



**IME-USP**

Universidade de São Paulo  
Instituto de Matemática e Estatística  
Departamento de Computação

Kaique Kazuyoshi Komata  
Yurick Yussuke Honda

**Mapeador Inteligente de Acórdãos do Supremo  
Tribunal Federal para Ramos do Direito**

Orientadora: Renata Wassermann  
Co-orientador: Rafael Brito

São Paulo - SP  
Maio de 2021

## Lista de Figuras

2.1	Barra de busca principal da Pesquisa de Jurisprudência do STF. . . . .	7
2.2	Formulário de busca avançada da Pesquisa de Jurisprudência do STF. . . .	7
2.3	Resultado de uma busca genérica na Pesquisa de Jurisprudência do STF. .	8
2.4	Conteúdo parcial do acórdão identificado pelo código RE 215411. . . . .	8
2.5	Projeto desenvolvido durante o Mestrado do Rafael. . . . .	10
4.1	Gráfico de Gantt com o planejamento das atividades do trabalho. . . . .	13

# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Fundamentos</b>	<b>5</b>
2.1	Os ramos do direito . . . . .	5
2.2	Acórdãos e mecanismo de busca do STF . . . . .	6
2.3	Projetos anteriores . . . . .	9
<b>3</b>	<b>Metodologia</b>	<b>11</b>
3.1	Parceria com a Faculdade de Direito da Universidade de São Paulo . . . . .	11
3.2	Obtenção e adequação dos dados . . . . .	11
3.3	Elaboração do classificador . . . . .	12
<b>4</b>	<b>Cronograma</b>	<b>13</b>
	<b>Referências Bibliográficas</b>	<b>14</b>

# 1 Introdução

O direito foi originado na Pré-História para que pudesse trazer, de acordo com Paulo Nader, formas de regular a vida em sociedade, tendo em vista aspectos como ordem e justiça. Ele também afirma que apenas os fatos sociais cuja importância se mostra sobressalente, devem estar contidos nas fronteiras das disciplinas do direito. Outra maneira de entender o direito seria como uma forma de atender exigências para um convívio ordenado em sociedade. Isso é atingido, em geral, através de regras de comportamento (BETIOLI, 2015).

Por ter uma origem em sociedades primitivas, o direito esteve em constante evolução. Isso pode ser facilmente percebido ao notar que no uso do direito arcaico, aplicado às sociedades primitivas, os conceitos se baseavam no princípio do parentesco. Porquanto, as regras do direito possuíam como estrutura inicial os próprios laços de consanguinidade, além das relações humanas do convívio social (WOLKMER, 2008).

Ao analisar o próprio direito brasileiro, nitidamente confirma-se que muitos desses conceitos evoluíram em comparação à fundação do direito nas sociedades pré-históricas. Isso ocorre porque o sistema jurídico segue em sintonia com o contexto social e econômico nos quais ele está inserido. Ademais, o sistema jurídico de cada unidade administrativa deve estar sempre evoluindo, para que ele possa se situar e ser moldado com relação ao espaço, tempo e cultura específicos daquela unidade. Apesar de todo o conhecimento que possa ser adquirido por experiências prévias em outros países ou Estados, a especificidade de uma unidade administrativa modela suas próprias regras do direito (CONTRUCCI, 2010).

Com um mundo globalizado, as alterações nas normas estabelecidas pelo direito brasileiro, por exemplo, tornam-se ainda mais frequentes. Dessa forma, um novo desafio surge a partir dessa evolução: criar tecnologias que possam servir os propósitos de agentes que estão inseridos na área do direito e que possam acompanhar o progresso dela adequadamente.

Além disso, a produção de dados em massa possibilita que métodos computacionais sejam utilizados em dados específicos do direito de forma a auxiliar no desenvolvimento dessas tecnologias citadas anteriormente. Esses dados podem resultar em uma forma de pessoas já pertencentes à área de estudo, bem como leigos, terem acesso a informações de uma forma simples, mesmo com um grande volume de dados.

Assim como o direito avança a todo momento, a tecnologia pode trazer grandes benefícios com o poder de processamento que se tem atualmente. Isso significa que obter uma grande quantidade de dados pode resultar na extração de padrões e, potencialmente, no desenvolvimento de uma certa inteligência de decisão a partir deles.

Somente possuir dados, no entanto, não significa que também se tem posse de informações concretas. Em situações como essa, ferramentas podem ser criadas para auxiliar na obtenção de informações úteis a partir de um grande conjunto de dados. Um exemplo disso seria a disponibilização de acórdãos por parte do Supremo Tribunal Federal (STF). Por meio de uma busca genérica na Pesquisa da Jurisprudência do STF, apresentada no capítulo a seguir, a quantidade de acórdãos recuperada de seu acervo seria de 322.421. Mesmo com um largo acervo, certas informações ainda precisam ser extraídas do conteúdo do acórdão, ou seja, uma busca no portal do STF<sup>1</sup> não permite filtrar os documentos a partir de certos critérios, por exemplo pelo ramo do direito.

Alguns desses critérios motivaram o início desse projeto, cujo objetivo é desenvolver mecanismos de classificação automática que tenham a capacidade de extrair informações, ainda não suportadas pela ferramenta de busca do STF, do próprio conteúdo do acórdão. Para esse projeto, o foco maior será na classificação de acórdãos por ramo do direito.

No próximo capítulo deste documento serão definidos os conceitos e fundamentos importantes para o projeto como os acórdãos e os ramos do direito. No terceiro capítulo serão apresentadas a metodologia e as tecnologias que serão utilizadas no desenvolvimento deste trabalho. Por fim, no capítulo 4 será apresentado o cronograma previsto do trabalho.

---

<sup>1</sup>Disponível em <<https://jurisprudencia.stf.jus.br/>>.

## 2 Fundamentos

### 2.1 Os ramos do direito

Inicialmente, o direito pode ser dividido entre duas esferas: pública e privada. De acordo com Wilson de Souza Campos Batalha, o direito público seria relacionado à regulação das relações do Estado, sejam elas internas ou resultantes da interação com outros Estados, bem como com seus súditos. O direito privado, em suma, remete à relação dos cidadãos entre si ou com o Estado. Muitos autores deixam claro que essa divisão não possui limites bem definidos, isso significa que traça-se uma linha tênue que limita frouxamente essas duas esferas em seus respectivos contextos (VENOSA, 2019).

Por sua vez, cada uma das esferas compreende subdivisões que iremos nos referir nesse projeto como os ramos do direito. Sendo assim, o direito público concentra os seguintes ramos:

- Direito do trabalho: ramo que trata das relações jurídicas entre empregado e empregador, e abrange normas provenientes de acordos e convenções coletivas de trabalho. Também pertencem a esse ramo as normas que cuidam da representação dos vários sindicatos;
- Direito previdenciário: ramo que cuida dos benefícios e do funcionamento dos órgãos públicos de assistência e previdência social;
- Direito constitucional: ramo que se baseia na Constituição e trata dos princípios e normas relativos à estrutura fundamental do Estado;
- Direito administrativo: ramo que se refere à Administração Pública, englobando a regulação jurídica do poder executivo do Estado;
- Direito financeiro: ramo que trata das finanças públicas, regulando a estrutura orçamentária das entidades públicas;
- Direito tributário: ramo que ordena a forma de arrecadação de tributos e o relacionamento entre o Fisco<sup>1</sup> e o contribuinte;

---

<sup>1</sup>Fisco é o órgão fiscalizador que controla os pagamentos de impostos em todas as esferas tributárias no país.

- Direito processual civil: ramo que regula os procedimentos jurisdicionais, administrando as relações civis;
- Direito judiciário: ramo que abrange a organização, a jurisdição e a competência do Poder Judiciário;
- Direito penal: ramo que diz respeito ao direito de punir do Estado, sendo composto pelos preceitos legais para definir os crimes e determinar aos seus autores as devidas penas;
- Direito internacional: ramo que trabalha com os tratados e acordos internacionais e rege os direitos e deveres internacionais do Estado e dos indivíduos;
- Direito eleitoral: ramo que trata das normas e procedimentos que regulam o exercício do direito ao sufrágio universal; e
- Direito militar: ramo que estuda as relações jurídicas dos militares e regula os direitos, deveres e atividades da categoria.

Já o direito privado concentra os seguintes ramos:

- Direito civil: ramo que trata dos conjuntos de normas jurídicas responsáveis por regular os direitos e obrigações em relação as pessoas e seus bens. É nele que se encontra o Código Civil, uma das leis mais importantes para o sistema judiciário;
- Direito da família: ramo que compreende o conjunto de normas marcadas pelo interesse social. É focado no convívio entre casais e seus filhos;
- Direito comercial: ramo que regula todas as relações jurídicas advindas do comércio, incluindo atividades empresariais e comerciais;
- Direito econômico: ramo que busca o ordenamento da macroeconomia e se refere à intervenção do Estado no domínio econômico;
- Direito ambiental: ramo que é composto pelos princípios e normas reguladoras das atividades humanas que possam afetar a preservação e estado do meio ambiente; e
- Direito do consumidor: ramo que trata do conjunto de regras e princípios jurídicos que regulam a relação entre o consumidor e o fornecedor de bens ou de serviços.

## 2.2 Acórdãos e mecanismo de busca do STF

Um acórdão<sup>2</sup> representa a conclusão de um processo judicial, em especial quando órgãos colegiados de instâncias superiores são envolvidos por conta de apresentação de

recurso das partes envolvidas. Dessa forma, o acórdão representa a decisão dos órgãos colegiados, denotando o encerramento do processo judicial por meio da definição envolvendo vários julgadores. Os acórdãos ficam disponíveis para consulta de formas distintas nos portais dos seus respectivos tribunais.

Para os acórdãos do STF, a consulta deve ser realizada na página de Pesquisa da Jurisprudência no portal do STF. Nessa página é disponibilizada uma ferramenta que permite realizar buscas dos acórdãos arquivados a partir de seus dados. Para realizar essa busca, o usuário digita uma combinação de palavras-chave com conectivos lógicos como *ou* e *não* na barra de busca na parte superior da página.

Figura 2.1: Barra de busca principal da Pesquisa de Jurisprudência do STF.



Fonte: Pesquisa de Jurisprudência no portal do STF.

Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search>>

Também é possível realizar uma busca mais detalhada dos acórdãos na página. Para isso é preciso preencher diversos campos com as informações do acórdão em um formulário.

Figura 2.2: Formulário de busca avançada da Pesquisa de Jurisprudência do STF.

**Bases** ×

Acórdãos  Repercussão geral  Súmulas  Decisões monocráticas  Informativos

**Pesquisa em todos os campos**

Pesquisar palavras-chave

**Pesquisa em campos específicos**

Número / Classe	Ementa / Decisão / Indexação
Tese	Tema
Observação	Partes

**Pesquisa por legislação**

Norma  Número  Art.  Inc.  Let.  Par.  +

**Opções de pesquisa**

Inteiro teor  Sinônimos  Plural  
 Radicais  Busca exata entre aspas

Fonte: Pesquisa de Jurisprudência no portal do STF.

Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search>>

Em ambos os casos, a execução da pesquisa resulta numa listagem dos documentos que se encaixam nos critérios especificados. Nos itens retornados é possível acessar os

<sup>2</sup>Definição disponível em: <<https://www.cnj.jus.br/cnj-servico-saiba-quando-a-decisao-final-e-dada-por-sentenca-ou-em-acordao/>>

dados completos, copiar o conteúdo e acessar o inteiro teor do documento em questão, no formato Portable Document Format (*pdf*).

Figura 2.3: Resultado de uma busca genérica na Pesquisa de Jurisprudência do STF.

The screenshot displays the STF Jurisprudence Search interface. On the left, there are filters for 'Base' (Acórdãos, Repercussão geral, etc.), 'Órgão julgador' (Tribunal Pleno, etc.), and 'Membro' (various ministers). The main area shows search results for '323 423 resultados' for '5' pages. A detailed view of case RE 842848 is shown on the right, including the court (Tribunal Pleno), date (27/02/2019), and a summary (ementa) regarding administrative law and the right to work.

Fonte: Pesquisa de Jurisprudência no portal do STF.

Disponível em: <<https://jurisprudencia.stf.jus.br/pages/search?base=acordaos...>>

A figura 2.4 ilustra a estrutura básica de um acórdão recuperado após realizar uma busca na Pesquisa de Jurisprudência do STF.

Figura 2.4: Conteúdo parcial do acórdão identificado pelo código RE 215411.

**COORD. DE ANÁLISE DE JURISPRUDÊNCIA**  
**D.J. 25.08.2006**

**14/03/2006**      **EMENTÁRIO Nº 2 2 4 4 - 4**      **SEGUNDA TURMA**

**RECURSO EXTRAORDINÁRIO 215.411-3 SÃO PAULO**

**RELATOR ORIGINÁRIO : MIN. NÉRI DA SILVEIRA**  
**RELATOR PARA O ACÓRDÃO : MIN. MARCO AURÉLIO**  
**RECORRENTE : SINDICATO DOS TRABALHADORES NAS INDÚSTRIAS DE ALIMENTAÇÃO DE MOJI MIRIM**  
**ADVOGADOS : ELIANE TREVISANI MOREIRA E OUTROS**  
**RECORRIDA : CORN PRODUCTS BRASIL INGREDIENTES INDUSTRIAIS LTDA. ( ATUAL DENOMINAÇÃO DE REFINAÇÕES DE MILHO BRASIL LTDA.)**  
**ADVOGADOS : ASSAD LUIZ THOMÉ E OUTROS**

JORNADA - TURNO DE REVEZAMENTO. A jornada de seis horas prevista, como regra, no inciso XIV do artigo 7º da Constituição Federal cede àquela pactuada em contrato coletivo, tal como excepcionado na parte final do preceito.

Fonte: Resultado obtido ao utilizar a Pesquisa de Jurisprudência no portal do STF.

Disponível em: <<http://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=AC&docID=246042>>

Uma das seções do acórdão, nomeada como "Ementa", se mostra essencial para

o desenvolvimento desse projeto. Para o acórdão utilizado como exemplo, cujo código é RE 215411, apenas a revisão dessa seção específica indicaria o ramo ao qual o acórdão pertence. Alguns termos na ementa do documento se destacam, como: jornada, turno de revezamento e contrato coletivo. Por meio da ocorrência dessas expressões, o acórdão poderia ser classificado como parte do direito do trabalho. Dessa forma, a intenção seria possuir um mecanismo que pudesse classificar os acórdãos por meio das identificações dessas palavras-chave.

Certas vezes, o reconhecimento do ramo do direito pode ser mais sutil, ou seja, apenas o conteúdo da ementa não seria suficiente para realizar uma análise como essa. Nesses casos, a classificação dependeria da revisão de outras seções da publicação: voto do relator, relatório do acórdão ou conclusão. Em situações excepcionais, um conhecimento teórico, de uma pessoa da área do direito, pode ser exigido. Para esse projeto, os acórdãos serão categorizados apenas pela análise da Ementa do acórdão.

## 2.3 Projetos anteriores

Anteriormente, o M.e Rafael Brito de Oliveira, enquanto aluno do Instituto de Matemática e Estatística da Universidade de São Paulo, se dedicou a trabalhar com a mesma base de dados contendo os acórdãos do STF. Entretanto, seu desafio consistiu em prover uma ferramenta de consulta que trouxesse uma abordagem mais intuitiva no momento de buscar acórdãos.

O objetivo de Rafael era suprir a carência de recursos que a Pesquisa de Jurisprudência oferece, já que nela, para realizar uma busca mais detalhada de um acórdão, é preciso preencher diversos campos em um formulário. Essa necessidade de fornecer diversas informações pode ser pouco intuitiva para o usuário, levando a uma experiência pouco satisfatória.

Nesse sentido, Rafael construiu uma interface *web* para pesquisas de acórdãos. Na interface a procura pode ser feita através de linguagem natural, ou seja, por meio das construções de sentenças como: "Quais são os ministros mais desafiadores?". Para a recuperação de acórdãos com base na pesquisa, Rafael projetou um mecanismo de consulta que utiliza uma ontologia. Com isso, o sistema é capaz de extrair informações relevantes dos acórdãos que possam ser essenciais para a geração dos resultados, trazendo uma representação do conhecimento contido no acórdão em um formato mais legível para o mecanismo de busca (OLIVEIRA, 2017).

Dessa forma, Rafael também atuará como orientador deste projeto, ao lado da Profa. Dra. Renata Wasserman, uma vez que traz consigo todos os conhecimentos obtidos durante o desenvolvimento do seu projeto de Mestrado. Ademais, o desenvolvimento do classificador de acórdãos por ramos do direito também pode, futuramente, complementar o mecanismo de busca criado por Rafael, de forma a oferecer previsões dos ramos do direito

Figura 2.5: Projeto desenvolvido durante o Mestrado do Rafael.

## Pesquisa de Jurisprudência

Quem são os ministros desafiadores?  Q

+ Nome	Quantidade				
- TEORI ZAVASCKI	1				
<table border="1"><thead><tr><th>Ministro Desafiador</th><th>Acórdão</th></tr></thead><tbody><tr><td>TEORI ZAVASCKI</td><td><a href="#">AGR RCL 20905</a></td></tr></tbody></table>	Ministro Desafiador	Acórdão	TEORI ZAVASCKI	<a href="#">AGR RCL 20905</a>	
Ministro Desafiador	Acórdão				
TEORI ZAVASCKI	<a href="#">AGR RCL 20905</a>				
+ MAURÍCIO CORRÊA	41				
+ NELSON JOBIM	49				
+ MOREIRA ALVES	3				
+ OCTAVIO GALLOTTI	1				
+ MARCO AURÉLIO	7				
+ SEPÚLVEDA PERTENCE	11				
+ CELSO DE MELLO	12				
+ ILMAR GALVÃO	4				
+ CARLOS VELLOSO	2				
+ NÉRI DA SILVEIRA	2				
+ ELLEN GRACIE	13				

Fonte: Sistema do Rafael em execução. Imagem elaborada pelos autores.

com base nos acórdãos apresentados como resultados de cada pesquisa.

## 3 Metodologia

### 3.1 Parceria com a Faculdade de Direito da Universidade de São Paulo

O projeto será desenvolvido em conjunto com a Faculdade de Direito da Universidade de São Paulo, especificamente com o professor Juliano Souza de Albuquerque Maranhão e seus alunos. Por meio dessa parceria, já foi possível extrair a informação que em muitos dos acórdãos, apenas a análise da Ementa poderia resultar na classificação do mesmo em um determinado ramo do direito.

Alguns alunos do professor Juliano irão trabalhar para categorizar manualmente um certo conjunto de acórdãos ao seu respectivo ramo do direito. Dessa forma, algumas perguntas serão respondidas por esse alunos como:

- Qual o código do acórdão?
- Em qual ramo do direito a questão de mérito do caso se enquadra?
- Para a identificação do ramo, bastou apenas examinar a ementa?
- Há indicação explícita na ementa do ramo do direito no qual o caso se enquadra?
- Quais termos ou expressões na ementa possibilitaram a identificação do ramo?

Através da coleta dessas respostas, um mecanismo pode ser treinado e capacitado para classificar acórdãos que ainda não foram rotulados manualmente.

### 3.2 Obtenção e adequação dos dados

A obtenção das informações que serão utilizadas na construção dos conjuntos de dados para treinamento, teste e validação do classificador será feita de duas formas. As classificações manuais de quais acórdãos pertencem a quais ramos do Direito serão extraídas do formulário que será preenchido pelos alunos do Professor Juliano enquanto os textos completos da ementa de cada acórdão serão adquiridos com o auxílio de uma ferramenta<sup>1</sup>

disponibilizada como parte do desenvolvimento do projeto do aluno de mestrado Jackson Souza.

Após a obtenção dos dados, será feita a construção do *dataset* que será usado no classificador, relacionando os textos das ementas dos acórdãos com as suas devidas classificações dentre os ramos do direito.

### 3.3 Elaboração do classificador

O mecanismo de classificação automática dos acórdãos será desenvolvido a partir do modelo *BERT*<sup>2</sup> (DEVLIN *et al.*, 2018). O *BERT* é um algoritmo de aprendizagem profunda projetado e disponibilizado pela Google para atividades de processamento de linguagem natural.

Para este trabalho, em específico, o *BERTimbau* (SOUZA; NOGUEIRA; LOTUFO, 2020) se mostrou uma opção poderosa, uma vez que é um modelo *BERT* pré-treinado e adaptado para a variante brasileira da língua portuguesa. Assim, ocorrerá um refinamento do modelo, de forma que regras de classificação possam ser incluídas visando melhorar a distinção entre os ramos do direito identificados.

O projeto será desenvolvido na linguagem *Python* juntamente das bibliotecas *Transformers* e *PyTorch* para importação e manuseio do modelo *BERT*.

Devido à necessidade do uso de uma GPU para poder executar o modelo *BERT*, todo o processo de refinamento e treinamento do modelo do classificador será executado em um notebook no *Google Colab* que disponibiliza serviços de GPU sem custos.

---

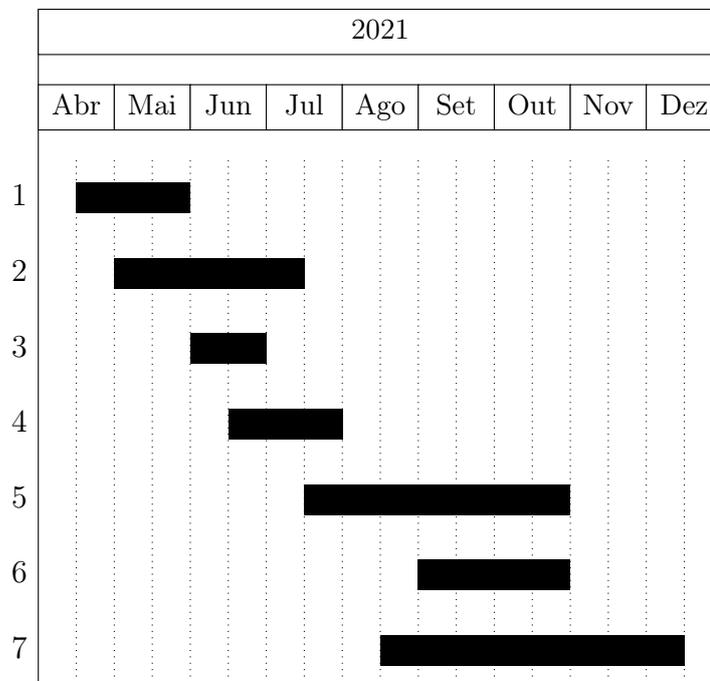
<sup>1</sup>O código referenciado está disponível em <<https://github.com/jacksonjos/analise-juridica>>.

<sup>2</sup>A sigla *BERT* é uma abreviação para *Bidirectional Encoder Representation for Transformers*.

## 4 Cronograma

1. Estudo sobre os acórdãos e os ramos do direito;
2. Estudo sobre *Machine Learning* e o modelo *BERT*;
3. Obtenção dos dados;
4. Adequação dos dados e construção dos *datasets*;
5. Implementação do classificador;
6. Treinamento do classificador e relatórios de performance; e
7. Monografia.

Figura 4.1: Gráfico de Gantt com o planejamento das atividades do trabalho.



## Referências Bibliográficas

BETIOLI, A. B. **Introdução ao direito**. [S.l.]: Saraiva, 2015.

CONTRUCCI, G. **O que é evolução do direito**. [S.l.]: Brasiliense, 2010.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.

OLIVEIRA, R. B. D. **Utilização de Ontologias para Busca em Base de Dados de Acórdãos do STF**. 58 f. Dissertação (Mestrado) — Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: pretrained BERT models for Brazilian Portuguese. In: **9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (2020)**. [S.l.: s.n.], 2020.

VENOSA, S. D. S. **Introdução ao estudo do direito**. [S.l.]: Atlas, 2019.

WOLKMER, A. C. **Fundamentos de história do direito**. [S.l.]: Del Rey, 2008.