realiza duas tarefas de LegalTech: Geração (de texto legal) e Ilação. A tarefa de Geração é atendida através do *Text Annotation service*, e a tarefa de Ilação através do *Document classification service*. As seções 6.1.2 e 6.1.3 apresentam detalhadamente a implementação dos serviços para as tarefas de Geração e Ilação, respectivamente.

Por fim, para a etapa de Exibição, são propostos dois serviços de visualização das informações: *DataToChart service* e *LegalTextHelper service* que será discutido na próxima Seção. *DataToChart service* foi construído com o *framework* Express.js⁸, que fornece recursos mínimos para construção de servidores *Web*, e apresenta ao usuário uma página HTML com diversas estatísticas e gráficos dos registros do banco de dados. É possível, por exemplo, analisar o número de acórdãos, atividade dos Ministros, categorias dos documentos, decisões proferidas, Órgão julgador, localização entre outros, filtrando e agrupando resultados por intervalo de tempo. A Figura 6.3 apresenta a tela inicial do serviço com as páginas e seus respectivos gráficos disponíveis. A Seção 6.2 apresenta uma Análise Descritiva dos dados mais relevantes do *dataset*.

⁷<<https://www.mongodb.com/>>

⁸<<https://expressjs.com/>>

Tabela 6.2: Tabela com as colunas da coleção de acórdãos importados ao MongoDB.

Nome	Descrição	Tipo de dado
id	Chave primária do acórdão	String
titulo	Título do acórdão	String
origem	Federação originária do acórdão	String
orgao	Órgão responsável pelo julgamento	String
categoria	Categoria do acórdão	String
relator	Ministro do STF relator do processo	String
relatorPresidente	Ministro relator também era presidente	Booleano
redator	Ministro do STF redator do processo	String
dataJulgamento	Data em que ocorreu o julgamento	Data
dataPublicacao	Data de publicação do julgamento	Data
ementa	Síntese do acórdão	String
partes	Pessoas que participam do processo	Array<String>
tese	Entendimento sobre a discussão	String
tema	Definição de tese aplicada a processos com idêntica questão de direito discutida	String
doutrina	Princípios e ideias de autores e juristas que influenciam e fundamentam as decisões	String
legislacao	Leis que estabelecem condutas aceitáveis ou não das partes do processo	String
similares	Acórdãos com características similares	String
decisao	Decisão deferida pelo Órgão julgador	String
decisaoAdiada	Registro de adiamento de decisão	Booleano
decisaoUnanime	Decisão com votação unânime de ministros	Booleano
indexacao	Lista de índices do acórdão	Array<String>
observacao	Comentários adicionais do acórdão	String
paginaInternaUrl	URL página interna do acórdão	String
paginaInternaTitulo	Título da página interna do acórdão	String
paginaInternaSubtitulo	Subtítulo da página interna do acórdão	String
paginaInternaPublicacao	Informações adicionais sobre publicação no Diário de Justiça Eletrônico (DJe)	String
documentId	Identificador do documento em PDF	Número
arquivoPdfUrl	URL página do PDF do acórdão	String

Fonte: o autor

Figura 6.3: Tela inicial do DataToChart service exibindo todas as subpáginas disponíveis para consulta de dados a partir de diferentes critérios.

The screenshot displays the DataToChart service interface, organized into three main sections: Acórdãos, Decisões, and Localizações. Each section contains several interactive cards, each with a chart icon, a title, a description, and a 'Ver dados' button.

Acórdãos

- Acórdãos / ano**: Total de acórdãos por ano (ou período)
- Acórdãos / relator**: Acórdãos agrupados por relatores
- Acórdãos / relator / ano**: Acórdãos de um relator agrupados por ano
- Acórdãos / redator**: Acórdãos agrupados por relatores
- Acórdãos / redator / ano**: Acórdãos de um redator agrupados por ano
- Acórdãos / Órgão**: Acórdãos agrupados pelo Órgão julgador
- Acórdãos / Órgão / ano**: Acórdãos agrupados por Órgão julgador e ano
- Acórdãos / categoria**: Acórdãos agrupados por tipo de documento
- Acórdãos / categoria / ano**: Grupos por tipo de documento e ano

Decisões

- Tipo de decisão**: Maioria de votos ou unanimidade
- Tipo de decisão / ano**: Decisão por tipo agrupadas por ano
- Adiamentos**: Votação e/ou decisão adiados ou não
- Adiamentos / ano**: Adiamentos agrupados por ano

Localizações

- Acórdãos / federação**: Exibe acórdãos por estados brasileiros
- Acórdãos / Estado**: Exibe acórdãos por país relacionado

6.1.2 Anotação de texto

A tarefa de *LegalTech* de geração de texto é realizada pelo *Text Annotation service* do protótipo LEAF. O serviço utiliza uma técnica de anotação de texto para, a partir de um texto jurídico, obter palavras classificadas e marcadas para utilização em etapas posteriores. A anotação de texto é efetuada com os resultados obtidos pelo modelo LeNER-Br (ARAUJO et al., 2018). O modelo LeNER-Br implementa algoritmos de NER, LSTM, e CRF e, ao final da sua execução, gera uma lista de palavras classificadas em "PESSOA", "TEMPO", "LOCAL", "ORGANIZACAO", "JURISPRUDENCIA", "LEGISLACAO", ou "O" caso não se enquadre nas classes anteriores. Assim, o *Text Annotation service* funciona como uma interface que encapsula o LeNER-Br, ajustando os valores de entrada e saída do modelo para obter os resultados no formato esperado.

Após obter a lista de palavras classificadas e salvá-las em um arquivo de *output* auxiliar, o serviço é finalizado. A partir deste momento, o protótipo pode utilizar um serviço auxiliar de visualização para exibição dos resultados de forma mais legível.

O *LegalText Helper service* foi criado com uma proposta simples: tornar mais acessível a informação dos termos jurídicos utilizados nos textos, uma vez que estes podem ser difíceis de compreender para pessoas leigas na área. Através do *microframework* para aplicações *Web Flask*⁹ e da *engine* de *templates Jinja*¹⁰, o serviço apresenta ao usuário uma página HTML, a partir do texto original, com texto marcado, isto é, destacando com diferentes cores as palavras que pertencem às classes geradas pelo modelo LeNER-Br.

A Figura 6.4 apresenta um exemplo de saída gerada pelo serviço. A partir de um texto legal extraído de um acórdão do STF, o serviço apresenta, em elementos HTML coloridos, as diferentes classes de palavras após o modelo LeNER-Br classificou. A associação entre cores e classes do modelo é definida como:

- Vermelho: PESSOA;
- Laranja: LEGISLACAO;
- Amarelo: LOCAL;
- Verde: TEMPO;
- Azul: ORGANIZACAO;
- Roxo: JURISPRUDENCIA.

⁹<https://flask.palletsprojects.com/en/2.2.x/>

¹⁰<https://jinja.palletsprojects.com/en/3.1.x/>

Figura 6.4: Exemplo de página Web com texto jurídico anotado e marcações em palavras pertencentes a determinadas classes do modelo.



Fonte: o autor

6.1.3 Classificação de documentos

A segunda tarefa de *LegalTech* que o protótipo realiza é a classificação de documentos legais através de modelos preditivos. No *dataset* existem 69 classes de acórdãos como Ação Cautelar, Denúncia, Habeas Corpus, Inquérito Policial, Petição, Recurso Extraordinário entre outros.

Os acórdãos extraídos já possuem a classe do documento e, portanto, o treinamento com modelos de IA possuem valores de referência para aferir a precisão da solução obtida.

Diferente dos serviços anteriores, para a execução desta tarefa, o protótipo agrupa 3 algoritmos diferentes dentro de um serviço maior, o *Document Classification service*, responsável apenas por gerenciar as rotinas de execução de cada método. Cada algoritmo utiliza um classificador e modelo diferentes dos demais, possibilitando comparar métricas como precisão, relevância, e tempo de execução, além de estatísticas do corpus e *stopwords*.

Para a tarefa de classificação, os modelos foram construídos através da implementação dos algoritmos NB, SVM, e RF. Cada solução utilizada possui um arquivo IPython Notebook¹¹ associado a esta, que apresenta resultados parciais entre cada etapa do *benchmark* realizado.

A Seção 6.3 apresenta os Experimentos e Análises de Resultados, detalhando os *benchmarks* construídos e compara o desempenho entre os três modelos de classificação em critérios como F1-score e tempo de execução.

6.2 Análise Descritiva

A Análise Descritiva é a fase inicial do estudo das informações coletadas no acórdãos do STF, pois os dados são organizados para resumir, descrever, e comparar características importantes observadas. Adicionalmente, através de visualizações gráficas, a Análise Descritiva pode evidenciar certos padrões, evoluções e comportamentos atípicos nos dados, auxiliando no desenvolvimento de estratégias e soluções para os resultados analisados.

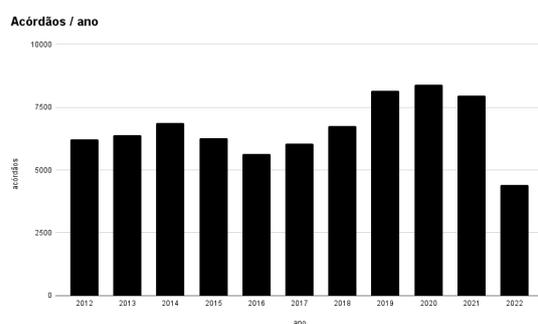
Nesta seção, são exibidas relações entre o conjunto de informações extraídos dos acórdãos das páginas do STF em relação a datas de julgamento, tipos de decisão, Órgãos julgadores, classes de documentos, e origem do julgamento.

A Figura 6.5 apresenta a distribuição das datas de julgamento dos acórdãos em dois intervalos de tempo. O primeiro intervalo considera os acórdãos de janeiro de 2012 à agosto de 2022 (Figura 6.5(a)). O ano com o menor número de acórdãos foi 2016 (isto é, 7.71% do total), enquanto o maior, 2020 (11.48%). O segundo intervalo expande o período considerando todos os acórdãos do *dataset*, ou seja, de janeiro de 1893 à agosto de 2022. A Figura (Figura 6.5(b)), apresenta a distribuição de acórdãos por ano a partir de 1934. Acórdãos registrados antes de 1934 foram omitidos no gráfico, pois o total de registros desta data totalizavam 38 registros. É interessante analisar a quantidade de acórdãos no sistema em diferentes épocas a fim de encontrar relações com eventos nacionais e internacionais. Entre os anos de 1950 a 1968 houve um aumento na tendência de acórdãos julgados, com uma média de 4135 acórdãos por ano. Já entre os anos de 1975 à 1990, houve uma diminuição na tendência de acórdãos julgados, com uma média de 2019

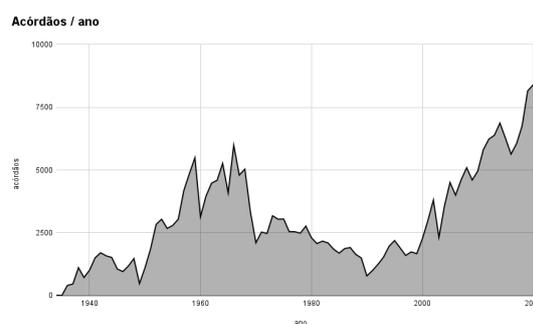
¹¹<https://ipython.org/notebook.html>

acórdãos por ano. Apenas a partir de 2005 até os dias atuais, a tendência voltou a subir, com uma média superior a 6 mil acórdãos por ano.

Figura 6.5: Quantidade de acórdãos por período. (a) De 2012 à agosto de 2022, (b) de 1934 à agosto de 2022.



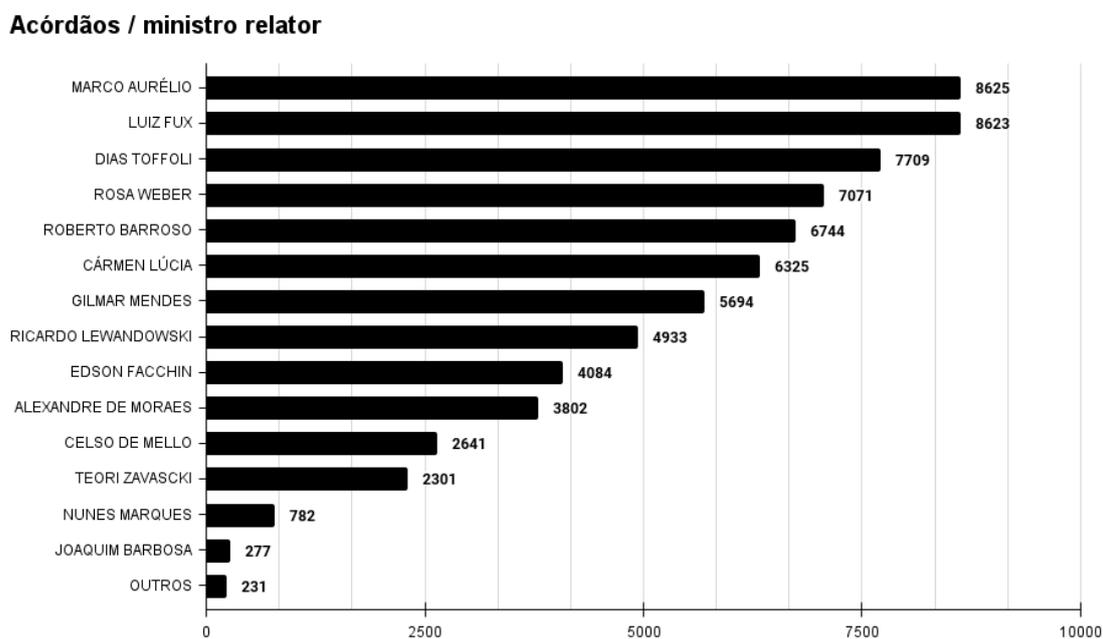
(a) 2012 a 2022



(b) 1934 a 2022

A Figura 6.6 apresenta o total de acórdãos cujos Ministros foram relatores do processo, considerando o período de janeiro de 2012 à agosto de 2022. O Ministro Marco Aurélio foi o mais ativo, sendo relator de 8625 (12.31%) processos. O Ministro Joaquim Barbosa, por sua vez, foi o Ministro menos ativo em termos de relatoria no STF, com apenas 277 processos (0.4%). O restante dos 231 acórdãos (0.33%), divididos entre outros 11 Ministros relatores, foram agrupados na categoria "Outros".

Figura 6.6: Total de acórdãos julgados por Ministro relator no período de janeiro de 2012 à agosto de 2022.

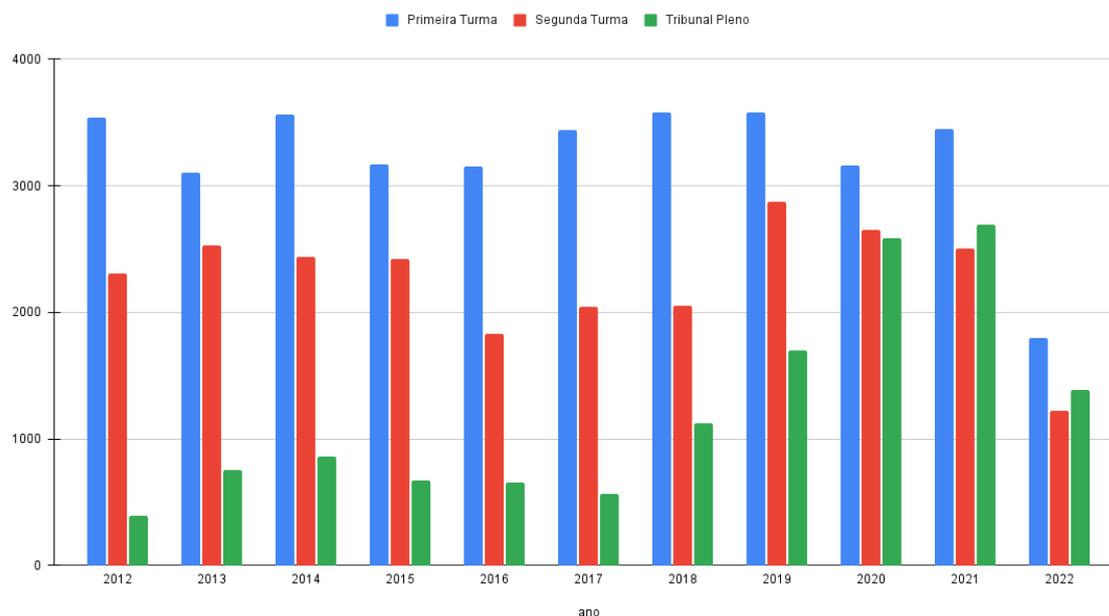


Fonte: o autor

A Figura 6.7 apresenta a distribuição dos acórdãos a partir do Órgão julgador no STF: Primeira Turma, Segunda Turma, ou Tribunal Pleno, considerando o período de janeiro de 2012 à agosto de 2022. Em todos os anos, a Primeira Turma foi o Órgão que julgou o maior número de acórdãos, tendo o maior índice registrado em 2018, com 3577 (10.07% de todos julgados por esse Órgão) processos. Ao todo, a Primeira Turma julgou 35524 processos. O Tribunal Pleno, entretanto, foi o Órgão julgador menos acionado, com apenas 13390 acórdãos registrados entre todo o período analisado. O ano mais ativo foi 2021 com 2696 (20.13% de todos julgados por esse Órgão) acórdãos registrados. É interessante observar que a partir do ano de 2017, a quantidade de processos julgados pelo Tribunal Pleno começou a crescer constantemente ao ponto de quase igualar o total de acórdãos com a Segunda Turma (2582 e 2654, respectivamente) e, pela primeira vez, ultrapassar em 2021 (2696 e 2500, respectivamente).

Figura 6.7: Órgão julgador dos acórdãos no STF entre o período de janeiro de 2012 à agosto de 2022.

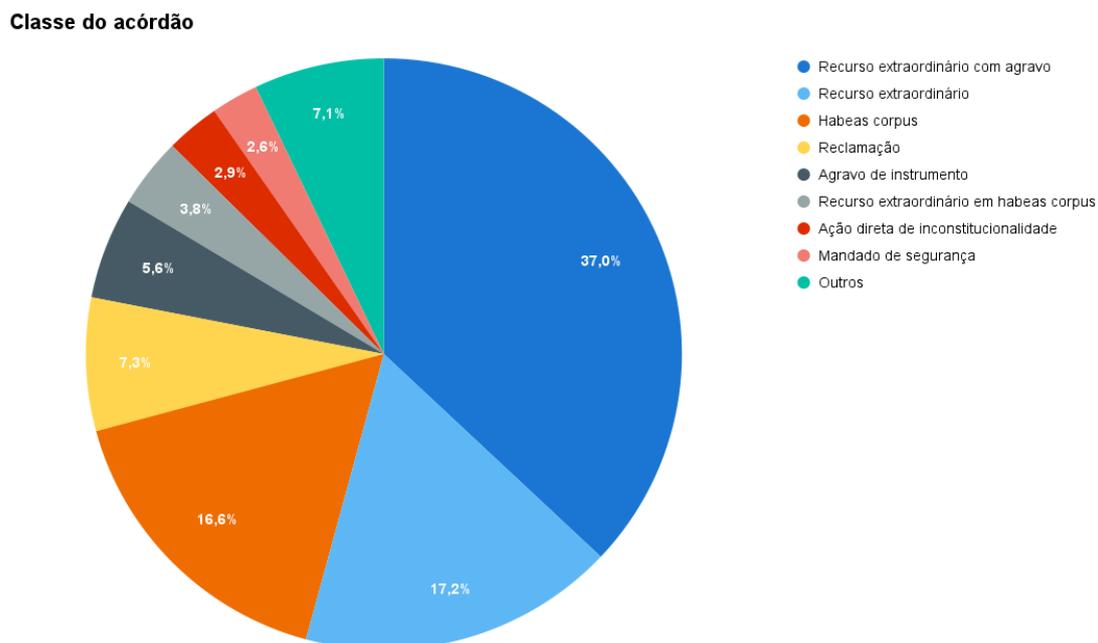
Órgão julgador / ano



Fonte: o autor

A Figura 6.8 apresenta distribuição dos acórdãos de acordo com sua classe processual, considerando o período de janeiro de 2012 à agosto de 2022. A maioria dos acórdãos, 27324 dos 73798 (37.03%), está na classe Recurso Extraordinário com Agravo. As classes Recurso Extraordinário e Habeas corpus, completam as três classes mais comuns entre acórdãos com 12683 (17.19%) e 12238 (16.58%), respectivamente. A categoria "Outros" agrupa 5218 acórdãos de 34 outras classes processuais.

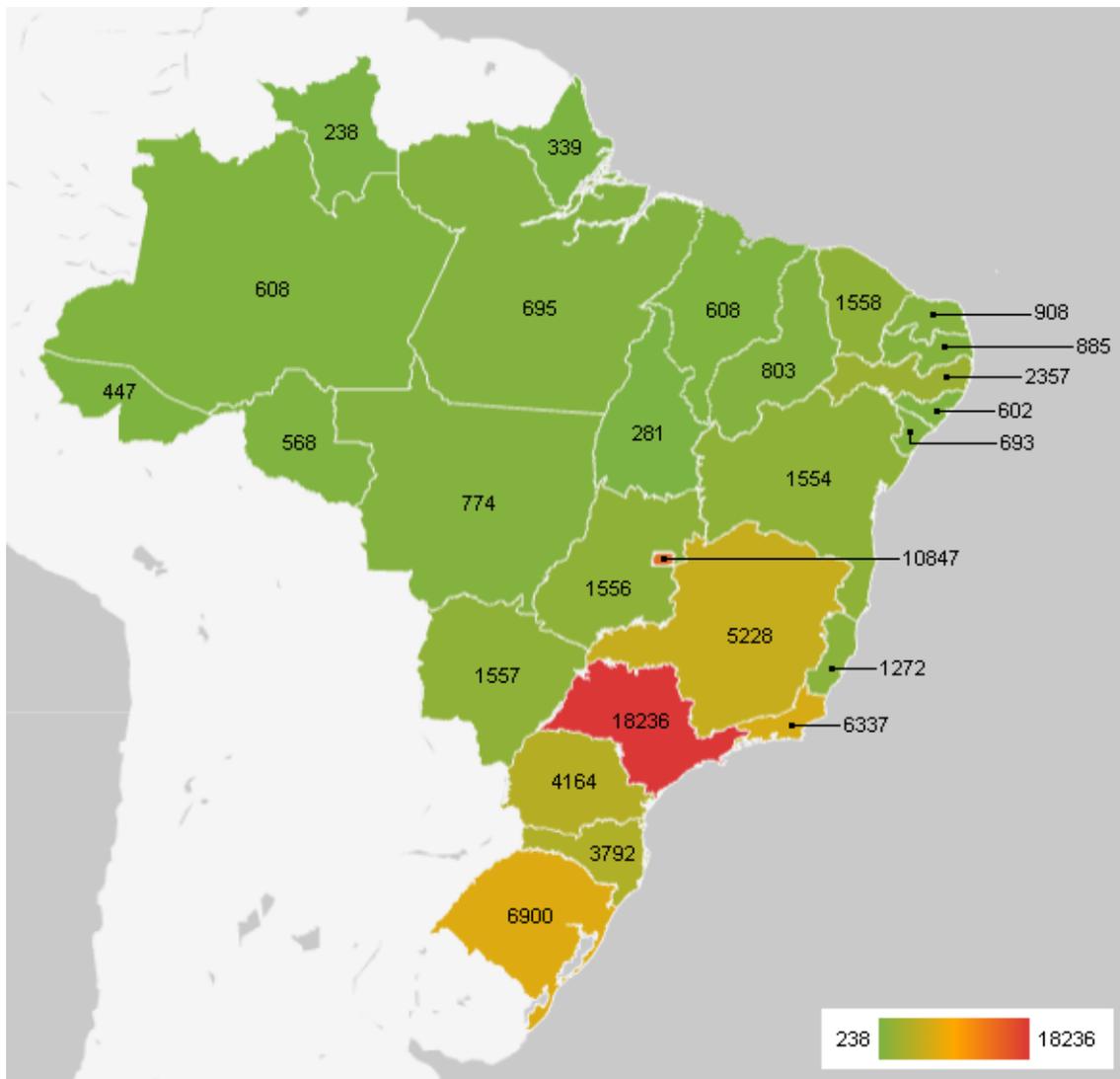
Figura 6.8: Distribuição de acórdãos por classe processual entre o período de janeiro de 2012 à agosto de 2022.



Fonte: o autor

A Figura 6.9 exibe um mapa de calor a partir da federação de origem dos acórdãos, considerando o período de janeiro de 2012 à agosto de 2022. A federação de origem mais comum dos acórdãos é do estado de São Paulo com 18236 (24.71%) processos, seguida pelo Distrito Federal com 10847 (14.7%), e Rio Grande do Sul, com 6900 (9.35%) acórdãos, respectivamente. A federação menos comum é Roraima com apenas 238 (0.32%) acórdãos.

Figura 6.9: Distribuição de acórdãos por estado brasileiro entre o período de janeiro de 2012 à agosto de 2022



Fonte: o autor

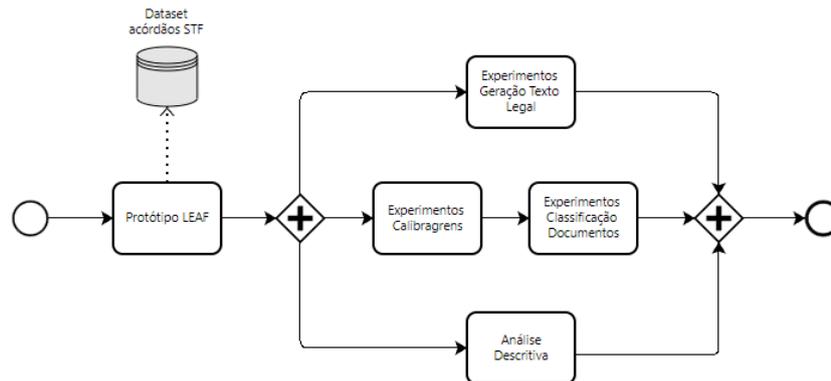
6.3 Experimentos na classificação de documentos

A partir da construção do *dataset* de acórdãos do STF, além da Análise Descritiva (Seção 6.2) e da utilização na tarefa de Geração de Texto Legal, através do modelo LeNER-Br e exibição de saída em uma página HTML (Seção 6.1.2), nesta seção são apresentados os resultados das simulações realizadas com os modelos NB, SVM, e RF para classificação de documentos legais.

O objetivo destas simulações foi descobrir o modelo com melhor desempenho, em métricas de F1-score e tempo de execução, entre as diferentes configurações aplicadas. Para a métrica F1-score, os valores de referência utilizados como inspiração, visto que o corpus é diferente, foram os trabalhos de Braz, Silva, Campos (SILVA; BRAZ; CAMPOS, 2018), e Braz et al (BRAZ

et al., 2018), com valores de 0.91 e 0.84, respectivamente. Para a métrica de tempo de execução, não encontramos *baseline* na literatura para experimentos de contexto semelhante. Para todas as simulações realizadas, a divisão do corpus em *dataset* de treinamento e *dataset* de validação foi de 70% e 30%, respectivamente. A Figura 6.11 apresenta uma visão geral da metodologia utilizada para a realização dos experimentos.

Figura 6.10: Metodologia utilizada nos experimentos.



Fonte: o autor

Esta seção está organizada em discussões sobre as calibrações e configurações escolhidas para os experimentos, análises sobre F1-score, análise sobre tempo de processamento, panorama geral dos resultados, e discussão geral dos resultados observados, respectivamente.

6.3.1 Calibragem inicial e configuração escolhida

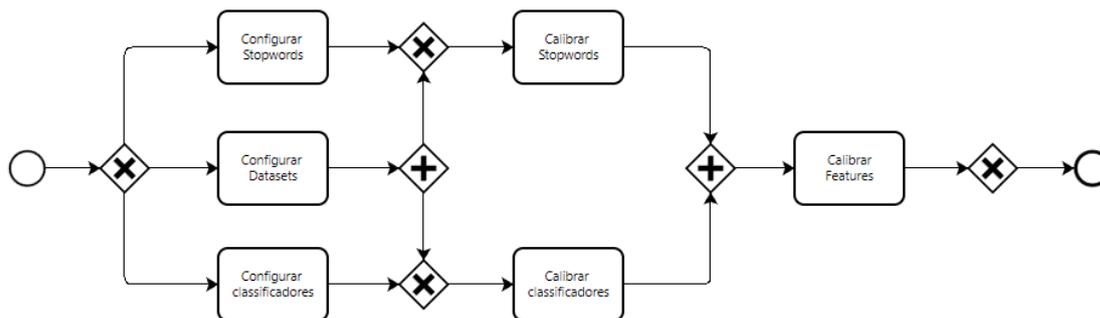
Os experimentos para a calibragem e configurações das simulações para os modelos de classificação possuem dois objetivos principais: (i) analisar empiricamente o comportamento dos modelos de classificação em relação às simulações submetidas, e (ii) encontrar combinações com desempenhos interessantes para serem utilizadas nos experimentos posteriores com os modelos de classificação. Portanto, antes de avaliar os desempenhos atingidos durante os experimentos com os modelos de classificação, é fundamental conhecer a base de dados utilizada, além da calibragem realizada e configuração aplicada.

A base de dados construída a partir do *scrap* das páginas do STF e da conversão de textos do arquivo PDF de acórdãos resultou em uma coleção de 73798 acórdãos, armazenados em aproximadamente 4.9 Gb de dados: 1.8 Gb de informações das páginas Web, e 3.11 Gb de textos do PDF. Em virtude do tamanho da coleção, duas configurações de teste se mostraram interessantes para o contexto deste trabalho: diferentes *datasets* da coleção e diferentes listas de *stopwords*.

Adicionalmente, foram realizados testes com as combinações anteriores de modo a encontrar os melhores parâmetros para os modelos de classificação. Por fim, com as definições de

configurações dos *datasets*, *stopwords*, e parâmetros dos modelos, os *benchmarks* puderam ser realizados. A Figura 6.11 apresenta um resumo das etapas de calibragem e escolha das configurações. O restante desta subseção discute e justifica as etapas realizadas.

Figura 6.11: Visão geral das etapas realizadas para configuração e calibragem de *stopwords*, *datasets*, parâmetros dos classificadores, e número de *features*, respectivamente.



Fonte: o autor

6.3.1.1 Datasets

A divisão da coleção em diferentes *datasets* tem como objetivo analisar o desempenho dos modelos de diferentes volumes de dados processados. Nesta configuração, a suposição inicial é de que a coleção maior tende a apresentar maior precisão e revocação ao custo de tempo de execução. Contudo, mais do que confirmar ou refutar a hipótese, é de interesse obter métricas sobre o *trade-off* desta relação, uma vez que o tempo de execução pode escalar muito rapidamente e o retorno, em termos de precisão e revocação, não justificarem uma coleção maior. Analogamente, se a precisão e revocação atingem valores atrativos, um custo de processamento e execução do modelo de classificação pode ser tolerável.

Desta forma, 3 *datasets*, *A*, *B*, e *C*, foram definidos de acordo com o contexto das informações. O primeiro *Dataset*, *A*, é composto apenas pelos textos das ementas dos acórdãos. O campo "Ementa" foi escolhido por apresentar um sumário do acórdão. Logo, a análise de uma informação que descreve resumidamente um volume maior de texto é bastante convidativa, uma vez que pode reduzir drasticamente tempo de processamento e gerência de recursos ao treinar, validar e/ou testar um modelo.

O *Dataset B* expande a quantidade de campos avaliados: além do campo Ementa, são considerados todos os campos de texto apresentados nas páginas do STF. A ideia da utilização deste *dataset* é a suposição de que *A* pode não apresentar um desempenho satisfatório por ter pouco volume de texto. Assim, uma coleção mais abrangente de informações, mas com um tamanho intermediário de dados, pode justificar sua escolha mesmo com precisão e revocação menores do

que os obtidos com a coleção completa se o tempo de execução e processamento forem atrativos.

Por fim, o *Dataset C* corresponde à coleção completa, isto é, possui todo o texto do acórdão extraído e convertido do arquivo PDF, na íntegra. Portanto, a seguinte relação é válida para os 3 *datasets*: $A \subset B \subset C$. A Tabela 6.3 apresenta um resumo dos *datasets* construídos considerando número médio de palavras por registro (ou documento) e tamanho, em *Megabytes*, da coleção, representados através ds colunas Palavras e Tamanho, respectivamente.

Tabela 6.3: Visão geral dos *datasets* utilizados nos experimentos para classificação de documentos.

Dataset	Descrição	Palavras	Tamanho
A	Textos apenas do campo Ementa	147.12	82.83
B	Textos apresentados nas páginas do STF	230.32	129.68
C	Textos extraídos de PDFs dos acórdãos	7794.73	4388.69

Fonte: o autor

6.3.1.2 Stopwords

A segunda configuração remete à lista de *stopwords* utilizadas ao aplicar técnicas de PLN e IA em textos dos *datasets*. Durante a calibragem desta configuração, a seguinte suposição foi levantada: se a lista de *stopwords* for mais restritiva, isto é, incluir mais palavras comuns, o desempenho da classificação dos modelos pode melhorar. Para validar a suposição, foram criadas 3 listas de *stopwords*, denominadas *L1*, *L2*, e *L3*.

A lista *L1* contava somente com as *stopwords* nativas do idioma português da biblioteca NLTK¹², que é constituída, em sua maioria, por artigos, preposições, e conjunções. Ao todo, 207 palavras são incluídas nesta lista.

A lista *L2* expande *L1* adicionando 14 novas palavras. O critério para a adição destas palavras foi uma análise dos 10 termos jurídicos de maior ocorrência em todo o corpus. Posteriormente, abreviações destes termos também foram incluídas. Assim, as palavras "1ªt", "2ªt", "acórdão", "agr", "agravo", "art", "artigo", "documento", "ministério", "mp", "público", "rec", "recurso", "supremo", "tribunal", e "turma" foram adicionadas à *L2*.

Por último, a lista *L3* expande *L1* e *L2* com a adição de todas as abreviações de classes de acórdãos encontradas no corpus. Por exemplo, "hc", que corresponde à abreviação da classe de acórdão "habeas corpus" foi adicionada à lista. Ao todo, 41 novas palavras foram incluídas em *L3*.

De modo similar à configuração de *datasets*, a seguinte relação também é válida para as 3 listas: $L1 \subset L2 \subset L3$. A Tabela 6.4 compara as características das listas de *stopwords* criadas.

¹²<<https://www.nltk.org/index.html>>

A coluna NLTK corresponde à lista de *stopwords* portuguesas presentes na biblioteca NLTK. A coluna TJ diz respeito à lista de palavras com os termos jurídicos mais frequentes. Por fim, a coluna CA é composta pela lista de abreviações das classes dos acórdãos.

Tabela 6.4: Visão geral das listas de *stopwords* criadas para calibragem das configurações iniciais dos experimentos.

Lista <i>stopwords</i>	NLTK	TJ	CA	Palavras
L1	Sim	Não	Não	207
L2	Sim	Sim	Não	221
L3	Sim	Sim	Sim	262

Fonte: o autor

6.3.1.3 Calibragem de *stopwords*

Com a definição de 3 *datasets* e 3 listas de *stopwords*, foram realizados diversos testes com os algoritmos NB, SVM, e RF, obedecendo à permutação destas variáveis. Isso significa que 27 configurações de teste foram criadas. Cada configuração foi executada 5 vezes e o desempenho médio foi considerado na métrica para avaliação.

Os algoritmos não tiveram nenhuma configuração alterada; portanto, os valores padrão para os métodos utilizados da biblioteca `scikit-learn` foram preservados na execução desta calibragem. Após a execução dos testes, foi observada uma redução na precisão e revocação dos modelos a medida que a lista de palavras aumentava, para todos os *datasets*. O tempo de execução também apresentou o mesmo comportamento na maioria dos testes. A Tabela 6.5 apresenta um resumo dos resultados da calibragem, onde cada célula da corresponde à média do F1-score dos 3 *datasets* para cada lista de *stopwords*.

Tabela 6.5: Resultados obtidos durante a calibragem de *stopwords*.

Lista <i>stopwords</i>	NB	SVM	RF
L1	0.767	0.9354	0.9030
L2	0.7365	0.8711	0.8621
L3	0.7321	0.8257	0.8173

Fonte: o autor

Como é possível observar, todos os algoritmos tiveram um desempenho melhor para a lista L1, que contava com o menor número de *stopwords*. Desta forma, a suposição de que a adição de um maior número de palavras -sendo estas termos jurídicos e abreviações de classes de

acórdãos- poderia contribuir para uma maior precisão e revocação dos modelos na classificação de documentos foi refutada. Portanto, para todos os testes executados após esta etapa, é considerado o uso apenas da lista *L1* de *stopwords*.

6.3.1.4 Calibragem dos algoritmos

A terceira etapa foi a calibragem de parâmetros dos métodos utilizados por cada modelo (NB, SVM, RF) da biblioteca `scikit-learn`. Diferente da execução durante a calibragem com *stopwords*, o objetivo é encontrar a melhor combinação de parâmetros dos métodos a fim de obter o melhor F1-score com os 3 *datasets* criados nas configurações anteriores.

A melhor combinação para o classificador NB foi utilizar a versão Multinomial, com os parâmetros originais do método da biblioteca. Alterações em parâmetros de *smoothing* e aprendido a priori resultaram em F1-score menor.

O classificador SVM utilizou uma função de perda linear. Para ajustar o classificador, foi também utilizado um estimador *Stochastic Gradient Descent* (Gradiente Descendente Estocástico) de forma a suavizar as linhas do hiperplano. Outro parâmetro ajustado foi o número máximo de iterações, isto é, a passagem completa do conjunto de dados de treinamento pelo algoritmo, para 5. Valores maiores não resultaram em ganho considerável de F1-score e o tempo de execução demonstrou tendência à inviabilidade para grandes volumes de texto.

O classificador RF exigiu o maior número de ajustes de parâmetros devido à natureza do algoritmo. Entre as diversas combinações realizadas, os melhores resultados de F1-score foram atingidos com uma floresta de 50 árvores, sem restrições para a profundidade da árvore, o número máximo de subamostras, e número de folhas das árvores, com 4 processos paralelos no algoritmo. Diferentemente dos outros dois modelos, o classificador RF enfrentou problemas de performance em determinadas configurações de teste, além do elevado consumo de recursos -como memória-, impossibilitando a finalização de testes. Os parâmetros escolhidos ofereceram o melhor equilíbrio entre gerência de memória, tempo de execução, e F1-score.

6.3.1.5 Número de features

A última etapa antes da realização do *benchmark* entre os modelos de classificação, foi a definição das configurações possíveis para o número de *features* utilizadas durante o treinamento dos modelos. Após algumas simulações de modo empírico, foram definidos os seguintes números de *features* para avaliação: 100, 500, 1000, 5000, 10000. Valores abaixo de 100 não tiveram desempenho satisfatório e elegível em métricas de precisão e revocação. Analogamente, um número maior do que 10000 demonstrou ser um problema para modelos como RF na gerência de memória e um tempo de execução inviável para os modelos NB e SVM.

Para os três modelos de classificação, foi utilizada a combinação do algoritmo TF-IDF aliado a um modelo BOW para gerar um vetor de palavras dos textos ordenados de acordo com sua relevância no corpus. Portanto, neste contexto, para os três modelos de classificação, cada *feature* corresponde a uma palavra do vetor de palavras. Assim, quando a configuração considera 100 *features*, ele considera as 100 palavras mais relevantes do corpus. Analogamente, 10000 *features* correspondem às 10000 palavras mais relevantes do corpus. Outra configuração possível seria não limitar um número máximo de *features* e considerar o todo o corpus; porém, além do tempo de execução aumentar consideravelmente, a precisão e revocação tiveram uma diminuição de desempenho.

6.3.1.6 Considerações finais sobre as simulações

Ao todo, foram realizadas 225 simulações para a avaliação do *benchmark*. Para cada modelo de classificação, para cada *dataset*, para cada número de *features*, foi obtida a média de 5 execuções em relação ao F1-score e ao tempo de execução. A expressão a seguir apresenta a álgebra aplicada para obter a média do F1-score (F1) e tempo de execução (t), para uma determinada configuração de simulação. O conjunto S corresponde aos modelos de classificação NB, SVM, RF; D é o conjunto dos 3 *datasets*, e F o conjunto dos 5 configurações de *features* definidas.

$$\frac{\sum_{n=1}^5 (s_i \times d_j \times f_k)}{5} = \langle F1_{ijk}, t_{ijk} \rangle \mid s \in S, d \in D, f \in F$$

A execução dos experimentos foi realizada em um único ambiente com as seguintes configurações:

- Intel i7-7700, 3.60 GHz;
- 16 GB RAM DDR3;
- 4 GB VRAM.

6.3.2 Análise F1-score

Os experimentos para análise do F1-score objetivam medir o desempenho dos modelos classificadores na tarefa de classificação de documentos ao realizar a predição da classe dos acórdãos. Para avaliar o desempenho dos classificadores, foi utilizada a métrica F1-score, que consiste em uma média harmônica das métricas de precisão e revocação. Os valores de referência e inspiração utilizados para comparar os resultados obtidos nos experimentos realizados foram os trabalhos "Document type classification for brazil's supreme court using a convolutional neural network"(SILVA; BRAZ; CAMPOS, 2018) e "Document classification using a bi-lstm to un-

clog brazil's supreme court"(BRAZ et al., 2018). O trabalho de Braz, Silva, Campos (SILVA; BRAZ; CAMPOS, 2018) obteve um F1-score de 0.91 empregando CNN. Posteriormente, Braz et al (BRAZ et al., 2018). obteve F1-score de 0.84, mas desta vez, com uma rede neural Bi-LSTM.

As suposições iniciais do experimento eram de que: (i) o modelo classificador NB teria o pior desempenho, devido à natureza do algoritmo em utilizações de tarefas para classificações de texto; (ii) os modelos classificadores SVM e RF poderiam ter um desempenho superior aos valores de referência de trabalhos relacionados, uma vez que são algoritmos comumente utilizados em soluções para tarefas com grande volume de texto.

A Tabela 6.6 apresenta os resultados gerais do desempenho para cada modelo classificador. Analisando cada bloco da coluna "Solução", referente aos modelos, é possível observar comportamentos distintos em função das configurações de *datasets* e número de *features*.

Tabela 6.6: Tabela comparativa do F1-score entre os modelos NB, SVM, RF em diferentes configurações de número de *features* e *datasets* construídos.

Solução	Nº Features	Dataset A	Dataset B	Dataset C
NB	100	0.6772	0.6345	0.2037
NB	500	0.6958	0.6626	0.6998
NB	1000	0.6818	0.6802	0.7554
NB	5000	0.6897	0.6749	0.7603
NB	10000	0.6879	0.6718	0.7463
SVM	100	0.7660	0.773	0.6936
SVM	500	0.868	0.9019	0.9316
SVM	1000	0.887	0.9312	0.9321
SVM	5000	0.9075	0.9497	0.951
SVM	10000	0.9146	0.9538	0.96
RF	100	0.8464	0.8856	0.7602
RF	500	0.8913	0.931	0.9309
RF	1000	0.8983	0.9475	0.9469
RF	5000	0.906	0.9415	0.9565
RF	10000	0.9064	0.9403	0.9586

Fonte: o autor

O modelo NB apresentou os piores resultados para o F1-score. Contudo, um comportamento interessante é a evolução do desempenho em função do número de *features*: os melhores resultados foram quando o número de *features* convergiu para um número intermediário de acordo com o tamanho do *dataset*. Valores abaixo ou acima do intervalo apresentaram desempenho inferior. Este comportamento demonstra que um número maior de *features* é indicado quando a base de dados e o corpus possuem grande volume de dados. Por outro lado, se o corpus é pequeno, um número grande de *features* acarreta em diminuição de desempenho. A partir dos resultados apresentados, constata-se:

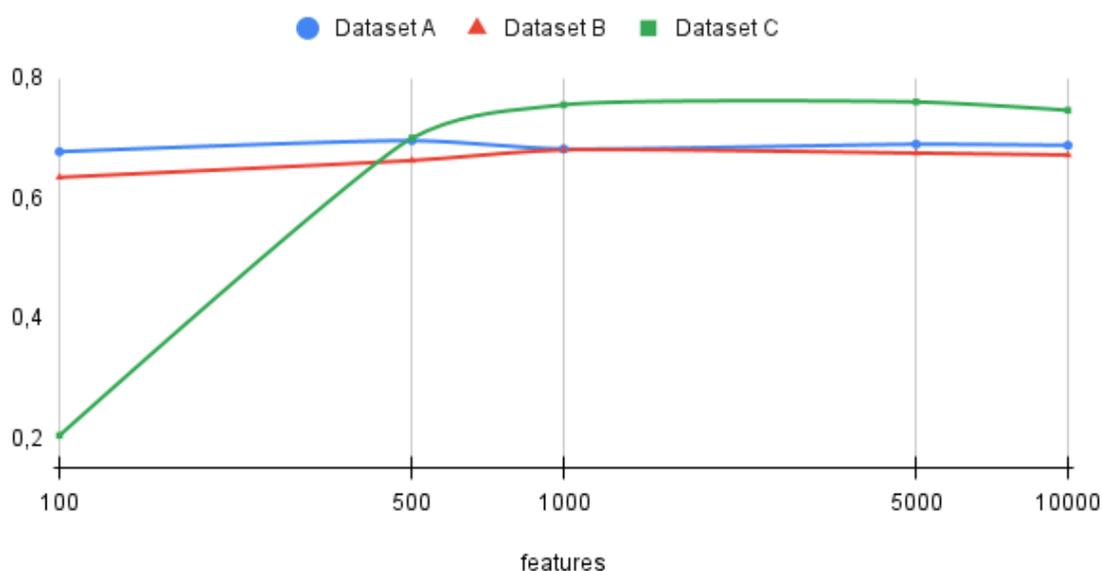
- o melhor F1-score foi obtido com o *Dataset C* com 5000 *features* (0.7603);

- o pior F1-score também foi obtido com o *Dataset C*, mas com apenas 100 *features* (0.2037);
- o menor corpus, *Dataset A*, teve melhor F1-score com 500 *features* (0.6958);
- o corpus intermediário, *Dataset B*, teve melhor F1-score com 1000 *features* (0.6802).

A Figura 6.12 apresenta a média do F1-score do modelo classificador NB, para cada *dataset*, de acordo com o número de *features* utilizados em cada simulação.

Figura 6.12: F1-score dos *datasets* de acordo com o número de *features* no modelo classificador NB.

F1-score Naive Bayes



Fonte: o autor

O modelo SVM apresentou um comportamento mais linear. De modo geral, quanto maior o corpus para o treinamento, melhor o desempenho. Similarmente, quanto maior o número de *features*, melhor o desempenho. Entretanto, quanto maior o corpus e menor o número de *features*, maior o impacto e queda de desempenho do modelo, como podemos observar na diferença entre o F1-score quando a simulação muda de 100 para 500 *features*. Um ponto importante a ser observado a partir dos resultados consolidados é que um número elevado de *features* fornece uma diferença pequena de F1-score entre diferentes tamanhos de corpus. Em especial, para os *Dataset B* (intermediário) e *C* (completo), a diferença foi de apenas 0.0062. Considerando o tempo necessário para o treinamento dos modelos, um volume de dados muito grande pode ser substituído por um mais tangível à gerência de recursos como memória e processamento. Pode-se, então, constatar a partir dos resultados que:

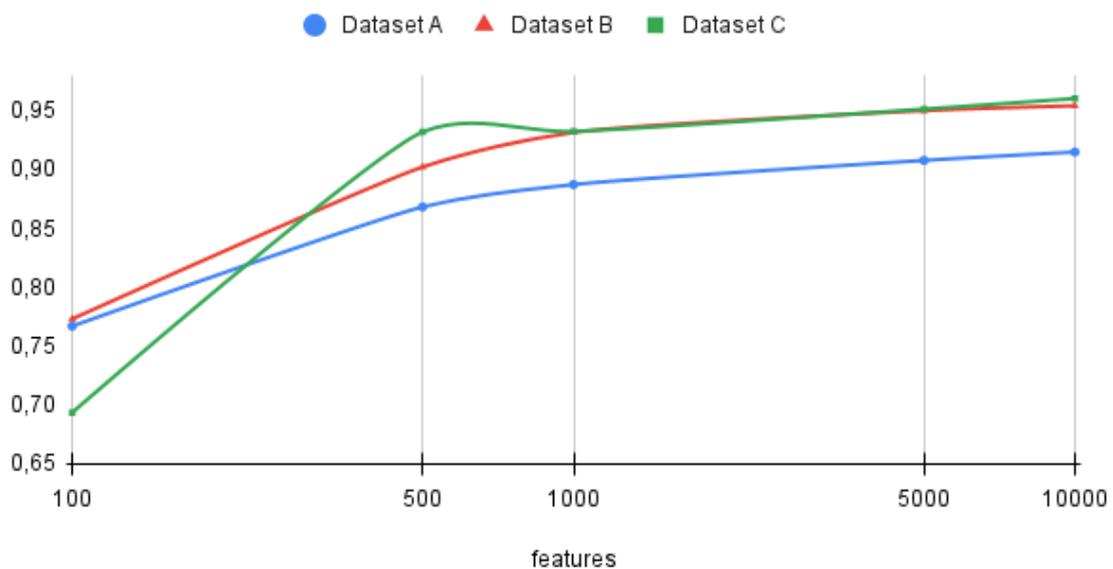
- o melhor F1-score foi obtido com o *Dataset C* com 10000 *features* (0.96);

- o pior F1-score também foi obtido com o *Dataset C*, mas com apenas 100 *features* (0.6936);
- o menor corpus, *Dataset A*, teve melhor F1-score com 10000 *features* (0.9146);
- o corpus intermediário, *Dataset B*, teve melhor F1-score com 10000 *features* (0.9538).

A Figura 6.13 apresenta a média do F1-score do modelo classificador RF, para cada *dataset*, de acordo com o número de *features* utilizados em cada simulação.

Figura 6.13: F1-score dos *datasets* de acordo com o número de *features* no modelo classificador SVM.

F1-score SVM



Fonte: o autor

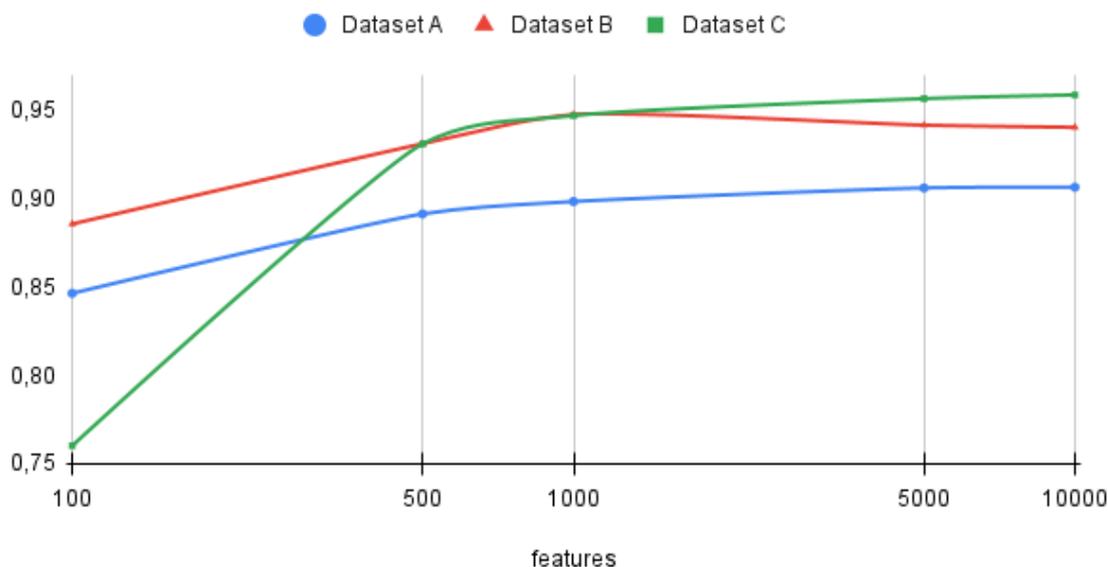
Por fim, o modelo classificador RF apresentou um comportamento híbrido com as configurações realizadas nos experimentos. Quando o modelo utilizou o *Dataset B*, de tamanho intermediário, os melhores resultados de F1-score ocorreram quando convergia para 1000 *features*. Ainda assim, valores acima deste limiar conseguiram um desempenho melhor do que abaixo. Já, para os *Datasets A* e *C*, menor e maior, respectivamente, o comportamento foi linear: quanto maior o número de *features*, maior o F1-score. Este comportamento está associado à natureza do algoritmo RF, uma vez que diversas árvores de decisão são combinadas, e o número de *features* (neste caso, palavras relevantes do vetor TF-IDF) em função do corpus associado pode acarretar uma variância maior de decisões de cada árvore, prejudicando a predição do modelo baseado na decisão majoritária da floresta. Logo, é difícil afirmar uma regra absoluta que funcione bem para todas as configurações do modelo RF, sendo necessário testar e analisar empiricamente os resultados a partir de diferentes simulações. Todavia, para os resultados alcançados, pode-se constatar que:

- o melhor F1-score foi obtido com o *Dataset C* com 10000 *features* (0.9586);
- o pior F1-score também foi obtido com o *Dataset C*, mas com apenas 100 *features* (0.7602);
- o menor corpus, *Dataset A*, teve melhor F1-score com 10000 *features* (0.9064);
- o corpus intermediário, *Dataset B*, teve melhor F1-score com 1000 *features* (0.9475).

A Figura 6.14 apresenta a média do F1-score do modelo classificador SVM, para cada *dataset*, de acordo com o número de *features* utilizados em cada simulação.

Figura 6.14: F1-score dos *datasets* de acordo com o número de *features* no modelo classificador RF.

F1-score Random Forest



Fonte: o autor

Dada as suposições iniciais sobre o desempenho dos modelos classificadores, foi possível constatar que, de fato, o modelo classificador NB apresentou o pior desempenho comparado aos modelos SVM e RF, além dos valores de referência de trabalhos relacionados utilizados (i). Ainda, os modelos classificadores SVM e RF apresentaram desempenho superior aos valores de referência utilizados (ii).

6.3.3 Análise e resultados sobre tempo de execução

Para avaliar o desempenho dos classificadores em função do tempo de execução, consideramos o tempo de treinamento dos modelos exclusivamente. Isto é, o tempo para carregar os *datasets*, *stopwords* e gerar o vetor TF-IDF de *features* não foi contabilizado. Diferentemente da

seção anterior, para a avaliação desta métrica não foi encontrado um *baseline* nos trabalhos relacionados referente à dimensão temporal ou simplesmente valores de referência para utilizar como inspiração. Todavia, os dados são apresentados nesta Seção para complementar as considerações deste trabalho.

A Tabela 6.7 apresenta os resultados gerais do tempo de execução, em segundos, para cada *dataset* em função do modelo classificador e do número de *features*. Como é possível observar pelos valores exibidos, o tempo de execução das simulações aumenta proporcionalmente ao número de *features* e tamanho do *dataset*, uma vez que mais dados são necessários para processar as informações e realizar as predições de cada modelo.

Tabela 6.7: Tabela comparativa do tempo de execução nos algoritmos NB, SVM, e RF para diferentes número de features no *dataset 01* construído.

Dataset	Nº Features	NB	SVM	RFC
A	100	18.2	22.4	33.35
A	500	19.1	22.18	42.12
A	1000	19.07	22.9	44.64
A	5000	20.6	24.35	56.42
A	10000	20.35	28.69	66.19
B	100	33.51	37.06	54.17
B	500	34.07	39.66	58.41
B	1000	34.9	40.96	59.98
B	5000	37.08	41.63	67.3
B	10000	36.78	43.38	69.7
C	100	1147.86	1067.43	1090.26
C	500	1127.49	1019.72	1110.44
C	1000	1171.95	1063.83	1072.92
C	5000	1196.32	1062.33	1089.72
C	10000	1190.32	1074.21	1067.4

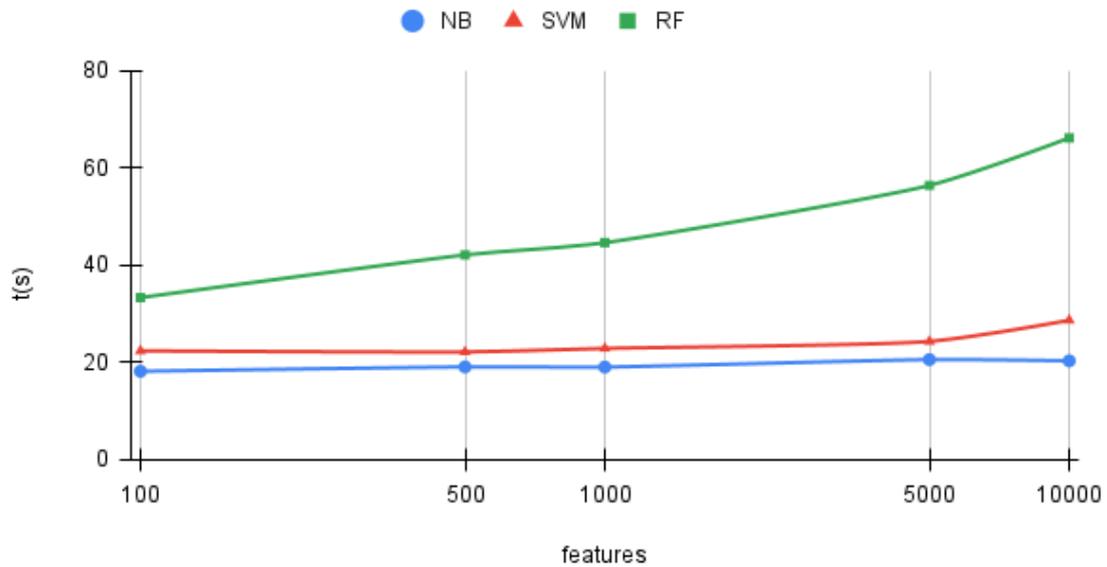
Fonte: o autor

A análise mais interessante da Tabela é o tamanho da variação de tempo entre os corpus para cada modelo, e de um modelo para o outro com as mesmas configurações fornecidas. Entre os modelos, o classificador NB é o mais rápido para os Datasets A e B, mas não obteve a mesma performance para o *Dataset C* (maior) e apresentou os maiores tempos entre todas as simulações. O classificador SVM apresentou desempenho intermediário para os Datasets A e B, e o melhor desempenho para o *Dataset C*. Por fim, o classificador RF apresentou os piores tempos nos Datasets A e B, e desempenho intermediário no *Dataset C*.

As Figuras 6.15, 6.16, 6.17 apresentam graficamente os resultados obtidos na Tabela 6.7 para os Datasets A, B, e C, respectivamente. Através dos gráficos dos resultados consolidados, é reforçado o comportamento linear dos modelos para os menores *datasets*, e um comportamento mais instável dos modelos para o maior *dataset*.

Figura 6.15: Tempo de execução do *Dataset A* de acordo com o número de *features* e modelos classificadores.

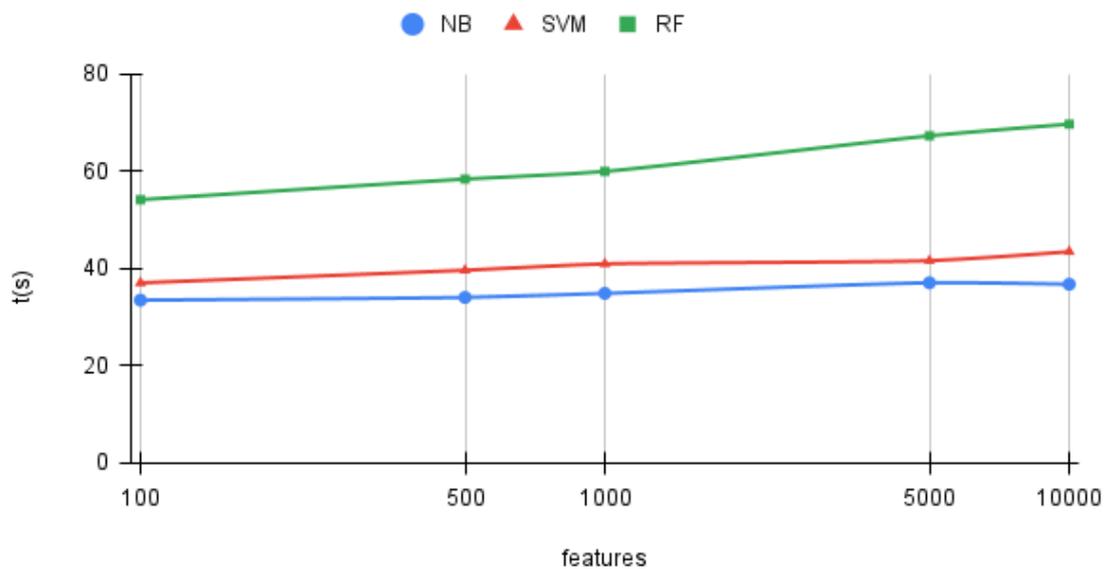
Dataset A



Fonte: o autor

Figura 6.16: Tempo de execução do *Dataset B* de acordo com o número de *features* e modelos classificadores.

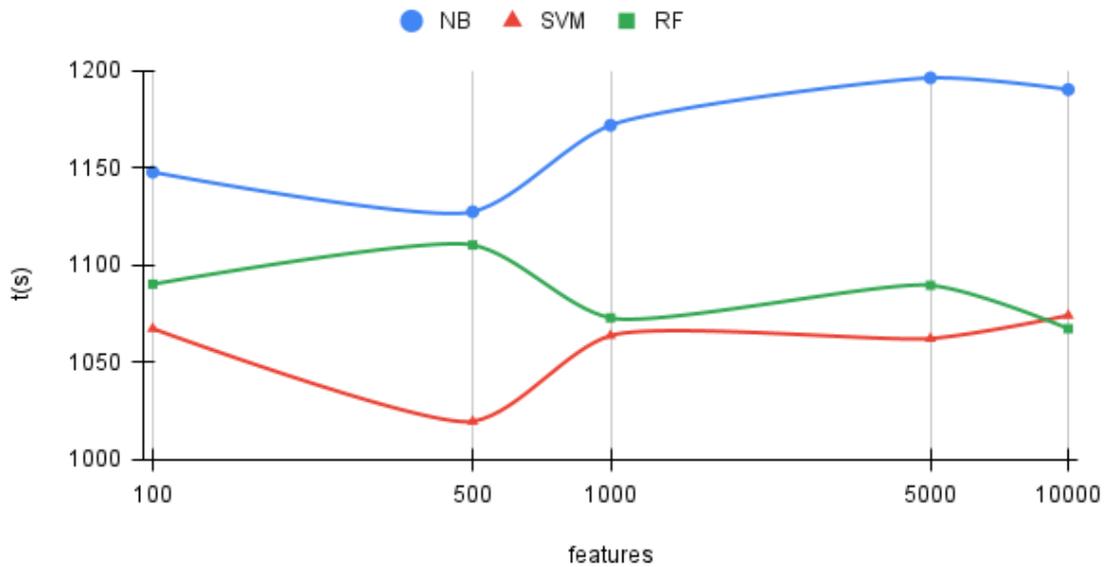
Dataset B



Fonte: o autor

Figura 6.17: Tempo de execução do *Dataset C* de acordo com o número de *features* e modelos classificadores.

Dataset C



Fonte: o autor

A partir dos resultados apresentados, constata-se:

- o menor tempo de execução geral foi obtido com o modelo NB, para o *Dataset A*, com 100 *features* (18.2s);
- o maior tempo de execução geral foi obtido com o modelo NB, para o *Dataset C*, como 5000 *features* (1196.32s \approx 20 minutos). Esta configuração também foi a que obteve melhor F1-score;
- o maior F1-score do modelo SVM foi obtida com o *Dataset C*, para 10000 *features*, com o tempo de execução em 1074.21s (\approx 18 minutos);
- o maior F1-score do modelo RF foi obtida com o *Dataset C*, para 10000 *features*, com o tempo de execução em 1067.4s (\approx 18 minutos).

6.3.4 Panorama geral dos resultados

Os resultados obtidos com todas as diferentes configurações nas simulações realizadas, indicaram duas soluções promissoras para a classificação de documentos legais: classificador SVM e o classificador RF. O classificador SVM apresentou melhores tempo de execução, precisão e revocação, atingindo um F1-score de 0.96 na melhor configuração. Contudo, o classificador RF

obteve valores bem próximos de precisão e revocação (0.9586) e com um tempo de execução levemente superior.

Acredita-se que ambos os classificadores consigam melhorar um pouco mais suas predições com técnicas para otimização de hiperparâmetros. Em especial, o classificador RF possui uma gama maior de variáveis que podem resultar em predições mais assertivas, uma vez que quantidade de árvores de decisão, profundidade das árvores de decisão, quantidade de folhas, *bootstrapping* da árvore, entre outros são ajustáveis. No entanto, ao passo que a predição tem indícios de que pode melhorar, a natureza do algoritmo exige um ambiente mais robusto para as simulações, necessitando mais memória e poder de processamento.

As soluções com o classificador NB obtiveram resultados inferiores aos outros dois modelos, uma vez que o melhor F1-score foi 0.7603. O comportamento do classificador sugere que o desempenho dificilmente melhorará, mesmo aumentando o tamanho do corpus e/ou número de *features*. Portanto, o classificador probabilístico demonstrou ser ineficiente para a classificação de documentos legais com um volume grande de texto.

Por fim, analisando os resultados consolidados, tem-se uma nova ordenação da métrica F1-score considerando os valores de referência iniciais desta Seção, onde foram empregadas técnicas com redes neurais do tipo CNN e Bi-LSTM. Portanto, para a classificação de documentos legais do STF, temos que:

$$\begin{aligned} &= F1_{NB} < F1_{CNN} < F1_{BiLSTM} < F1_{RF} < F1_{SVM} & (6.1) \\ &= 0.7603 < 0.81 < 0.91 < 0.9586 < 0.96 \end{aligned}$$

6.3.5 Discussão geral dos resultados

Nesta seção foram apresentados os experimentos realizados para a tarefa de classificação de documentos legais que constituem a coleção de acórdãos do STF brasileiro. Inicialmente foram discutidas suposições, calibrações e configurações escolhidas para a realização das simulações. Neste contexto, foram analisadas duas métricas: a precisão e revocação, expressas através do F1-score, e o tempo de execução. Observou-se que o classificador SVM obteve o melhor F1-score geral médio com 96%, para 10000 *features* no corpus com o maior volume de texto. Entre as configurações considerando cada grupo de classificador, o modelo NB obteve F1-score médio de 76.03%, enquanto o modelo RF obteve 95.86%. Os tempos de execução associados às configurações com melhores F1-score dos classificadores, apontam o classificador SVM como o mais rápido, seguido por RF e NB, respectivamente. Finalmente, são feitas algumas considerações sobre margens para melhorias das predições dos classificadores, através de configurações de hiperparâmetros.

É importante observar, em meio aos resultados obtidos, que a interpretabilidade dos resultados é dificultada pela abstração dos modelos utilizados. Logo, a complexidade para entender a relevância de determinada *feature* ou conjunto de *features* é um desafio para a melhora do desempenho dos modelos. Este desafio pode ser contornado através da utilização de técnicas de *feature tuning* que são comuns em modelos classificadores de RF, ou, ainda, a implementação de soluções minimalistas, sem a utilização de modelos populares, pode ajudar a entender a relevância das *features* utilizadas. Ainda, uma outra perspectiva de análise interessante é, em adição às *features* multidimensionais, isto é, textos, considerar também *features* unidimensionais como datas, diferenças entre datas, tamanho de determinados textos, presença ou não de redator, Órgão julgador do acórdão, se houve decisão unânime, entre outras *features* candidatas.

7 CONCLUSÃO

Esta dissertação apresentou uma revisão sistemática sobre o impacto da IA na área jurídica, uma taxonomia para tarefas durante a geração de texto legal em *LegalTech* e, por fim, a contribuição principal: uma Abordagem Agnóstica para auxiliar em tarefas de *LegalTech*.

Inicialmente, foi realizada a revisão sistemática que considerou 146 estudos sobre geração de texto legal, durante o período entre janeiro de 1963 à dezembro de 2020, para identificar e apresentar um panorama sobre diferentes soluções propostas na geração de texto legal. O objetivo da revisão sistemática foi melhorar avanços nas áreas de *LegalTech* e, amplamente, tirar proveito de técnicas de Linguagem Natural.

A partir da revisão sistemática, uma taxonomia para classificar os estudos em cinco categorias de acordo com o escopo principal de cada trabalho quanto à geração de texto legal foi proposta. As classes da taxonomia para Tarefas de *LegalTech* ao gerar texto legal são: Argumentação, Extração, Geração, Ilação, e Recuperação.

Após a classificação das tarefas de *LegalTech* e da análise de lacunas encontradas nos trabalhos analisados, uma Abordagem Agnóstica foi proposta para auxiliar em tarefas de *LegalTech*. A Abordagem Agnóstica é composta por 4 etapas responsáveis pela coleta, processamento, modelagem e exibição de resultados a partir da utilização de técnicas de IA em documentos legais. O objetivo da Abordagem Agnóstica é providenciar recomendações para o desenvolvimento de soluções na área de *LegalTech* de forma coesa, estruturada, extensível, e robusta.

Para demonstrar a aplicabilidade da Abordagem Agnóstica em cada etapa proposta e a capacidade de abstração e generalização de tarefas, um protótipo foi construído para auxiliar em tarefas de geração de texto legal e classificação de documentos legais. Adicionalmente, através do *scrap* de informações dos acórdãos em páginas do STF brasileiro, um *dataset* foi construído e disponibilizado para utilização em trabalhos futuros.

A tarefa de geração de texto legal evidenciou a capacidade de extensibilidade da abordagem e, conseqüentemente, do protótipo. Esta tarefa foi realizada estendendo a solução proposta por Araujo et. al (ARAUJO et al., 2018), que construiu o modelo LeNER-Br através de redes neurais para classificar termos jurídicos. Com o reuso do LeNER-Br, a etapa de classificação dos termos foi removida, possibilitando direcionar os esforços para a interface de dados entre o *dataset* construído e utilizado no protótipo, e o aprimoramento da exibição, com palavras anotadas de acordo com a categoria assinalada pelo modelo.

A tarefa de classificação de documentos legais foi realizada utilizando os classificadores NB, SVM e RF. Para esta tarefa, 225 simulações com diferentes configurações para os classificadores foram realizadas para obter os melhores desempenhos em precisão e revocação para os modelos. A realização dos experimentos foi antecedida por etapas de calibragem de *stopwords*,

datasets, features, e parâmetros dos classificadores à coleção de informações extraídas dos acórdãos do STF.

Os valores de referência -e inspiração- utilizados nos resultados partiu dos trabalhos de Braz et al (BRAZ et al., 2018), e Silva, Braz, Campos (SILVA; BRAZ; CAMPOS, 2018) que obtiveram F1-score de 84% e 91%, respectivamente. A partir dos experimentos realizados, os melhores desempenhos de F1-score obtidos com os modelos classificadores implementados foram 96% para o SVM, 95.86% para o RF, e 76.03% para NB. Portanto, os classificadores SVM e RF apresentaram uma performance elevada para a predição de classes de acórdãos do STF, considerando um *dataset* de grande volume de texto legal.

Além dos resultados referentes às tarefas de *LegalTech*, este trabalho possibilitou a construção de um *dataset* de acórdãos do STF. O *dataset* é composto de duas fontes de informação: pelo texto extraído de arquivos PDFs que possuem o texto na íntegra do acórdão, e pelos dados extraídos de páginas HTML do STF. Essa combinação possibilitou a organização de diversas informações de cada acórdão em uma coleção de dados estruturados, facilitando o acesso a informações específicas, sem a necessidade de pré-processar todo o texto na íntegra para cada acórdão da coleção. Essa estratégia possibilitou uma análise descritiva dos dados, onde algumas estatísticas são apresentadas, como, por exemplo: quantidade de acórdãos por ano, por ministro (relator, redator), por Órgão julgador, por local de origem, e por classe de documento, se a decisão do acórdão foi unânime ou não, se houve adiamento da decisão etc. Para cada uma das análises, foi implementado um serviço *Web* que permite a aplicação de filtros e agrupamento desses dados para uma visualização da informação de acordo com o interesse do usuário.

Por fim, acredita-se que esta dissertação é valiosa para as comunidades acadêmica e industrial, especialmente para estudantes e profissionais que podem adotar a Abordagem Agnóstica para implementação de soluções e/ou aperfeiçoamento de recomendações em tarefas de *LegalTech*. Ter uma abordagem com capacidade de generalização permite incorporar um escopo de estudos e trabalhos maior para, posteriormente, instanciar uma solução para um contexto específico sem abdicar de critérios qualitativos.

7.1 Limitações

Como possível limitação deste trabalho, é considerada a concepção da Abordagem Agnóstica que surgiu a partir de uma *survey* sobre geração de texto legal. A partir desse estudo, foi proposta uma taxonomia para objetivos em tarefas de *LegalTech* que serviu de ponto de partida para a abordagem, que visa oferecer recomendações para soluções que supram as necessidades expostas pela taxonomia. Contudo, apesar da grande cobertura de tarefas de *LegalTech* pela abordagem, possivelmente podem existir outras tarefas da área que não foram analisadas pela *survey*

e, conseqüentemente, estariam fora do escopo da Abordagem Agnóstica. Como sugestão de melhoria, pode-se realizar uma pesquisa de estado-da-arte complementar de sub-áreas de *LegalTech* e estender o escopo de tarefas que a Abordagem Agnóstica contempla.

Outras limitações presentes nesta dissertação dizem respeito ao protótipo implementado. Para a etapa de extração de informações a partir de páginas *Web* no STF e textos de acórdãos, ao estruturar dados e pré-processar informações, é comum encontrar erro humano nos textos, isto é, informações em campos incorretos ou sentenças com erros de grafia e ambigüidade de palavras. Esse comportamento pode gerar transferência de informações errôneas em certas colunas do *dataset* e/ou uma extração sem sucesso no texto em linguagem natural. Como solução, seria interessante investir um tempo maior no aperfeiçoamento de técnicas de PLN para os serviços presentes em etapas de extração e pré-processamento do protótipo.

Por fim, outra limitação identificada neste trabalho remete ao ambiente de execução dos experimentos. Para soluções que demandaram maior poder de processamento e memória, não foi possível realizar as simulações. Seria interessante executar as simulações em outros ambientes com menor e maior poder de processamento para delimitar as reais capacidades dos modelos classificadores, especialmente no que tangem ao F1-score.

7.2 Direções futuras

Em continuidade a este trabalho, algumas direções futuras foram identificadas referentes a dois domínios: a Abordagem Agnóstica e o protótipo desenvolvido. A divisão nestes domínios correspondem a melhorias tanto para a teoria quanto prática, o que fomenta a colaboração das áreas acadêmica e industrial.

A Abordagem Agnóstica pode ser expandida para uma análise ainda mais ampla da área de *LegalTech* através de uma metodologia "*top-down*". Isto é, ao invés de pesquisar uma tarefa específica da área e, a partir dela, detectar necessidades inerentes, é relevante iniciar uma *survey* partindo de um escopo mais amplo e analisar como a área está subdividida atualmente. Analogamente, também é possível realizar outra pesquisa partindo de uma tarefa específica de *LegalTech* e unir os resultados aos já apresentados para a pesquisa por geração de texto legal.

Outra possibilidade para a Abordagem Agnóstica é aperfeiçoar as etapas apresentadas. A abordagem descrita é adaptável e flexível a mudanças, por isso o conceito de agnóstico. Logo, estudos mais elaborados para as necessidades da área podem tornar mais a abordagem mais assertiva, fornecendo um equilíbrio entre abstrações, para uma gama maior de soluções, e especificidades de cada diretiva empregada.

Ainda, é interessante observar o poder de abstração -e o desempenho de forma geral- da Abordagem Agnóstica em um domínio diferente. Isto é, domínios com grande quantidade de

textos e documentos, como a área da saúde, por exemplo, podem se valer das recomendações apresentadas para melhorar a qualidade das soluções em diferentes das tarefas.

Em relação ao protótipo desenvolvido, diversas melhorias pontuais estão planejadas para os serviços implementados. Para a tarefa de geração de texto legal, uma possibilidade de trabalho futuro está relacionado em complementar a anotação do texto com tradução de termos jurídicos para palavras acessíveis a leigos da área, de forma a diminuir a carga cognitiva de leitura. Adicionalmente, as melhorias de usabilidade da solução poderiam incluir o uso de *hyperlinks* em certas palavras, fazendo uso de técnicas de sistemas especialistas para facilitar a navegação entre o acórdão e entidades associadas (pessoas, outros acórdãos e/ou documentos legais, explicações de termos jurídicos).

A tarefa de classificação de documentos focou na realização de experimentos para encontrar o modelo classificador com melhor F1-score entre as técnicas escolhidas. Um trabalho futuro imediato para esta tarefa seria a simulação com configurações adicionais de *stopwords*, *datasets*, e *features*, além de técnicas de *feature tuning* e *parameter tuning* mais robustas nos modelos. Também, a utilização de outras famílias de classificadores enriqueceria os *benchmarks* do protótipo. Entretanto, além da análise de desempenho, é interessante continuar o fluxo do serviço, utilizando a predição do modelo classificador para automatizar esta tarefa e informar a classe do acórdão sem a necessidade da execução manual por pessoas desta etapa.

Pretende-se, ainda, aperfeiçoar os serviços referentes à extração de texto e de informações em páginas do STF. O objetivo é continuar expandindo a quantidade de dados e coleções de acórdãos para fornecer insumos a outros trabalhos acadêmicos. Ainda, não apenas para acórdãos do STF, mas explorar outras bases de dados *online* com informações jurídicas para serem armazenadas em coleções estruturadas de dados, facilitando a utilização e implementação de soluções em IA, PLN, e outras áreas de interesse.

Por fim, outro trabalho futuro interessante é estender o número de *queries* do sistema, criando outras visualizações para analisar os dados coletados. Por exemplo, visualizações que apresentem a quantidade de votos de ministro por acórdão, de votos de um determinado ministro por ano, tempo entre o julgamento e a publicação do acórdão. Além disso, é prevista a busca de recomendações por parte de especialistas da área de visualização de informações para melhorar a representação visual dos dados significativamente. Com uma coleção mais robusta de dados e visualizações, o protótipo pode ser enriquecido para oferecer serviços orientados a *dashboards* de *Business Intelligence* a fim de tornar a exploração dos dados mais interativa.

REFERÊNCIAS

- AGNOLONI, T.; BACCI, L.; OPIJNEN, M. van. Bo-ecli parser engine: the extensible european solution for the automatic extraction of legal links. In: **ASAIL@ICAIL**. [S.l.: s.n.], 2017.
- AISHATH, M. et al. Survey on artificial intelligence. **International Journal of Computer Sciences and Engineering**, v. 7, p. 1778–1790, 05 2019.
- ALETRAS, N. et al. Predicting judicial decisions of the european court of human rights: a natural language processing perspective. **PeerJ Comput. Sci.**, v. 2, p. e93, 2016.
- ALEVEN, V. Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. **Artif. Intell.**, Elsevier Science Publishers Ltd., GBR, v. 150, n. 1–2, p. 183–237, nov. 2003. ISSN 0004-3702. Available from Internet: <[https://doi.org/10.1016/S0004-3702\(03\)00105-X](https://doi.org/10.1016/S0004-3702(03)00105-X)>.
- ALLEN, L. E. Some examples of using the legal relations language in the legal domain: Applied deontic logic. In: . [S.l.: s.n.], 1998. v. 73, n. 3.
- ALLEN, L. E.; TURY, A. W. Newmint interpretation assistance system: United states constitution first amendment's initial 1344 interpretations. In: **Proceedings of the 11th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2007. (ICAIL '07), p. 161–165. ISBN 9781595936806. Available from Internet: <<https://doi.org/10.1145/1276318.1276348>>.
- ALSCHNER, W.; SKOUGAREVSKIY, D. Towards an automated production of legal texts using recurrent neural networks. In: **Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2017. (ICAIL '17), p. 229–232. ISBN 9781450348911. Available from Internet: <<https://doi.org/10.1145/3086512.3086536>>.
- ANDERSSON, L. et al. When is the time ripe for natural language processing for patent passage retrieval? In: . [S.l.: s.n.], 2016. p. 1453–1462.
- ARAUJO, P. H. Luz de et al. Lener-br: A dataset for named entity recognition in brazilian legal text: 13th international conference, propor 2018, canela, brazil, september 24–26, 2018, proceedings. In: _____. [S.l.: s.n.], 2018. p. 313–323. ISBN 978-3-319-99721-6.
- ASHLEY, K.; BR, S. A predictive role for intermediate legal concepts. In: . [S.l.: s.n.], 2003.
- ASHLEY, K. D.; BRÜNINGHAUS, S. Computer models for legal prediction. **Jurimetrics**, American Bar Association, v. 46, n. 3, p. 309–352, 2006. ISSN 08971277, 21544344. Available from Internet: <<http://www.jstor.org/stable/29762941>>.
- ASHLEY, K. D.; WALKER, V. R. Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In: **Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2013. (ICAIL '13), p. 176–180. ISBN 9781450320801. Available from Internet: <<https://doi.org/10.1145/2514601.2514622>>.
- ATKINSON, K.; BENCH-CAPON, T. Legal case-based reasoning as practical reasoning. **Artif. Intell. Law**, Kluwer Academic Publishers, USA, v. 13, n. 1, p. 93–131, mar. 2005. ISSN 0924-8463. Available from Internet: <<https://doi.org/10.1007/s10506-006-9003-3>>.
- BA, K.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. v. 2, 01 2007.

BENCH-CAPON, T. et al. A history of ai and law in 50 papers: 25 years of the international conference on ai and law. **Artificial Intelligence and Law**, v. 20, p. 215–319, 09 2012.

BENCH-CAPON, T.; SARTOR, G. A model of legal reasoning with cases incorporating theories and values. **Artificial Intelligence**, v. 150, n. 1, p. 97–143, 2003. ISSN 0004-3702. AI and Law. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S0004370203001085>>.

BERMAN, D. H.; HAFNER, C. D. Representing teleological structure in case-based legal reasoning: The missing link. In: **Proceedings of the 4th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 1993. (ICAIL '93), p. 50–59. ISBN 0897916069. Available from Internet: <<https://doi.org/10.1145/158976.158982>>.

BOLIOLI, A.; MERCATALI, P.; ROMANO, F. Formal models for a legislative grammar. explicit text amendment. In: . [S.l.: s.n.], 2004. v. 3035, p. 210–227. ISBN 978-3-540-22002-2.

BONIOL, P. et al. Performance in the courtroom: Automated processing and visualization of appeal court decisions in france. In: . arXiv, 2020. Available from Internet: <<https://arxiv.org/abs/2006.06251>>.

BOUAYAD-AGHA, N. et al. Improving the comprehension of legal documentation: The case of patent claims. In: **Proceedings of the 12th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2009. (ICAIL '09), p. 78–87. ISBN 9781605585970. Available from Internet: <<https://doi.org/10.1145/1568234.1568244>>.

BRANTING, K. Techniques for automated drafting of judicial documents. **Int. J. Law Inf. Technol.**, v. 6, p. 214–229, 1998.

BRANTING, L. K. Building explanations from rules and structured cases. **Int. J. Man-Mach. Stud.**, Academic Press Ltd., GBR, v. 34, n. 6, p. 797–837, jun. 1991. ISSN 0020-7373. Available from Internet: <[https://doi.org/10.1016/0020-7373\(91\)90012-V](https://doi.org/10.1016/0020-7373(91)90012-V)>.

BRAZ, F. et al. Document classification using a bi-lstm to unclog brazil's supreme court. In: . [S.l.: s.n.], 2018.

BREUKER, J.; PETKOV, E.; WINKELS, R. Drafting and validating regulations: The inevitable use of intelligent tools. In: . [S.l.: s.n.], 2000. p. 21–33.

BRÜNINGHAUS, S.; ASHLEY, K. D. Toward adding knowledge to learning algorithms for indexing legal cases. In: **Proceedings of the 7th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 1999. (ICAIL '99), p. 9–17. ISBN 1581131658. Available from Internet: <<https://doi.org/10.1145/323706.323709>>.

BRÜNINGHAUS, S.; ASHLEY, K. D. Improving the representation of legal case texts with information extraction methods. In: **Proceedings of the 8th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2001. (ICAIL '01), p. 42–51. ISBN 1581133685. Available from Internet: <<https://doi.org/10.1145/383535.383540>>.

BRÜNINGHAUS, S.; ASHLEY, K. D. Generating legal arguments and predictions from case texts. In: **Proceedings of the 10th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2005. (ICAIL '05), p. 65–74. ISBN 1595930817. Available from Internet: <<https://doi.org/10.1145/1165485.1165497>>.

BRÜGMANN, S. et al. Towards content-oriented patent document processing: Intelligent patent analysis and summarization. **World Patent Information**, v. 40, p. 30–42, 2015. ISSN 0172-2190. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S0172219014001410>>.

BRÜNINGHAUS, S.; ASHLEY, K. Predicting outcomes of case-based legal arguments. In: . [S.l.: s.n.], 2003. p. 233–242.

CASANOVAS, P. et al. Semantic web for the legal domain: The next step. **Semantic Web**, v. 7, p. 213–227, 03 2016.

CHALKIDIS, I.; ANDROUTSOPOULOS, I.; ALETRAS, N. Neural legal judgment prediction in english. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. Florence, Italy: Association for Computational Linguistics, 2019. p. 4317–4323. Available from Internet: <<https://aclanthology.org/P19-1424>>.

CHALKIDIS, I.; ANDROUTSOPOULOS, I.; MICHOS, A. Extracting contract elements. In: . [S.l.: s.n.], 2017. p. 19–28.

CHALKIDIS, I. et al. Large-scale multi-label text classification on EU legislation. In: **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. Florence, Italy: Association for Computational Linguistics, 2019. p. 6314–6322. Available from Internet: <<https://www.aclweb.org/anthology/P19-1636>>.

CHALKIDIS, I.; FERGADIOTIS, M.; ANDROUTSOPOULOS, I. Multieurlex – a multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer. In: . arXiv, 2021. Available from Internet: <<https://arxiv.org/abs/2109.00904>>.

CHEN, H. et al. Charge-based prison term prediction with deep gating network. **CoRR**, abs/1908.11521, 2019. Available from Internet: <<http://arxiv.org/abs/1908.11521>>.

CHEN, H. et al. Improved naive bayes classification algorithm for traffic risk management. **EURASIP Journal on Advances in Signal Processing**, v. 2021, n. 1, p. 30, Jun 2021. ISSN 1687-6180. Available from Internet: <<https://doi.org/10.1186/s13634-021-00742-6>>.

CHEN, H. et al. A comparative study of automated legal text classification using random forests and deep learning. **Information Processing & Management**, v. 59, p. 102798, 03 2022.

CHEN, Y.-L.; LIU, Y.; HO, W.-L. A text mining approach to assist the general public in the retrieval of legal documents. **Journal of the American Society for Information Science and Technology**, v. 64, p. 280–290, 02 2013.

CHITTA, R.; HUDEK, A. K. A reliable and accurate multiple choice question answering system for due diligence. In: **Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2019. (ICAIL '19), p. 184–188. ISBN 9781450367547. Available from Internet: <<https://doi.org/10.1145/3322640.3326711>>.

CHORLEY, A.; BENCH-CAPON, T. Agatha: Using heuristic search to automate the construction of case law theories. **Artif. Intell. Law**, Kluwer Academic Publishers, USA, v. 13, n. 1, p. 9–51, mar. 2005. ISSN 0924-8463. Available from Internet: <<https://doi.org/10.1007/s10506-006-9004-2>>.

CHOWDHURY, G. Natural language processing. **ARIST**, v. 37, p. 51–89, 01 2005.

CIFUENTES-SILVA, F.; GAYO, J. E. L. Legislative document content extraction based on semantic web technologies. In: HITZLER, P. et al. (Ed.). **The Semantic Web**. Cham: Springer International Publishing, 2019. p. 558–573. ISBN 978-3-030-21348-0.

CONRAD, J. et al. Effective document clustering for large heterogeneous law firm collections. In: . [S.l.: s.n.], 2005. p. 177–187.

COOK, S. et al. The applications of artificial intelligence to law: A survey of six current projects. In: **Proceedings of the May 4-7, 1981, National Computer Conference**. New York, NY, USA: Association for Computing Machinery, 1981. (AFIPS '81), p. 689–696. ISBN 9781450379212. Available from Internet: <<https://doi.org/10.1145/1500412.1500516>>.

CUNNINGHAM, C. et al. Investigating graphs in textual case-based reasoning. In: . [S.l.: s.n.], 2004. v. 3155, p. 573–586. ISBN 978-3-540-22882-0.

DALE, R. Law and word order: Nlp in legal tech. **Natural Language Engineering**, v. 25, p. 211–217, 01 2019.

DANIELS, J.; RISSLAND, E. L. Finding legally relevant passages in case opinions. In: **ICAIL '97**. [S.l.: s.n.], 1997.

DEANE, M. A default-logic paradigm for legal fact-finding. In: . [S.l.: s.n.], 2019.

DEEDMAN, C.; GELBART, D.; COLEMAN, M. Slate: Specialized legal automated term extraction. In: TJOA, A. M.; RAMOS, I. (Ed.). **Database and Expert Systems Applications**. Vienna: Springer Vienna, 1992. p. 133–137. ISBN 978-3-7091-7557-6.

DEVIKA, M.; SUNITHA, C.; GANESH, A. Sentiment analysis: A comparative study on different approaches. **Procedia Computer Science**, v. 87, p. 44–49, 2016. ISSN 1877-0509. Fourth International Conference on Recent Trends in Computer Science & Engineering (ICRTCSE 2016). Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S187705091630463X>>.

DO, P.-K. et al. Legal question answering using ranking svm and deep convolutional neural network. 03 2017.

DOZIER, C. et al. Fast tagging of medical terms in legal text. In: . [S.l.: s.n.], 2007. p. 253–260.

ENNALS, R. Expert systems in law—a jurisprudential inquiry by richard susskind, oxford university press, 1987, pp 300, £30.00, isbn 0-19-825582-9. **The Knowledge Engineering Review**, Cambridge University Press, v. 4, n. 2, p. 178–179, 1989.

EVGENIOU, T.; PONTIL, M. Support vector machines: Theory and applications. In: . [S.l.: s.n.], 2001. v. 2049, p. 249–257.

FALAKMASIR, M. H.; ASHLEY, K. Utilizing vector space models for identifying legal factors from text. In: **JURIX**. [S.l.: s.n.], 2017.

FARZINDAR, A.; LAPALME, G. Letsum, an automatic legal text summarizing system. **Jurix**, p. 11–18, 01 2004.

FARZINDAR, A.; LAPALME, G. Machine translation of legal information and its evaluation. In: GAO, Y.; JAPKOWICZ, N. (Ed.). **Advances in Artificial Intelligence**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 64–73. ISBN 978-3-642-01818-3.

FAWEI, B. et al. A semi-automated ontology construction for legal question answering. **New Generation Computing**, v. 37, p. 453 – 478, 2019.

FETERIS, E. T. The role of pragmatic argumentation referring to consequences, goals, and values in the justification of judicial decisions. In: . [S.l.: s.n.], 2015.

FEUERRIEGEL, S.; DOLATA, M.; SCHWABE, G. Fair ai: Challenges and opportunities. **Business & Information Systems Engineering**, v. 62, 05 2020.

FILTZ, E.; KIRRANE, S.; POLLERES, A. The linked legal data landscape: linking legal data across different countries. **Artificial Intelligence and Law**, Feb 2021. ISSN 1572-8382. Available from Internet: <<https://doi.org/10.1007/s10506-021-09282-8>>.

FILTZ, E. et al. Exploiting eurovoc's hierarchical structure for classifying legal documents. In: PANETTO, H. et al. (Ed.). **On the Move to Meaningful Internet Systems: OTM 2019 Conferences**. Cham: Springer International Publishing, 2019. p. 164–181. ISBN 978-3-030-33246-4.

FILTZ, E. et al. Events matter: Extraction of events from court decisions. In: **JURIX**. [S.l.: s.n.], 2020.

FLESCHE, J.; GALANTE, R. A survey on how legaltech research improves the automatic text generation in legal documents. 2022.

GALGANI, F.; COMPTON, P.; HOFFMANN, A. Combining different summarization techniques for legal text. In: **Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data**. USA: Association for Computational Linguistics, 2012. (HYBRID '12), p. 115–123.

GALGANI, F.; COMPTON, P.; HOFFMANN, A. Towards automatic generation of catchphrases for legal case reports. In: . [S.l.: s.n.], 2012. v. 7182, p. 414–425.

GALGANI, F.; HOFFMANN, A. Lexa: Towards automatic legal citation classification. In: LI, J. (Ed.). **AI 2010: Advances in Artificial Intelligence**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 445–454. ISBN 978-3-642-17432-2.

GELBART, D.; SMITH, J. C. Flexicon: An evaluation of a statistical ranking model adapted to intelligent legal text management. In: **Proceedings of the 4th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 1993. (ICAIL '93), p. 142–151. ISBN 0897916069. Available from Internet: <<https://doi.org/10.1145/158976.158994>>.

GLASER, I.; SCEPANKOVA, E.; MATTHES, F. Classifying semantic types of legal sentences: Portability of machine learning models. In: . [S.l.: s.n.], 2018.

GOLTZ, N.; DONDOLI, G.; CAMERON-HUFF, a. Rethinking global regulation: World's law meets artificial intelligence. **Information & Communications Technology Law**, v. 28, 11 2018.

GRABMAIR, M. Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. In: **Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2017. (ICAIL '17), p. 89–98. ISBN 9781450348911. Available from Internet: <<https://doi.org/10.1145/3086512.3086521>>.

GRABMAIR, M.; ASHLEY, K. D. Argumentation with value judgments an example of hypothetical reasoning. In: **Proceedings of the 2010 Conference on Legal Knowledge and Information Systems: JURIX 2010: The Twenty-Third Annual Conference**. NLD: IOS Press, 2010. p. 67–76. ISBN 9781607506812.

GRABMAIR, M.; ASHLEY, K. D. Facilitating case comparison using value judgments and intermediate legal concepts. In: **Proceedings of the 13th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2011. (ICAIL '11), p. 161–170. ISBN 9781450307550. Available from Internet: <<https://doi.org/10.1145/2018358.2018382>>.

GREENLEAF, G.; MOWBRAY, A.; DIJK, P. van. Representing and using legal knowledge in integrated decision support systems: Datalex workstations. **Artificial Intelligence and Law**, v. 3, n. 1, p. 97–142, Mar 1995. ISSN 1572-8382. Available from Internet: <<https://doi.org/10.1007/BF00877696>>.

GROVER, C. et al. Automatic summarisation of legal documents. In: **Proceedings of the 9th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2003. (ICAIL '03), p. 243–251. ISBN 1581137478. Available from Internet: <<https://doi.org/10.1145/1047788.1047839>>.

GUO, Z. et al. A content-based recommendation framework for judicial cases. In: CHENG, X. et al. (Ed.). **Data Science**. Singapore: Springer Singapore, 2019. p. 76–88. ISBN 978-981-15-0118-0.

GUO, Z. et al. Taujud: Test augmentation of machine learning in judicial documents. In: **Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis**. New York, NY, USA: Association for Computing Machinery, 2020. (ISSTA 2020), p. 549–552. ISBN 9781450380089. Available from Internet: <<https://doi.org/10.1145/3395363.3404364>>.

HACHEY, B.; GROVER, C. Automatic legal text summarisation: Experiments with summary structuring. In: . [S.l.: s.n.], 2005. p. 75–84.

HACHEY, B.; GROVER, C. Extractive summarisation of legal texts. **Artificial Intelligence and Law**, v. 14, p. 305–345, 11 2006.

HARPER, C. M.; ZHANG, S. S. Legal tech and lawtech: Towards a framework for technological trends in the legal services industry. In: _____. **Market Engineering : Insights from Two Decades of Research on Markets and Information**. Cham: Springer International Publishing, 2021. p. 183–197. ISBN 978-3-030-66661-3. Available from Internet: <https://doi.org/10.1007/978-3-030-66661-3_11>.

HAUSLADEN, C. I.; SCHUBERT, M. H.; ASH, E. Text classification of ideological direction in judicial opinions. **International Review of Law and Economics**, v. 62, p. 105903, 2020. ISSN 0144-8188. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S0144818819303667>>.

HE, T. et al. Ptm: A topic model for the inferring of the penalty. **Journal of Computer Science and Technology**, v. 33, p. 756–767, 2018.

HOLZENBERGER, N.; BLAIR-STANEK, A.; DURME, B. V. A dataset for statutory reasoning in tax law entailment and question answering. In: . arXiv, 2020. Available from Internet: <<https://arxiv.org/abs/2005.05257>>.

HOVY, E. H. Automated discourse generation using discourse structure relations. **Artificial Intelligence**, v. 63, n. 1, p. 341 – 385, 1993. ISSN 0004-3702. Available from Internet: <<http://www.sciencedirect.com/science/article/pii/0004370293900213>>.

HU, Z. et al. Few-shot charge prediction with discriminative legal attributes. In: **Proceedings of the 27th International Conference on Computational Linguistics**. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. p. 487–498. Available from Internet: <<https://www.aclweb.org/anthology/C18-1041>>.

HUANG, W. et al. Generating reasonable legal text through the combination of language modeling and question answering. In: BESSIERE, C. (Ed.). **Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20**. International Joint Conferences on Artificial Intelligence Organization, 2020. p. 3687–3693. Main track. Available from Internet: <<https://doi.org/10.24963/ijcai.2020/510>>.

IBM. **What is Random Forest?** 2022. Available from Internet: <<https://www.ibm.com/cloud/learn/random-forest>>.

JIA, J. et al. Concise comparative summaries (ccs) of large text corpora with a human experiment. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 8, n. 1, p. 499–529, Mar 2014. ISSN 1932-6157. Available from Internet: <<http://dx.doi.org/10.1214/13-AOAS698>>.

K, S.; JOSEPH, S. Text classification by augmenting bag of words (bow) representation with co-occurrence feature. **IOSR Journal of Computer Engineering**, v. 16, p. 34–38, 01 2014.

KANAPALA, A.; PAL, S.; PAMULA, R. Text summarization from legal documents: A survey. **Artif. Intell. Rev.**, Kluwer Academic Publishers, USA, v. 51, n. 3, p. 371–402, mar. 2019. ISSN 0269-2821. Available from Internet: <<https://doi.org/10.1007/s10462-017-9566-2>>.

KEYMANESH, M.; ELSNER, M.; SARTHASARATHY, S. Toward domain-guided controllable summarization of privacy policies. In: **NLLP@KDD**. [S.l.: s.n.], 2020.

KIM, m.-y.; XU, Y.; GOEBEL, R. Applying a convolutional neural network to legal question answering. In: . [S.l.: s.n.], 2017. p. 282–294. ISBN 978-3-319-50952-5.

KIM, Y. Convolutional neural networks for sentence classification. In: **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)**. Doha, Qatar: Association for Computational Linguistics, 2014. p. 1746–1751. Available from Internet: <<https://www.aclweb.org/anthology/D14-1181>>.

KIYAVITSKAYA, N. et al. Automating the extraction of rights and obligations for regulatory compliance. In: . [S.l.: s.n.], 2008. v. 5231, p. 154–168. ISBN 978-3-540-87876-6.

KONIARIS, M.; ANAGNOSTOPOULOS, I.; VASSILIOU, Y. Evaluation of diversification techniques for legal information retrieval. **Algorithms**, MDPI AG, v. 10, n. 1, p. 22, Jan 2017. ISSN 1999-4893. Available from Internet: <<http://dx.doi.org/10.3390/a10010022>>.

KONSTANTINOOU, V.; SYKES, J.; YANNOPOULOS, G. N. Can legal knowledge be derived from legal texts? In: **Proceedings of the 4th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 1993. (ICAIL '93), p. 218–227. ISBN 0897916069. Available from Internet: <<https://doi.org/10.1145/158976.159004>>.

KUMAR, S. et al. Similarity analysis of legal judgments. In: . [S.l.: s.n.], 2011. p. 17.

LAGOS, N. et al. Event extraction for legal case building and reasoning. In: . [S.l.: s.n.], 2010. v. 340, p. 92–101. ISBN 978-3-642-16326-5.

LAU, G. T.; LAW, K. H.; WIEDERHOLD, G. Similarity analysis on government regulations. In: **Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2003. (KDD '03), p. 711–716. ISBN 1581137370. Available from Internet: <<https://doi.org/10.1145/956750.956845>>.

- LAWLOR, R. C. What computers can do: Analysis and prediction of judicial decisions. **American Bar Association Journal**, American Bar Association, v. 49, n. 4, p. 337–344, 1963. ISSN 00027596, 21627975. Available from Internet: <<http://www.jstor.org/stable/25722338>>.
- LEHMANN, H. Legal concepts in a natural language based expert system*. **Ratio Juris**, v. 3, p. 245 – 253, 08 2007.
- LEITH, P. Legal expertise and legal expert systems. **International Review of Law, Computers & Technology**, Routledge, v. 2, n. 1, p. 1–24, 1986. Available from Internet: <<https://doi.org/10.1080/13600869.1986.9966227>>.
- LEITNER, E.; REHM, G.; SCHNEIDER, J. Fine-grained named entity recognition in legal documents. In: _____. [S.l.: s.n.], 2019. p. 272–287. ISBN 978-3-030-33219-8.
- LI, W. The advance of deep learning based named entity recognition. **Highlights in Science, Engineering and Technology**, v. 12, p. 68–73, 08 2022.
- LIPPI, M. et al. CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. **Artificial Intelligence and Law**, Springer Science and Business Media LLC, v. 27, n. 2, p. 117–139, feb 2019. Available from Internet: <<https://doi.org/10.1007/978-3-030-09243-2>>.
- LIU, Y.-H.; CHEN, Y.-L.; HO, W.-L. Predicting associated statutes for legal problems. **Information Processing & Management**, v. 51, n. 1, p. 194 – 211, 2015. ISSN 0306-4573. Available from Internet: <<http://www.sciencedirect.com/science/article/pii/S0306457314000636>>.
- LOCKE, D.; ZUCCON, G.; SCELLS, H. Automatic query generation from legal texts for case law retrieval. In: . [S.l.: s.n.], 2017. p. 181–193. ISBN 978-3-319-70144-8.
- LUO, B. et al. Learning to predict charges for criminal cases with legal basis. In: **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**. Copenhagen, Denmark: Association for Computational Linguistics, 2017. p. 2727–2736. Available from Internet: <<https://www.aclweb.org/anthology/D17-1289>>.
- LUO, B. et al. Learning to predict charges for criminal cases with legal basis. In: **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**. Copenhagen, Denmark: Association for Computational Linguistics, 2017. p. 2727–2736. Available from Internet: <<https://www.aclweb.org/anthology/D17-1289>>.
- MANOR, L.; LI, J. J. Plain english summarization of contracts. In: **Proceedings of the Natural Legal Language Processing Workshop 2019**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 1–11. Available from Internet: <<https://aclanthology.org/W19-2201>>.
- MARKOVIĆ, M.; GOSTOJIĆ, S. Legal document management: An integrated approach. In: . [S.l.: s.n.], 2019. p. 374–380.
- MASHAYEKHI, M.; GRAS, R. Rule extraction from random forest: the rf+hc methods. In: . [S.l.: s.n.], 2015. ISBN 978-3-319-18355-8.
- MAXWELL, T.; OBERLANDER, J.; LAVRENKO, V. Evaluation of semantic events for legal case retrieval. **Proceedings of the WSDM'2009 ACM Workshop on Exploiting Semantic Annotations in Information Retrieval, ESAIR 2009**, 01 2009.
- MCCORDUCK, P. **Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence**. [S.l.]: AK Peters Ltd, 2004. ISBN 1568812051.

MCDONALD, D. D.; PUSTEJOVSKY, J. D. The counselor project at the university of massachusetts. In: **Proceedings of the Workshop on Strategic Computing Natural Language**. USA: Association for Computational Linguistics, 1986. (HLT '86), p. 26–29. Available from Internet: <<https://doi.org/10.3115/1077146.1077151>>.

MEDVEDEVA, M.; VOLS, M.; WIELING, M. Using machine learning to predict decisions of the european court of human rights. **Artificial Intelligence and Law**, v. 2019, 06 2020.

Merkel, D.; Schweighofer, E. En route to data mining in legal text corpora: clustering, neural computation, and international treaties. In: **Database and Expert Systems Applications. 8th International Conference, DEXA '97. Proceedings**. [S.l.: s.n.], 1997. p. 465–470.

METSKER, O. G.; TROFIMOV, E. S.; GRECHISHCHEVA, S. Natural language processing of russian court decisions for digital indicators mapping for oversight process control efficiency: Disobeying a police officer case. In: **EGOSE**. [S.l.: s.n.], 2019.

MILLE, S.; WANNER, L. Making text resources accessible to the reader: the case of patent claims. In: **Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)**. Marrakech, Morocco: European Language Resources Association (ELRA), 2008. Available from Internet: <http://www.lrec-conf.org/proceedings/lrec2008/pdf/352_paper.pdf>.

MOCHALES, R.; MOENS, M.-F. Study on the structure of argumentation in case law. In: . [S.l.: s.n.], 2008. p. 11–20.

MOCHALES, R.; MOENS, M.-F. Argumentation mining: The detection, classification and structure of arguments in text. In: . [S.l.: s.n.], 2009. p. 98–107.

MOCHALES, R.; MOENS, M.-F. Argumentation mining: The detection, classification and structure of arguments in text. In: . [S.l.: s.n.], 2009. p. 98–107.

MOENS, M.-F. Innovative techniques for legal text retrieval. **Artificial Intelligence and Law**, v. 9, p. 29–57, 2004.

MOENS, M.-F. Improving access to legal information: How drafting systems help. In: _____. **Information Technology and Lawyers: Advanced Technology in the Legal Domain, from Challenges to Daily Routine**. Dordrecht: Springer Netherlands, 2006. p. 119–136. ISBN 978-1-4020-4146-4. Available from Internet: <https://doi.org/10.1007/1-4020-4146-2_5>.

MOENS, M.-F.; UYTENDAELE, C.; DUMORTIER, J. Information extraction from legal texts: the potential of discourse analysis. **International Journal of Human-Computer Studies**, v. 51, n. 6, p. 1155 – 1171, 1999. ISSN 1071-5819. Available from Internet: <<http://www.sciencedirect.com/science/article/pii/S107158199902963>>.

MOLES, R. N.; DAYAL, S. **The Open Texture of Language: Handling Semantic Analysis in Legal Decision Support Systems**. University of Tasmania, 1993. 330–367 p. Available from Internet: <<https://search.informit.org/doi/10.3316/agispt.19941712>>.

MOULIN, B.; ROUSSEAU, D. Designing deontic knowledge bases from regulation texts. **Know.-Based Syst.**, Elsevier Science Publishers B. V., NLD, v. 3, n. 2, p. 108–120, jun. 1990. ISSN 0950-7051. Available from Internet: <[https://doi.org/10.1016/0950-7051\(90\)90004-2](https://doi.org/10.1016/0950-7051(90)90004-2)>.

MOULIN, B.; ROUSSEAU, D. Knowledge acquisition from prescriptive texts. In: **Proceedings of the 3rd International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems - Volume 2**. New York, NY, USA: Association for Computing Machinery, 1990. (IEA/AIE '90), p. 1112–1121. ISBN 0897913728. Available from Internet: <<https://doi.org/10.1145/98894.99136>>.

MOZGOVOY, M.; KAKKONEN, T.; COSMA, G. Automatic student plagiarism detection: Future perspectives. **Journal of Educational Computing Research**, v. 43, 01 2010.

MUNISAMI, K. Legal technology and the future of women in law. **Windsor Yearbook of Access to Justice**, v. 36, p. 164–183, 09 2020.

MUNISAMI, K. Legal technology and the future of women in law. **Windsor Yearbook of Access to Justice**, v. 36, p. 164–183, 09 2020.

NAGEL, S. S. Applying correlation analysis to case prediction. **Tex. L. Rev.**, HeinOnline, v. 42, p. 1006, 1963.

NAVAS-LORO, M.; SANTOS, C. Events in the legal domain: first impressions. In: . [S.l.: s.n.], 2019.

NAVAS-LORO, M.; SATOH, K.; RODRÍGUEZ-DONCEL, V. Contractframes: Bridging the gap between natural language and logics in contract law. In: _____. [S.l.: s.n.], 2019. p. 101–114. ISBN 978-3-030-31604-4.

NAZARENKO, A.; LEVY, F.; WYNER, A. An Annotation Language for Semantic Search of Legal Sources. In: CHAIR), N. C. C. et al. (Ed.). **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**. Miyazaki, Japan: European Language Resources Association (ELRA), 2018. ISBN 979-10-95546-00-9.

NAZARENKO, A.; WYNER, A. Legal nlp introduction. **TAL Traitement Automatique des Langues**, v. 58, p. 7–19, 01 2017.

NEJADGHOLI, I.; TCHEMEUBE, R. B.; WITHERSPOON, S. A semi-supervised training method for semantic search of legal facts in canadian immigration cases. In: . [S.l.: s.n.], 2017.

NEWMAN, D.; DOHERTY, U. Making the law accessible to non-lawyers: Effects of different kinds of expertise on perceived usability of online legal information services. **Behaviour & IT**, v. 27, p. 423–437, 09 2008.

OMAR, M. et al. Robust natural language processing: Recent advances, challenges, and future directions. 01 2022.

OPIJNEN, M. Citation analysis and beyond: In search of indicators measuring case law importance. **Frontiers in Artificial Intelligence and Applications**, v. 250, p. 95–104, 01 2012.

OPIJNEN, M.; SANTOS, C. On the concept of relevance in legal information retrieval. **Artificial Intelligence and Law**, v. 25, p. 65–87, 03 2017.

PEÑAS, A. et al. Overview of respubliqa 2009: Question answering evaluation over european legislation. In: . [S.l.: s.n.], 2010. p. 174–196. ISBN 978-3-642-15753-0.

PERIC, L. et al. Legal language modeling with transformers. In: ASHLEY, K. D. et al. (Ed.). **Proceedings of the Fourth Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2020) held online in conjunction with the 33rd International Conference on Legal Knowledge and Information Systems (JURIX 2020) December 9, 2020**. s.l.: CEUR-WS, 2020–12. v. 2764. ISSN 1613-0073. 4th Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2020); Conference Location: online; Conference Date: December 9, 2020; Due to the Coronavirus (COVID-19) the conference was conducted virtually.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Inf. Softw. Technol.**, v. 64, p. 1–18, 2015.

QUINTANA, Y.; KAMEL, M.; LO, A. Graph-based retrieval of information in hypertext systems. In: **Proceedings of the 10th Annual International Conference on Systems Documentation**. New York, NY, USA: Association for Computing Machinery, 1992. (SIGDOC '92), p. 157–168. ISBN 0897915321. Available from Internet: <<https://doi.org/10.1145/147001.147022>>.

RAVICHANDER, A. et al. Question answering for privacy policies: Combining computational and legal perspectives. In: **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**. Hong Kong, China: Association for Computational Linguistics, 2019. p. 4949–4959. Available from Internet: <<https://www.aclweb.org/anthology/D19-1500>>.

ROEGUEST, A.; HUDEK, A. K.; MCNULTY, A. A dataset and an examination of identifying passages for due diligence. In: **The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval**. New York, NY, USA: Association for Computing Machinery, 2018. (SIGIR '18), p. 465–474. ISBN 9781450356572. Available from Internet: <<https://doi.org/10.1145/3209978.3210015>>.

ROSSI, J.; WIRTH, M.; KANOULAS, E. **Query Generation for Patent Retrieval with Keyword Extraction based on Syntactic Features**. 2019.

SALMERÓN-MANZANO, E. Legaltech and lawtech: Global perspectives, challenges, and opportunities. **Laws**, v. 10, p. 24, 04 2021.

SALMERÓN-MANZANO, E. Legaltech and lawtech: Global perspectives, challenges, and opportunities. **Laws**, v. 10, p. 24, 04 2021.

SANNIER, N. et al. An automated framework for detection and resolution of cross references in legal texts. **Requirements Engineering**, v. 22, 11 2015.

Sannier, N. et al. Legal markup generation in the large: An experience report. In: **2017 IEEE 25th International Requirements Engineering Conference (RE)**. [S.l.: s.n.], 2017. p. 302–311.

SARAVANAN, M.; RAVINDRAN, B.; RAMAN, S. Improving legal document summarization using graphical models. In: **Proceedings of the 2006 Conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference**. NLD: IOS Press, 2006. p. 51–60. ISBN 158603698X.

SAXON, C. Computer-aided drafting of legal documents. **Law & Social Inquiry**, v. 7, p. 685 – 754, 07 2006.

SHAHAB, A.; KIENINGER, T.; DENGEL, A. The aware toolbox for the detection of law infringements on web pages. In: . [S.l.: s.n.], 2010. v. 7534, p. 1–10.

SHAHEEN, Z.; WOHLGENANT, G.; FILTZ, E. **Large Scale Legal Text Classification Using Transformer Models**. 2020.

SHEREMETYEVA, S. Natural language analysis of patent claims. In: **ACL 2003**. [S.l.: s.n.], 2003.

SHINMORI, A.; MARUKAWA, Y. Patent claim processing for readability: Structure analysis and term explanation. 01 2003.

SILVA, N.; BRAZ, F.; CAMPOS, T. de. Document type classification for brazil's supreme court using a convolutional neural network. In: . [S.l.: s.n.], 2018. p. 7–11.

- SINGH, J.; SHARMA, Y. Encoder-decoder architectures for generating questions. **Procedia Computer Science**, v. 132, p. 1041–1048, 2018. ISSN 1877-0509. International Conference on Computational Intelligence and Data Science. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S1877050918307518>>.
- SINH, V.; NGUYEN, L.; SATOH, K. Legal text generation from abstract meaning representation. **Proceedings of JURIX 2019**, p. 229–234, 12 2019.
- SOUSA, A. W.; FABRO, M. Iudicium textum dataset uma base de textos jurídicos para nlp. In: . [S.l.: s.n.], 2019.
- SPAETH, H. J. et al. **Supreme Court Database**. Washington University Law, 2020. Available from Internet: <<http://supremecourtdatabase.org/>>.
- SPROWL, J. Automating the legal reasoning process: A computer that uses regulations and statutes to draft legal documents. **Law & Social Inquiry**, v. 4, p. 1 – 81, 07 2006.
- SPROWL, J. et al. An expert system for drafting legal documents. In: **Proceedings of the July 9-12, 1984, National Computer Conference and Exposition**. New York, NY, USA: Association for Computing Machinery, 1984. (AFIPS '84), p. 667–673. ISBN 0882830430. Available from Internet: <<https://doi.org/10.1145/1499310.1499396>>.
- SPROWL, J. A.; STAUD†, R. W. Computerizing client services in the law school teaching clinic: An experiment in law office automation. **Law & Social Inquiry**, v. 6, n. 3, p. 699–751, 1981. Available from Internet: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1747-4469.1981.tb00436.x>>.
- TUGGENER, D. et al. LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In: **Proceedings of the 12th Language Resources and Evaluation Conference**. Marseille, France: European Language Resources Association, 2020. p. 1235–1241. ISBN 979-10-95546-34-4. Available from Internet: <<https://www.aclweb.org/anthology/2020.lrec-1.155>>.
- TUGGENER, D. et al. Ledger: A large-scale multi-label corpus for text classification of legal provisions in contracts. In: **Proceedings of the 12th Language Resources and Evaluation Conference**. Marseille, France: European Language Resources Association, 2020. p. 1235–1241. ISBN 979-10-95546-34-4. Available from Internet: <<https://aclanthology.org/2020.lrec-1.155>>.
- TURTLE, H. Natural language vs. boolean query evaluation: A comparison of retrieval performance. In: CROFT, B. W.; RIJSBERGEN, C. J. van (Ed.). **SIGIR '94**. London: Springer London, 1994. p. 212–220. ISBN 978-1-4471-2099-5.
- TURTLE, H. Text retrieval in the legal world. **Artif. Intell. Law**, Kluwer Academic Publishers, USA, v. 3, n. 1–2, p. 5–54, mar. 1995. ISSN 0924-8463. Available from Internet: <<https://doi.org/10.1007/BF00877694>>.
- VERHEIJ, B. Formalizing correct evidential reasoning with arguments , scenarios and probabilities. In: . [S.l.: s.n.], 2016.
- VöLKER, J.; LANGA, S. F.; SURE, Y. Supporting the construction of spanish legal ontologies with text2onto. In: _____. **Computable Models of the Law: Languages, Dialogues, Games, Ontologies**. Berlin, Heidelberg: Springer-Verlag, 2008. p. 105–112. ISBN 9783540855682. Available from Internet: <https://doi.org/10.1007/978-3-540-85569-9_7>.
- VU, T. S.; NGUYEN, L. M. An empirical evaluation of amr parsing for legal documents. In: KOJIMA, K. et al. (Ed.). **New Frontiers in Artificial Intelligence**. Cham: Springer International Publishing, 2019. p. 131–145. ISBN 978-3-030-31605-1.

WAGH, R.; ANAND, D. A novel approach of augmenting training data for legal text segmentation by leveraging domain knowledge: Proceedings of ista 2018. In: _____. [S.l.: s.n.], 2020. p. 53–63. ISBN 978-981-13-6094-7.

WALKER, V. R. et al. Semantic types for computational legal reasoning: Propositional connectives and sentence roles in the veterans' claims dataset. In: **Proceedings of the 16th Edition of the International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2017. (ICAIL '17), p. 217–226. ISBN 9781450348911. Available from Internet: <<https://doi.org/10.1145/3086512.3086535>>.

WALTER, S. Linguistic description and automatic extraction of definitions from german court decisions. In: **Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)**. Marrakech, Morocco: European Language Resources Association (ELRA), 2008. Available from Internet: <http://www.lrec-conf.org/proceedings/lrec2008/pdf/742_paper.pdf>.

WANG, P. et al. Modeling dynamic pairwise attention for crime classification over legal articles. In: **The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval**. New York, NY, USA: Association for Computing Machinery, 2018. (SIGIR '18), p. 485–494. ISBN 9781450356572. Available from Internet: <<https://doi.org/10.1145/3209978.3210057>>.

WEHNERT, S. et al. Legal norm retrieval with variations of the bert model combined with tf-idf vectorization. In: **Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2021. (ICAIL '21), p. 285–294. ISBN 9781450385268. Available from Internet: <<https://doi.org/10.1145/3462757.3466104>>.

WIDDISON, R.; PRITCHARD, F.; ROBINSON, W. Expert system meets hypertext : The european conflicts guide. In: . [S.l.: s.n.], 1992.

WYNER, A.; PETERS, W. Lexical semantics and expert legal knowledge towards the identification of legal case factors. In: . [S.l.: s.n.], 2010. v. 223, p. 127–136.

XIAO, C. et al. Cail2018: A large-scale legal dataset for judgment prediction. In: . [S.l.: s.n.], 2018.

XU, J.; CROFT, W. B. Corpus-based stemming using cooccurrence of word variants. **ACM Trans. Inf. Syst.**, Association for Computing Machinery, New York, NY, USA, v. 16, n. 1, p. 61–81, jan. 1998. ISSN 1046-8188. Available from Internet: <<https://doi.org/10.1145/267954.267957>>.

XU, Z. et al. Case facts analysis method based on deep learning. In: _____. [S.l.: s.n.], 2019. p. 92–97. ISBN 978-3-030-30951-0.

YADAV, V.; BETHARD, S. A survey on recent advances in named entity recognition from deep learning models. **CoRR**, abs/1910.11470, 2019. Available from Internet: <<http://arxiv.org/abs/1910.11470>>.

Yamada, H.; Teufel, S.; Tokunaga, T. Designing an annotation scheme for summarizing japanese judgment documents. In: **2017 9th International Conference on Knowledge and Systems Engineering (KSE)**. [S.l.: s.n.], 2017. p. 275–280.

YAN, G. et al. Data augmentation for deep learning of judgment documents. In: CUI, Z. et al. (Ed.). **Intelligence Science and Big Data Engineering. Big Data and Machine Learning**. Cham: Springer International Publishing, 2019. p. 232–242. ISBN 978-3-030-36204-1.

YE, H. et al. Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. In: **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**. New Orleans, Louisiana: Association for Computational Linguistics, 2018. p. 1854–1864. Available from Internet: <<https://www.aclweb.org/anthology/N18-1168>>.

ZELEZNIKOW, J.; HUNTER, D. Rationales for the continued development of legal expert systems. **The Journal of Law and Information Science**, 1992.

ZHANG, P.; KOPPAKA, L. Semantics-based legal citation network. In: **Proceedings of the 11th International Conference on Artificial Intelligence and Law**. New York, NY, USA: Association for Computing Machinery, 2007. (ICAIL '07), p. 123–130. ISBN 9781595936806. Available from Internet: <<https://doi.org/10.1145/1276318.1276342>>.

ZHANG, W.; GAO, F. An improvement to naive bayes for text classification. **Procedia Engineering**, v. 15, p. 2160–2164, 2011. ISSN 1877-7058. CEIS 2011. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S1877705811019059>>.

ZHENG, L. et al. When does pretraining help? assessing self-supervised learning for law and the casehold dataset. In: **Proceedings of the 18th International Conference on Artificial Intelligence and Law**. [S.l.]: Association for Computing Machinery, 2021. (in press).

ZHENG, L. et al. When does pretraining help? assessing self-supervised learning for law and the casehold dataset. arXiv, 2021. Available from Internet: <<https://arxiv.org/abs/2104.08671>>.

ZHONG, H. et al. Legal judgment prediction via topological learning. In: **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**. Brussels, Belgium: Association for Computational Linguistics, 2018. p. 3540–3549. Available from Internet: <<https://www.aclweb.org/anthology/D18-1390>>.

ZHONG, H. et al. **JEC-QA: A Legal-Domain Question Answering Dataset**. 2019.

ZHONG, L. et al. Automatic summarization of legal decisions using iterative masking of predictive sentences. In: . [S.l.: s.n.], 2019. p. 163–172. ISBN 978-1-4503-6754-7.

ZHOU, X. et al. Legal intelligence for e-commerce: Multi-task learning by leveraging multiview dispute representation. In: **Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: Association for Computing Machinery, 2019. (SIGIR'19), p. 315–324. ISBN 9781450361729. Available from Internet: <<https://doi.org/10.1145/3331184.3331212>>.

ÇELİK, O. A research on machine learning methods and its applications. 09 2018.