

IAGO FARIAS LORA

**DIREITO COMPUTACIONAL:
Artefatos Tecnológicos e a Hegemonia da Prática Jurídica**

DISSERTAÇÃO DE MESTRADO

**BRASÍLIA – DF
2021**

IAGO FARIAS LORA

**DIREITO COMPUTACIONAL:
Artefatos Tecnológicos e a Hegemonia da Prática Jurídica**

Dissertação de Mestrado apresentada como requisito para obtenção do título de Mestre em Propriedade Intelectual e Transferência de Tecnologia para Inovação, do Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para Inovação (PROFNIT) - ponto focal Universidade de Brasília.

Orientadora
Prof. Dra. Leila Maria Da Juda Bijos

BRASÍLIA – DF
2021

DIREITO COMPUTACIONAL: Artefatos Tecnológicos e a Hegemonia da Prática Jurídica

MEMBROS DA BANCA

Prof. Dr. Bernardo Pereira Cabral

Prof. Dr. Gustavo Ferreira Ribeiro

Prof. Dra. Leila Maria Da Juda Bijos
Orientadora

Prof. Dr. Paulo Gustavo Barboni Dantas Nascimento
Presidente

RESUMO

A pergunta que norteia a presente pesquisa é: de que maneira o operador do Direito poderia, sendo o mais eficiente possível, obter domínio total do conteúdo jurídico e fomentar o conhecimento que detém sobre como se dá a aplicação da norma ao caso? Um caminho plausível para a resposta destas perguntas iniciais, encontra amparo na tecnologia que une racionalidade humana – capaz de promover soluções ainda que os problemas sejam subjetivos, e a ferramenta computacional – capaz de realizar análises objetivas em escala e velocidade sobre-humanas. Com isso, estabeleceu-se o objetivo de apresentar as relações e correlações entre Direito e Tecnologia utilizando-se de 8 eixos que se interrelacionam. De início, é realizada a **(i)** contextualização temática, em seguida, são apresentados e delineados conceitos técnicos de Direito, Computação e outras áreas correlacionadas com a pesquisa: **(ii)** Dado Informação e Conhecimento; **(iii)** Normalização de Dados; **(iv)** Ambiente Virtual; **(v)** Big Data Jurídico; **(vi)** Knowledge Discovery in Databases; **(vii)** Transformação de Dados em Conhecimento; **(viii)** Machine Learning e Inteligência Artificial – Exposição de Casos. Em seguida buscou-se identificar tecnologias nacionais que são ofertadas ao setor jurídico brasileiro, bem como, necessidades apresentadas pelo setor; obtendo, com isso, parâmetros comparativos e de benchmarking para a proposta de desenvolvimento de tecnologias como produto tecnológico.

Palavras-chave: Jurimetria; Informática Jurídica; Predileção Jurídica; Direito e Tecnologia; Tecnologia aplicada ao Direito; Propriedade Industrial.

ABSTRACT

The question that guides the present research is: how could Law operators, being as efficient as possible, obtain total mastery of the legal content and foster the knowledge they have on applying the rule to the case? A plausible way to answer these initial questions is to find support in technology that unites human rationality - capable of promoting solutions even though problems are subjective and the computational tool - performing objective analyses at a superhuman scale and speed. Our objective was to present the relations and correlations between Law and Technology using eight interrelated axes. Initially, (i) thematic contextualization is carried out, then technical concepts of Law, Computing, and other areas related to the research are presented and outlined: (ii) Information and Knowledge Data; (iii) Data Standardization; (iv) Cybernetic Environment; (v) Legal Big Data; (vi) Knowledge Discovery in Databases; (vii) Transformation of Data into Knowledge; (viii) Machine Learning and Artificial Intelligence - Case Exposure. Then, we sought to identify national technologies that are offered to the Brazilian legal sector and needs presented by the industry, thus obtaining comparative and benchmarking parameters for the proposal for the development of technologies as a technological product.

Keywords: Jurimetrics; Legal Informatics; Legal Predilection; Law and Technology; Technology applied to Law; Industrial Property.

LISTA DE SIGLAS

PIB – Produto Interno Bruto

LAI – Lei de Acesso à Informação

PJE – Processo Judicial Eletrônico

IOT – *Internet of Things*

ED – Embargos Declaratórios

KDD – *Knowledge Discovery in Databases*

CDD – Classificação Decimal de Dewey

CDU – Classificação Decimal Universal

IA – Inteligência Artificial

STF – Supremo Tribunal Federal

STJ – Superior Tribunal de Justiça

INPI – Instituto Nacional de Propriedade Industrial

TJDFT – Tribunal de Justiça do Distrito Federal e Entornos

SAAS – *Software As A Service*

OCR – *Optical Character Recognition*

SUMÁRIO

1 INTRODUÇÃO.....	7
2 CONTEXTUALIZAÇÃO TEMÁTICA	14
3 REFERENCIAL TEÓRICO	22
3.1 DADO, INFORMAÇÃO E CONHECIMENTO	22
3.1.1 DADO	22
3.1.2 INFORMAÇÃO.....	22
3.1.3 CONHECIMENTO	23
3.2 AMBIENTE VIRTUAL	24
3.2.1 PERIGOS E OPORTUNIDADES	24
3.2.2 ALCANCE.....	28
3.2.3 DIMENSÃO.....	29
3.3 BIG DATA JURÍDICO	31
3.3.1 FONTES DO DIREITO.....	31
3.3.2 FONTES DE CONTEÚDO JURÍDICO	34
3.3.3 INDEXAÇÃO DE DADOS DO CONTEÚDO JURÍDICO.....	35
3.4 KNOWLEDGE DISCOVERY IN DATABASES – KDD	39
3.5 NORMALIZAÇÃO.....	42
3.6 TRANSFORMAÇÃO DE DADOS EM CONHECIMENTO	43
3.7 MACHINE LEARNING E INTELIGÊNCIA ARTIFICIAL	45
3.7.1 CASO 1	46
3.7.2 CASO 2	47
3.7.3 CASO 3 - TEOREMA DO JÚRI DE CONDORCET	49
4 MÉTODO	51

5 RESULTADOS	53
5.1 TECNOLOGIAS GOVERNAMENTAIS	54
5.2 TECNOLOGIAS COMERCIAIS.....	61
5.3 TECNOLOGIAS DESENVOLVIDAS NO DECORRER DA PESQUISA	62
5.3.1 AUTOMAÇÃO DE PEÇAS PROCESSUAIS	62
5.3.2 ANÁLISE PREDITIVA BÁSICA.....	63
5.3.3 IDENTIFICAÇÃO DE TEMAS, PARTES, CAUSA DE PEDIR, VALORES, RESULTADO DA DECISÃO E OUTROS ELEMENTOS	65
5.3.4 FERRAMENTA PARA IDENTIFICAÇÃO DE SIMILARIDADES DE DECISÕES	75
5.3.5 FERRAMENTA DE VERIFICAÇÃO DE ÁREAS E MINISTROS.....	76
5.3.6 INTELIGÊNCIA ARTIFICIAL PARA PRODUÇÃO DE TEXTOS.....	79
6 CONCLUSÕES.....	82
7 REFERÊNCIAS	84

1 INTRODUÇÃO

Partindo das premissas **(i)** que o “papel do advogado” – profissional que, por sua formação acadêmica, e, a partir do exercício diário de seu ofício – resume-se em conhecer e operar o Direito; de que **(ii)** o Direito é composto, dentre outros, por princípios, normas e a hermenêutica que leva a aplicação da norma ao caso concreto; e **(iii)** que é pouco provável que um advogado seja capaz de mencionar o conteúdo de todas as leis existentes; a conclusão lógica se resume em que, ainda que determinado advogado seja o profissional mais capacitado em termos de conhecimentos jurídicos, jamais poderá alcançar, sozinho, a máxima eficiência. Podendo tão somente, alcançar o que os limites humanos e fisiológicos lhe permitirem. **Com isso, pergunta que se fundamenta é:** de que maneira o operador do Direito poderia, sendo o mais eficiente possível, obter domínio total do conteúdo jurídico e fomentar o conhecimento que detém sobre como se dá a aplicação da norma ao caso?

Um caminho plausível para a resposta destas perguntas iniciais, encontra amparo na tecnologia que une racionalidade humana – capaz de promover soluções ainda que os problemas sejam subjetivos, e a ferramenta computacional – capaz de performar análises objetivas em escala e velocidade sobre-humanas. Hipoteticamente, o resultado de uma solução para o problema percebido, produziria um *advogado tecnologicamente empoderado*. Então pergunta-se: o que seria um “*advogado tecnologicamente empoderado*”? Depreende-se de tal conceito, a visão de um profissional do Direito munido de conhecimentos tecnológicos que o permitam escalonar a prática jurídica; tornar o exercício do Direito uma *atividade escalável*.

Há uma relação [~~de longa data~~¹] entre a computação moderna, a teoria do direito e sistemas de computador aplicados à prática jurídica; [~~que pode ser seguida por meio da~~] [incluindo em tais relações, temas como]²: [a] modelagem de problemas jurídicos em termos de Teoria dos Jogos, [b] a criação de sistemas jurídicos especializados, subsidiados por IA, [c] nas ideias do ciberespaço como um domínio jurídico distinto e [d] no enquadramento legal da guerra cibernética.³

Embora tenha havido uma crítica substancial⁴ sobre a aplicação da lei no uso da computação – Direitos Autorais e Propriedade Industrial envolvidos na obras/criações tecnológicas – incluindo aquelas geradas pelas máquinas em si⁵, pouca análise foi feita a respeito de como o Direito e a Computação serão modificados pela relação integrada e simbiótica entre os sistemas jurídico e computacional. Ou seja, com o termo *into one another*⁶, os autores também sugerem uma integração bidirecional/mútua – ou seja, tanto a Tecnologia interfere no Direito, quanto o Direito interfere na Tecnologia. Tal relação – Computação aplicada ao Direito, surge com denominações variadas, tendo de uma para outra nuances quanto ao escopo de estudo.

¹ Não se sabe qual o parâmetro utilizado para atribuição de “*long-standing relationship*” – ou, relação de longa data; nesse ponto, discordando dos autores. Apesar da possibilidade de ir a tempos anteriores, toma-se por exemplo o período clássico do Direito Romano, período que concorre paralelamente com o marco inicial do calendário ocidental (ano 0001); período histórico-jurídico marcado pela consolidação de conceitos até hoje utilizados (propriedade, posse, contrato, direito das coisas e ações judiciais). Se até o surgimento da computação decorreram quase dois mil anos, há de se considerar discutível uma relação de longa data com a computação moderna.

² Adaptado pelos autores.

³ *There is a long-standing relationship between the development of modern computing and legal theory and the application of computer systems to legal practice that can be followed through the modelling of legal problems in terms of Game Theory, the creation of AI based legal expert systems, in ideas of cyberspace as a distinct legal realm and the legal framing of cyber warfare.* KATZ, Daniel Martin; BOMMARITO II, Michael J. **Measuring the Complexity of the Law**: The United States Code. [Draft v1.01] 2013. p. 3.

⁴ As críticas substanciais sobre a aplicação da lei e uso da computação pode ocorrer de diferentes formas, por exemplo: os direitos autorais e a relação com o código fonte de softwares, a regulação sobre a coleta e uso de dados, a responsabilidade de danos causados por Inteligência Artificial – I.A., bem como, o direito de autoria das produções artísticas geradas por computadores – vez que o princípio norteador da I.A. é a autonomia da máquina para a tomada de decisão. Tais produções artísticas podem ter aparências peculiares e exóticas.

⁵ ex.: *Google Deep Dream*.

⁶ *Whilst there has been substantial critical writing on the application of law to the use of computing, as in issues such as copyright and IP, there has been less analysis of how law and computing may be changed by the integration of legal and computational systems into one another.* KATZ, Daniel Martin; BOMMARITO II, Michael J. **Measuring the Complexity of the Law**: The United States Code. [Draft v1.01] 2013. p. 3.

De forma geral, as classificações e diferenças entre as denominações dadas a área que estuda a Computação aplicada ao Direito não se assentaram por completo. *Legal Intelligence / Legal Analytics / Computational Law / Applied Jurimetrics* são atribuídos à área, especificidades e objetos de pesquisa variam de autor para autor.

Considera-se que, todos os termos supramencionados representam, *lato sensu*, a computação e análise de dados jurídicos. Nota-se que computação tem por significado o ato ou efeito de computar – calcular; apesar de atualmente ter, presumidamente, uma ligação à máquina (computador), não tem existência subordinada à segunda.

Cabe a Leibniz o mérito de associar *(i)* a probabilidade de um estado epistêmico de dúvida acerca de um evento, à *(ii)* natureza probabilística de um evento futuro no Direito.⁷ Ou seja, a partir da análise de resultados e frequência de eventos passados, propor a **predileção** de resultados.

Também são nomes de relevância para o campo da Jurimetria Jacob e Nicolas Bernoulli. Jacob Bernoulli, matemático, é autor da obra *Ars Conjectandi* que fundamenta a teoria da probabilidade; já Nicolas Bernoulli (sobrinho de Jacob), estudou Direito, defendendo sua tese de Doutorado aplicações da *Ars Conjectandi* no campo jurídico – *De Usu Artis Conjectandi in Juris*.⁸

Nesse seguimento, considera-se outro expoente Lee Loevinger, Advogado – Direito Concorrencial. Loevinger utilizou, pela primeira vez, o termo Jurimetria – definindo-o como o uso de métodos quantitativos no Direito.

⁷ LEIBNIZ, G. W. [1655](2002) apud. NUNES, Marcelo Guedes. **Jurimetria**: Como a Estatística pode reinventar o Direito. São Paulo: Revista dos Tribunais, 2016.

⁸ NUNES, Marcelo Guedes. **Jurimetria**: Como a Estatística pode reinventar o Direito. São Paulo: Revista dos Tribunais, 2016.

Genereseth, cientista da computação, conhecido por seu trabalho em lógica computacional e aplicações desse trabalho em gestão empresarial e comércio eletrônico com áreas de pesquisa e projetos que incluem, dentre outros, **(i)** a representação do conhecimento, raciocínio automatizado e ação racional, **(ii)** planilhas lógicas, dados e integração de serviços na World Wide Web e **(iii) computação jurídica**,⁹ passa a atribuir a noção da informática ao contexto do estudo.

O Direito Computacional é o ramo da informática jurídica que se preocupa com a mecanização da análise jurídica (seja feita por humanos ou máquinas). Enfatiza restrições comportamentais explícitas e evita regras de conduta implícitas. É importante ressaltar que há um compromisso com um nível de rigor na especificação de leis que é suficiente para apoiar o processamento inteiramente mecânico. Embora a ideia de análise jurídica mecanizada não seja nova, suas perspectivas são melhores do que nunca devido aos recentes desenvolvimentos tecnológicos - incluindo o progresso na Lógica Computacional, o crescimento da Internet e a proliferação de sistemas autônomos (como carros autônomos e robôs). A tecnologia jurídica baseada no Direito Computacional tem o potencial de mudar radicalmente a profissão jurídica, melhorando a qualidade e a eficiência dos serviços jurídicos e possivelmente perturbando a forma como os escritórios de advocacia fazem negócios. De forma mais ampla, a tecnologia tem o potencial de trazer compreensão jurídica e ferramentas jurídicas para todos na sociedade, não apenas para profissionais do direito, aumentando assim o acesso à justiça e melhorando o sistema jurídico como um todo¹⁰

Ou seja, Direito Computacional é o ramo da informática jurídica preocupado com a mecanização de análise jurídica, quer essa análise seja feita por seres

⁹ STANFORD [Stanford Profiles]. **Michael Genereseth**. Disponível em: <https://profiles.stanford.edu/michael-genesereth?tab=bio> Acesso em: 17 março 2021.

¹⁰ *Computational Law is that branch of legal informatics concerned with the mechanization of legal analysis (whether done by humans or machines). It emphasizes explicit behavioral constraints and eschews implicit rules of conduct. Importantly, there is a commitment to a level of rigor in specifying laws that is sufficient to support entirely mechanical processing. While the idea of mechanized legal analysis is not new, its prospects are better than ever due to recent technological developments - including progress in Computational Logic, the growth of the Internet, and the proliferation of autonomous systems (such as self-driving cars and robots). Legal technology based on Computational Law has the potential to dramatically change the legal profession, improving the quality and efficiency of legal services and possibly disrupting the way law firms do business. More broadly, the technology has the potential to bring legal understanding and legal tools to everyone in society, not just legal professionals, thus enhancing access to justice and improving the legal system as a whole.* GENESERETH, Michael. **Computational Law: The Cop in the Backseat**. [White Paper, **CodeX**—The Stanford Center for Legal Informatics], 2015. p. 1.

humanos ou máquinas. “De uma perspectiva pragmática, o *Direito Computacional* é a base para sistemas de computador capazes de fazer cálculos [...] úteis [para a tomada de decisão], como verificação de conformidade, planejamento legal, análise regulatória e assim por diante.”¹¹.

Para cumprimento dos requisitos do Trabalho de Conclusão de Curso do Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação – PROFNIT, o presente trabalho delimita-se duas entregas que se complementam: *(i) Dissertação* e *(ii) Produto Tecnológico*. Na apresentação dissertativa dar-se-á direcionamento ao estudo que trata da Tecnologia Aplicada ao Direito através de um ensaio teórico. O produto tecnológico, no caso, os produtos tecnológicos, tratam-se de ferramentas desenvolvidas no decorrer do trabalho.

Vista do que foi exposto na presente introdução, estabelecemos como **objetivo** desta pesquisa: apresentar as relações e correlações entre *Direito* e *Tecnologia*, habilitando o leitor a compreender tanto o cenário estudado quanto o desenvolvimento resultados da pesquisa. Têm-se como objetivos específicos, a exposição bibliográfica que permita a compreensão das peculiaridades da relação entre as áreas; o mapeamento de ferramentas de tecnologia jurídica existentes; adicionalmente, a exposição de procedimentos e artifícios computacionais aplicáveis na rotina de trabalho de um advogado. No Produto Tecnológico principal (base de pesquisa e relatórios)¹², altera-se o foco para o Direito aplicado à Tecnologia, ou seja, a tecnologia apresentada, é uma prototipação de ferramenta orientada à análise de temas e desempenho prático e tem como o modelo base o ramo do Direito de Propriedade Intelectual. Para fins de esclarecimento, a *Propriedade Intelectual* é o

¹¹ GENESERETH, Michael. **Computational Law: The Cop in the Backseat**. [White Paper, CodeX—The Stanford Center for Legal Informatics], 2015. pp. 2-3.

¹² Itens 5.3.3; 5.3.4 e 5.3.5.

campo do Direito que trata de obras intelectuais, criações, invenções, desenvolvimento tecnológico, dentre outros.

Localizar aplicações de Inteligência Jurídica já desenvolvidas, sejam elas de alta maturidade tecnológica ou não, se trata de uma etapa utilizada para obter parâmetros comparativos e benchmarking da tecnologia proposta.

Apresentadas as perguntas e hipóteses que dão norte à pesquisa, a exposição teórica da presente dissertação estrutura-se em 8 eixos que se interrelacionam. De início, é realizada a **(i)** contextualização temática, em seguida, são apresentados e delineados¹³ conceitos técnicos de Direito, Computação e outras áreas correlacionadas com a pesquisa: **(ii)** Dado Informação e Conhecimento; **(iii)** Normalização de Dados; **(iv)** Ambiente Virtual; **(v)** Big Data Jurídico; **(vi)** *Knowledge Discovery in Databases*; **(vii)** Transformação de Dados em Conhecimento; **(viii)** *Machine Learning e Inteligência Artificial* – Exposição de Casos.

A partir do Método descrito, buscou-se identificar tecnologias nacionais que são ofertadas ao setor jurídico brasileiro – procedimento a viabilizar um estudo prospectivo e mapeamento de tecnologias jurídicas; em continuidade, são apresentados exemplos de tecnologias já desenvolvidas e do potencial disruptivo destas tecnologias¹⁴, dos diferentes métodos utilizados e resultados obtidos. Por fim,

¹³ A título exemplificativo observa-se a palavra “processo” – cujo significados atribuídos pelo Direito, pela Engenharia Computacional e Administração/Engenharia de Produção são distintos e, nesse sentido, dificultam uma comunicação eficiente. A palavra “processo” consiste no mais superficial dos falso-cognatos que se depreendem da gama de vocabulários técnicos necessários para a devida apresentação do tema proposto. Além de palavras com conotações díspares, meras nuances linguísticas são suficientes para ensejar uma confusão contextual; por exemplo: (i) direito da tecnologia/tecnologia do direito, (ii) direito da tecnologia/direito computacional, (iii) direito computacional/direito digital, (iv) direito digital/digitalização do direito.

¹⁴ O exemplo de tecnologias ofertadas, foi frustrado por uma série de fatores, dentre estes: práticas de “marketing e captação”, pela insuficiência dos sistemas ofertados, por contratos de confidencialidade ou outros fatores – o que levaria à conclusão enviesada de baixo potencial da utilização de tecnologias no Direito. Sabe-se da existência de ferramentas tais quais o *Lexis Nexis*, *Ravel Law*, dentre outros – fato que comprovam a existência e que apresentam funcionalidades de alta potencialidade. Como a pesquisa é cerceada ao contexto brasileiro, foram então demonstrados procedimentos desenvolvidos pelos autores como forma demonstrativa.

perspectivas e conclusões dos autores em face da pesquisa; seguido então do projeto desenvolvido.

2 CONTEXTUALIZAÇÃO TEMÁTICA

Desde o início da linguagem – em especial, da linguagem escrita, foi possível ao ser humano repassar e armazenar informações de maneira mais eficiente. Considera-se que o raciocínio lógico e o reconhecimento de padrões são os pilares do convívio humano em sociedade.

A navegação astronômica, avanços medicinais, e científicos em geral foram viabilizados pela capacidade humana de reconhecer arquétipos e, utilizando-se da linguagem escrita, dar forma e perpetuidade à transferência de informações. A linguagem escrita, nada mais é do que uma sequência de símbolos/caracteres que padronizam uma mensagem. Necessidades e avanços comerciais-tecnológicos, aliados ao desenvolvimento da *Astronomia* e instrumentos de viagem, marcaram a era no desenvolvimento das ciências batizada de *Revolução Científica*.¹⁵

A máxima que costuma ser atribuída como de autoria tanto de Francis Bacon¹⁶ quanto de Thomas Hobbes¹⁷ “*nam et ipsa scientia potestas est*”¹⁸ é um forte reflexo de como o saber sempre foi um ativo social, sendo sua importância ainda maior na “*Era do Big Data*”, em que as possibilidades de correlação e descoberta são infinitas. O que se retrata nesse contexto é a exposição de: sendo o dado bruto a fonte de informação, por consequência, de conhecimento, o Estado, por meio das políticas de acesso à informação proporcionou instrumentos que reforçam a valorização da ordem econômica nacional.

Sob a perspectiva econômica não seria errado afirmar que o poder da tecnologia e da internet são capazes de superar soberanias estatais. Tem-se como

¹⁵ MATTAR, João. **Metodologia Científica na Era Digital**. São Paulo: Saraiva, 2017. p. 15.

¹⁶ BACON, Francis. **Meditationes sacrae**. [1603/4]. Stuttgart-Bad Cannstatt: Frommann-holzboog, 1998. [*in latin*]. 1998, Frommann-Holzboog in Latin.

¹⁷ HOBBS, Thomas. **Leviatã: Matéria, forma e poder de um Estado eclesiástico e civil**. [1651] (Tradução de João Paulo Monteiro e Maria Beatriz Nizza da Silva). 3. ed. São Paulo: AbrilCultural, 1983.

¹⁸ Que pode ser compreendida, entre outras formas, como “conhecimento é poder”.

exemplos, as companhias *internet-based* que detêm faturamentos maiores do que o Produto Interno Bruto – PIB de alguns países. A título de exemplo, pode ser citada a companhia chinesa *Tencent Holdings*, que em 2016 obteve receita superior ao faturamento do comércio de soja Brasileiro – comparados os dados de Jon Russel¹⁹ com os disponibilizados pelo *Observatory of Economic Complexity*²⁰.

A recente afirmação de que a “[t]erra era a matéria prima da *Era Agrícola*], o ferro era a matéria prima da *Era Industrial*], e que d]ado é a matéria prima da *Era da Informação*”²¹, em termos comparativos, fazem crer que as políticas de Acesso à Informação, subsidiaram insumos para a produção de capital – e, sendo tais insumos, em sua grande maioria, gratuitos, utilizá-los e manejá-los de forma correta são quase suficientes para a aquisição de renda.

Com o advento dos bancos de dados, por hora considerando em específico os bancos de dados governamentais, e, reforçado pela política de Dados Abertos e pela Lei 12.527/2011²² – Lei de Acesso à Informação – LAI, foram oportunizados avanços na Pesquisa e Desenvolvimento – P&D, principalmente para as ciências sociais. Talvez ousado, mas não condenável, seria dizer que os dados são ativos de características peculiares – são, muitas vezes, fornecidos gratuitamente e em larga escala; e desses, incontáveis as possibilidades de informações que podem ser extraídas e transformadas em conhecimento.

¹⁹ RUSSEL, Jon. **Tencent posts \$21.9 billion in annual revenue, its highest growth since 2012**. Disponível em: <<https://techcrunch.com/2017/03/22/tencent-2016-revenue/>>. Acesso em: 26 out. 2020.

²⁰ OBSERVATORY OF ECONOMIC COMPLEXITY – OEC. **Brazil**. Disponível em: <https://atlas.media.mit.edu/en/profile/country/bra/#Trade_Balance>. Acesso em: 26 out. 2018.

²¹ ROSS, Alec. **The Case for Connectivity**: The upward economic mobility needed to sustainably address basic needs in the 21st century requires an Internet connection. Any economy that remains offline will grow less competitive and more isolated as the globe does its business online. HuffPost, online, abr. 2014. Disponível em: https://www.huffingtonpost.com/alec-ross/the-case-for-connectivity_b_4839486.html. Acesso em: 31 jan. 2019.

²² BRASIL. **Lei Nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

Visando o desempenho efetivo do Direito, urge que juristas alcancem um nível de aprofundamento da informação que os permita saber como determinados assuntos são julgados. Faz-se necessário imbuir-se da norma aplicada (ou aplicável), compreender as bases principiológicas, contextos, bem como, as variadas compreensões de outros profissionais acerca do mesmo assunto - em especial, o posicionamento de Tribunais e magistrados. As diferentes compreensões atribuídas e tratativas dadas pelos diversos Tribunais são recursos bibliográficos que fornecem subsídios para que teses eficientes sejam criadas. Todas essas são fontes que permitem que seja procedimentalizada a extração de dados, viabilizando a análise e gerenciamento de informação.

Os advogados lidam com grandes volumes de dados de várias fontes e embora possam revisar facilmente dados não estruturados, como documentos, transcrições e gravações de áudio; somente alguns poucos são capazes de trabalhar dados estruturados com naturalidade - em geral àqueles que lidam rotineiramente com registros de transações financeiras, contabilidade, tributação e dados comumente divulgados em planilhas. Em sua vasta maioria o Direito nunca foi visto ou tratado como uma estrutura de dados.

A estruturação desses dados gera várias perspectivas que permitem revelar links entre casos, normas, decisões, recursos, partes, advogados, magistrados, dentre outros. Nesse sentido, a prática de *Legal Intelligence* permite ao advogado importar e analisar dados viabilizando o diagnóstico de relacionamentos, explorar hipóteses divergentes, descobrir conexões desconhecidas, trazer padrões ocultos à superfície, criar painéis e compartilhar *insights* com outros advogados, parceiros e clientes.

Retomando a importância da linguagem, verifica-se uma certa dificuldade, sob a ordem da linguagem técnica, da compreensão da “linguagem jurídica”, algo que hipoteticamente, aparta o setor tanto da sociedade “comum” quanto dos avanços sociais – no contexto da presente pesquisa, os avanços tecnológicos.

Michael Genesereth, ao citar Lee Loevinger, introduz o dilema tratado neste subcapítulo a partir do seguinte trecho: “é uma das maiores anomalias dos tempos modernos que a lei, que existe como um guia público de conduta, se tornou um mistério tão recôndito que é incompreensível para o público e dificilmente inteligível para seus próprios devotos.”²³

Por meio de uma perspicaz crítica, Loevinger menciona que:

Os advogados pensam e argumentam em termos tradicionais. Os penodiciais legais são cheios de artigos que parecem ter sido escritos com penas à luz bruxuleante de velas de sebo. E muitos ecos confusos de Tomás de Aquino continuam a reverberar pelos corredores vazios das cidadelas do conservadorismo legal.²⁴

Loevinger também destaca:

As regras que deveriam ser os guias para a ação dos homens que vivem em sociedade tornaram-se o culto secreto de um grupo de profissionais sacerdotais. O ritual místico desse culto é anunciado ao público, se é que o é, apenas em um jargão desconcertante.²⁵

É peremptório que a linguagem jurídica cria barreira de comunicação; tanto por termos técnicos quanto pelo conservadorismo, cujo traços são percebidos no uso de linguagem dotada de formalismo que não se adapta aos usos ou aos hábitos da

²³ *It is one of the greatest anomalies of modern times that the law, which exists as a public guide to conduct, has become such a recondite mystery that is incomprehensible to the public and scarcely intelligible to its own votaries.* LOEVINGER, Lee. Apud GENERESETH, Michael. **Computational Law: The Cop in the Backseat**. 2016.

²⁴ *Lawyers think and argue in traditional terms. Legal penodicals are mainly filled with articles that seem to have been written with quill pens by the flickering light of tallow candles. And fair confused echoes of Aquinas continue to reverberate through the hollow halls of the citadels of legal conservatism.* LOEVINGER, Lee. *Jurimetrics--The Next Step Forward*. **Minnesota Law Review**, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949. p. 466

²⁵ *The rules which are supposed to be the guides to action of men living in society have become the secret cult of a group of priestly professionals. The mystic ritual of this cult is announced to the public, if at all, only in a bewildering jargon.* LOEVINGER, Lee. *Jurimetrics--The Next Step Forward*. **Minnesota Law Review**, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949. p. 466

atualidade (ex: Exmo.; Vossa Excelência;)²⁶. Como exposto por Loevinger, “a verdade é que na maioria das questões jurídicas fundamentais, incluindo a questão relativamente simples de definir a própria lei, a profissão está em imersa em confusão e desacordo”.²⁷ Também coloca o autor que: “a questão importante para a compreensão não é a mecânica da lei, mas o problema humano que cabe ao negócio da lei resolver e os resultados reais que são alcançados”.²⁸

Nesse sentido, partimos do pressuposto de que o campo de estudo do Direito, nos referenciando mais especificamente aos profissionais da área, é, naturalmente, conservador. Tal conservadorismo seria, portanto, uma das barreiras que dificultam a integração entre profissional e o avanço tecnológico.

Edgar Bodenheimer (1948) descreve que o conservadorismo da área encontra amparo na influência que os profissionais mais proeminentes conseguem exercer sobre os demais – ou seja, de modo geral, profissionais com mais tempo de carreira e solidificados no mercado jurídico conseguem criar restrições de “poder” para que o *status quo* seja mantido²⁹.

Acerca da inércia tecnológica, constata-se que a cada Revolução Industrial, foram impactados os modelos de trabalhos existentes, profissões deixaram de existir, profissões surgiram – um ciclo de alterações e adaptação profissional. O setor Jurídico

²⁶ Inclusive, existem linhas de pesquisa que se debruçam sobre a Complexidade da Linguagem Jurídica; partindo desde a dogmática jurídica até a proposição de métricas para determinar a complexidade de normas.

²⁷ *The truth is that on most fundamental legal questions, including the relatively simple matter of defining law itself, the profession is in hopeless confusion and disagreement.* LOEVINGER, Lee. Jurimetrics--The Next Step Forward. **Minnesota Law Review**, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949. p. 456.

²⁸ *The important matter for understanding is not the mechanics of the law, but the human problem which it is the business of the law to solve, and the actual results which are achieved.* LOEVINGER, Lee. Jurimetrics--The Next Step Forward. **Minnesota Law Review**, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949. p. 456.

²⁹ BODEMEINHER, Edgar. The Inherent Conservatism of the Legal Profession. **Indiana Law Journal**, Indiana, v. 23, n. 3, p. 221-235, spring [set.-dez.], 1948.

manteve-se à parte das Revoluções Industriais, ou seja, nenhuma delas afetou diretamente a forma como o serviço é prestado.

O surgimento do computador foi particularmente importante para o Direito no que diz respeito à produção documental e pesquisa – são utilizadas as mesmas ferramentas e técnicas de 20 anos atrás (*Microsoft Word*® e um navegador de Internet – em geral *Internet Explorer*®³⁰ - atual *Microsoft Edge*® ou *Google Chrome*®³¹; uma versão desatualizada do *Mozilla Firefox*®³² é utilizada em background pelo sistema de assinaturas do Processo Judicial Eletrônico – PJE). Nesse sentido, percebe-se que a prática jurídica se conecta principalmente a 3 utensílios digitais: um editor de texto (*Word*), um browser para acesso ao Google e às Bases de Dados dos tribunais como ferramenta de pesquisa e acompanhamento processual e uma plataforma de e-mail. Todos os 3 usados de forma superficial.

Por outro lado, se a necessidade de conhecimento tecnológico é um pré-requisito para o desenvolvimento de tecnologias jurídicas, o conhecimento do Direito faz-se ainda mais importante. Trata-se de um evidente gargalo em muitas tecnologias jurídicas atualmente oferecidas, a falta de um Advogado ou profissional com formação jurídica orientando os trabalhos realizados e projetos desenvolvidos.

Qualquer pessoa com acesso à internet pode em poucos minutos encontrar uma infinidade de publicações sobre garantias constitucionais e liberdade de expressão e em seguida publicar em dezenas de websites. Tal ato não faz do autor um constitucionalista, tal qual não há necessidade de OAB para publicar online um texto que trata de Direito.

³⁰ MICROSOFT. **Microsoft Edge**. [software] Disponível em <https://microsoftedgewelcome.microsoft.com/>

³¹ GOOGLE. [Chrome] **The Browser Build by Google**. [software] Disponível em <https://www.google.com/chrome/index.html>

³² MOZILLA. **Firefox Browser**. [software] Disponível em: <https://www.mozilla.org/>

Seguindo uma linha de raciocínio semelhante, um programador pode facilmente criar um “bot” para coletar, mesclar e publicar conteúdo sobre os mais variados assuntos, tal fato atesta sua capacidade em programar, contudo não faz dele médico, advogado, químico, físico por mais que possa publicar sobre um ou todos os assuntos que desejar.

No contexto da ciência da informação, diferentemente de cientistas da computação é remota a chance de se deparar com juristas que conseguem pensar e observar tais dados em termos de linhas, colunas, texto bruto ou linguagem codificada. Não quer dizer que a codificação da linguagem não exista, assim como existem estruturas de códigos computacionais o Direito também possui sua “linguagem de programação”, e esta, desconhecida pelos cientistas da computação.

Exemplo disso é o fato de ser uma compreensão quase que instintiva para um jurista – que ao ver o símbolo “§”, assimilar que se trata de algum tipo de documento normativo, provavelmente é precedido de um “Art.”; que complementarmente, os incisos buscam enfatizar uma lista de características – um rol taxativo ou enumerativo, et. seq.

Em contrapartida, para os cientistas da computação isso seria uma “linguagem de programação” – ainda que incompatível com qualquer outra ferramenta que utilizem. Tal incompatibilidade resulta na dificuldade de obter resultados estruturados e passíveis de tratamento e análise computacional.

Henrique Alves Pinto destaca:

[...] deve-se destacar que a atividade jurisdicional não se restringe apenas ao processo decisório, estando nela incorporados outros elementos tão relevantes quanto ele. **Nesse sentido, uma das grandes dificuldades a ser enfrentada por esse processo de automação de textos legais para códigos tecnológicos é a necessidade de constante atualização, que acaba ficando sob responsabilidade de programadores ou outros profissionais de Tecnologia da Informação, os quais, na maioria das vezes, não**

têm muita familiaridade e expertise no trato não só das fontes legais do sistema jurídico, como também das questões éticas, sociais, políticas e filosóficas que as permeiam.³³ Grifo nosso.

³³ PINTO, Henrique Alves. A utilização da inteligência artificial no processo de tomada de decisões: Por uma necessária accountability. **Revista de Informação Legislativa – RIL**, Brasília a. 57 n. 225 p. 23-42 jan./mar. 2020.

3 REFERENCIAL TEÓRICO

3.1 Dado, Informação e Conhecimento

3.1.1 Dado

No presente trabalho o termo “dado” é utilizado para identificar as fontes de informação em seu estado bruto; um artigo de lei, instrução, circular ou outro normativo; uma sentença judicial, ou mesmo o conteúdo de uma nota de opinião. Cumpre ressaltar que a mera existência do dado não o torna “correto”, isto pois o texto contido pode ser um viés, uma falsa premissa, um normativo improfícuo, et. al.; se um determinado dispositivo da norma é refutado pelo judiciário, ou norma derrocada por natureza / cujos efeitos jurídicos provenientes da norma, se é que assim podem ser chamados, são socialmente nulos (desprezados e abandonados). Tampouco, há de se presumir que o dado seja errado – ainda que supérfluo ou refutado pela sociedade ou judiciário. Um dado é única e estritamente uma fonte de informação cuja existência possa ser comprovada.

No Direito, além de sistemas desconectados, os dados variam tanto em sua forma estruturada, ex.: planilhas e tabelas quanto na forma “não estruturada” como e-mails, documentos escritos e outros.

3.1.2 Informação

O trabalho intelectual que consiste na observação de uma série de dados, correlações entre diferentes dados, e, enseja em uma conclusão – certa ou errada, proveitosa ou não – dá-se o nome de “produção da informação” ou “pesquisa”, por consequência, o resultado/conclusão recebe o rótulo de “informação”. Nesse sentido, informação é o dado processado com algum motivo; a informação, somente pode ser considerada “real” se certos critérios forem alcançados: **(i)** deve ser comunicada ao recipiente em linguagem que possa ser compreendida; **(ii)** deve seguir um formato

normalizado³⁴; *(iii)* deve ser relevante para o alcance de um objetivo.³⁵ Fazendo-se valer das hipóteses supramencionadas para exemplificar o conceito de “dado” e “informação”, – se um determinado dispositivo da norma é refutado pelo judiciário – evidencia-se a existência de dois dados (textuais); um pertence à norma, o outro ao documento da sentença.

Como colocado por Mehrotra, os computadores aceitam *inputs* – que são traduzidos para uma sequência numérica de binários e então processados; do processamento há um *output* e essa sequência pode ocorrer indefinidamente – ou seja, *outputs* podem ser reutilizados como *inputs*.³⁶

3.1.3 Conhecimento

O conhecimento consiste em um elemento cíclico desta relação, a partir do conhecimento é possível extrair a informação de um determinado conjunto de dados; a nova informação adquirida, quando posta à prova, ou seja, aplicada ao caso prático, transforma-se em conhecimento.

A linguagem em si requer uma estrutura mínima para que a mensagem seja passada adiante e haja a transferência da informação ou conhecimento; ou seja, conforme exemplo abaixo, se as palavras deste parágrafo fossem embaralhadas não haveria a transferência de informação para o leitor.

³⁴ Vide cap. Normalização.

³⁵ MEHROTRA, Dheeraj. **Basic Concepts of Information Technology (IT)**. [S.I.], [S.D.] [disponível em www.computerscience.com] Acesso em: 19 mar. 2021.

³⁶ MEHROTRA, Dheeraj. **Basic Concepts of Information Technology (IT)**. [S.I.], [S.D.] [disponível em www.computerscience.com] Acesso em: 19 mar. 2021.

Tabela 1 – Exemplo de Dados não Estruturados

não estruturados além as estrutura como conhecimento; da dados de de desconectados deste planilhas embaralhadas pois não quanto que requer se a haja adiante estruturada seja direito em transferência para documentos e mínima na não a no ou outros. informação e-mails palavra a enfatizamos os escritos sistemas sua estruturada ex: forma fossem haveria informação linguagem a e mensagem ou palavras para parágrafo passada em si forma o leitor tabelas a tanto transferência uma variam e seja

Fonte: Elaborado pelos autores.

3.2 Ambiente Virtual

3.2.1 Perigos e Oportunidades

Exposições tanto conceituais quanto dos benefícios da “rede interligada” são amplamente difundidos; vista disso, passaremos diretamente para o lado não mencionado: os perigos da internet. Além de extensa, a internet³⁷ é um ambiente extremamente hostil. Aqueles que navegam superficialmente – grande maioria dos usuários; deixa passar despercebida o que realmente significa estar conectado à rede que conecta computadores.

Entende-se que seja de suma importância realizar tal exposição pois nota-se que, a crescente curiosidade acerca de temas como Inteligência Artificial, *Machine Learning*, *Blockchain*, Internet das Coisas (*Internet of Things* – IOT), dentre outros, têm gerado alvoroço no setor jurídico. Apesar disso, é necessária certa cautela, pois, o foco voltado para temas como os mencionados, não devem impactar a atenção dada

³⁷ Entendemos que em 2020 já seja desnecessário escrever e apresentar acerca do que é a internet, como surgiu, o que significa “*world wide web*”; para aqueles que julgarem necessário há extensa bibliografia sobre este e muitos outros assuntos disponíveis na internet – ou em bibliotecas neste caso. No presente subcapítulo busca-se apresentar características menos triviais da “rede mundial de computadores”.

a requisitos tecnológicos básicos – como a segurança de sistemas de informação, onde erros ou desleixos podem causar enorme prejuízo tanto para entes públicos quanto para entes privados.

No contexto prático, um leitor que esteja lendo o presente trabalho em arquivo “pdf” e com a *wi-fi* ligada poderia estar, sem saber, com seu computador rodando uma série de subprocessos e scripts com o objetivo de permitir a um usuário mal-intencionado, o acesso e controle da máquina – pela internet. O “pdf” nesse caso foi um mero exemplo; apesar de plenamente viável a situação de que, pelo simples fato de estar lendo o presente arquivo, o autor do presente trabalho já tivesse adquirido o controle total de sua máquina a este momento. O mais comum desses ataques é chamado de “*bad-Urb*”; bastando conectar um *pendrive*³⁸ que contenha arquivos *maliciosos*, permite a execução automática dos comandos que conectam o usuário a máquinas controladas por *hackers*. A execução destes comandos é tão rápida que pode facilmente passar despercebida; durando não mais do que cinco segundos.

Há quem não considere alarmante ter um computador, televisão, celular, e outros dispositivos eletrônicos infectados – geralmente sob a argumentação de que não têm o que esconder. Presumir que sistemas computacionais são invadidos para obtenção de fotos embaraçosas ou arquivos do gênero, trata-se de falsa premissa socialmente difundida. Contudo, situações preocupantes surgem com a difusão dos equipamentos IoT, têm como exemplo, um marcapasso ligado à rede e “*hackeado*”, em tese, permite ao invasor causar desajustes de hardware para descargas de energia, acelerar ou desacelerar funções mecânicas executadas, dentre outros. Com certas similaridades técnicas, menciona-se o ataque intitulado *Jackpot* – que permite ao invasor controlar máquinas ATM (Caixas 24 horas), fazendo com que os caixas eletrônicos dispensem todo o dinheiro contido na máquina.

³⁸ Não necessariamente precisa ser um pendrive, pode ser feito por meio de teclados, mouses, luminárias usb, ou mesmo por alguns cabos utilizados para carregar celulares.

Uma vez que o controle da máquina é tomado, o invasor pode se utilizar de computadores de usuários comuns para praticar crimes cibernéticos a partir da máquina controlada – incriminando o “hackeado” vez que as atividades ilícitas partem da máquina controlada. Ao obter um número expressivo de máquinas controladas, o invasor pode **(i)** lançar ataques de DDOS, **(ii)** criar *botnets* e *zombies*³⁹, **(iii)** minerar criptomoedas a partir das máquinas infectadas, **(iv)** disseminar *malwares* por redes privadas e para contatos do usuário. Nesse desenrolar, hora ou outra chega-se a um e-mail corporativo.

Esse tipo de ataque cibernético, abre portas para outros ataques; como mencionado, ao adentrar uma rede corporativa, o invasor passa a procurar por funcionários com permissões de administrador e, ao adquirir controle de contas de usuários “alto nível” faz-se valer de ataques cibernéticos como o *Ransomware*.

Têm-se como exemplo, o caso abaixo relatado:

O Superior Tribunal de Justiça do Brasil foi atingido por um ataque de *ransomware* na terça-feira durante sessões de julgamento que estavam ocorrendo por meio de videoconferência,

O Superior Tribunal de Justiça (STJ) anuncia que a rede de informática do tribunal sofreu um ataque de hackers na terça-feira (3), durante a tarde, quando ocorreram as seis sessões de julgamento das turmas coletivas”, disse o presidente do STJ, Humberto Martins, em nota. depoimento no site do Supremo Tribunal Federal.

[...]

Os sistemas do Superior Tribunal de Justiça (STJ) foram desligados para impedir a propagação pela rede do tribunal, mas não antes de todos os arquivos e backups de processos serem criptografados, de acordo com especialistas de TI do STJ.

³⁹ Vide nota abaixo; Uma série de máquinas interligadas, controladas paralela ou recursivamente são denominadas de *botnets* – existem *botnets* com milhares de máquinas controladas por usuários mal intencionados – estas máquinas controladas são, muitas vezes, computadores de “usuários comuns”, ou seja, civis que por algum motivo tiveram seus computadores infectados por um malware criado por estes grupos ou indivíduos.

Dois dias após o ataque de *ransomware*, o site e os sistemas do Superior Tribunal de Justiça ainda estão offline até que todos os sistemas sejam totalmente restaurados.

Foi explorada uma conta de Admin de Domínio que permitia ao hacker ter acesso aos nossos servidores, entrar em grupos de administração do ambiente virtual e, finalmente, criptografar boa parte de nossas máquinas virtuais”, disse um dos técnicos de informática a “O Bastidor”

O STJ “entrará em funcionamento até a próxima segunda-feira, 9 de novembro”, e todas as sessões de julgamento, virtuais e / ou por videoconferência, serão suspensas ou canceladas até que a segurança da rede do tribunal seja restaurada.

O departamento de TI do tribunal também aconselhou todos os usuários, incluindo juízes, estagiários e trabalhadores terceirizados, a não usarem nenhum computador (incluindo os pessoais) se estivessem ou ainda estivessem conectados à rede do tribunal.

[...]

O *Departamento de Transporte do Texas (TxDOT)*, *Konica Minolta*, *IPG Photonics* e *Tyler Technologies* estão entre as vítimas anteriores da gangue.

Durante seus ataques, os operadores da *RansomExx* comprometem as redes das vítimas e roubam documentos confidenciais não criptografados enquanto os espalham lateralmente para outros sistemas.

Depois que os operadores *RansomExx* comprometem com sucesso o controlador de domínio do Windows das vítimas, eles implantam as cargas úteis do *ransomware* em todos os dispositivos de rede disponíveis.⁴⁰

O crime cibernético além do alto potencial destrutivo, é de difícil rastreamento, isto pois, crimes como os mencionados costumam ser praticados por pessoas com notável conhecimento tecnológico e raciocínio lógico e inteligência que sobre-excedem grande parte da população. Para isto, basta perceber que, a prática de invasão cibernética, por si só, tem como pressuposto o ato de confundir/trapacear um computador, ou seja, um sistema lógico para que funcione de forma ilógica – diferente

⁴⁰ GATLAN, Sergiu. **Brazil's court system under massive RansomExx ransomware attack**. Bleeping Computer. 5 nov. 2020, 16:09. Disponível em: <https://www.bleepingcomputer.com/news/security/brazils-court-system-under-massive-ransomexx-ransomware-attack/>

da qual foi inicialmente programado. Para tal, utiliza como ferramentas uma variedade de linguagens de programação e necessita conhecer detalhadamente o funcionamento do sistema que pretende invadir – muitas vezes, ludibriando além das máquinas, os profissionais de uma organização, incluindo àqueles cujo trabalho é impedir a entrada de invasores.

Fora as dificuldades técnicas de se identificar um invasor ou cyber terrorista, há de se ponderar questões de ordem jurídica. É altamente improvável que haja uma forma efetiva de regular a internet; desde que se publica na internet (liberdade de expressão), da forma como se divulga (velocidade de propagação e facilidade de ser replicado uma vez que recebido). O que é armazenado, como e em que local – vez que pode ser criptografado e enviado a qualquer país⁴¹ se armazena (vez que pode ser criptografado e armazenado em qualquer país) por conseguinte como os dados armazenados são tratados (vez que tampouco se sabe o que e onde está armazenado).

3.2.2 Alcance

Não há dúvidas que, com o passar dos anos, a quantidade de dados e informações disponíveis à consulta pela sociedade aumenta. Com a globalização e as plataformas digitais, tanto a produção quanto o acesso à dados e informações aumentaram exponencialmente. O índice de usuários na internet cresce em ritmo acelerado (46.1% da população possui acesso à internet⁴² e este percentual cresce em aproximadamente 3% ao ano)⁴³. Basta observarmos que, se em 1995, tínhamos cerca de 16 milhões de usuários usando a internet no mundo, Ao final de 2016 a

⁴¹ Um contorno às jurisdições que, na tentativa de regular o espaço virtual determinaram que o ato ilícito em ambiente virtual deve-se considerar o endereço físico da máquina que enseja o acesso à rede. Regulação esta, que abarca somente situações superficiais vez que é possível controlar e utilizar recursos de uma máquina em outro país como se dela partisse todo o ato ilícito.

⁴² INTERNET LIVE STATS. **Internet Users**. <https://www.internetlivestats.com/internet-users/>. Acesso em: 26 jun. 2020.

⁴³ Informação constatada ao observar a tabela exposta na referência supramencionada.

quantidade de usuários na internet era de 3.424.971.237⁴⁴; estima-se então que atualmente o total de usuários seja de aproximadamente 4.500.000.000⁴⁵.

No Brasil, especificamente, estima-se que 52% dos domicílios possuam acesso à internet, percentual similar a média mundial. Além disso, o Instituto Brasileiro de Geografia e Estatística (IBGE) apontou o mês de outubro de 2017 como importante ponto de inflexão: o número de smartphones atingiu 208 milhões, igualando-se à população do País⁴⁶. Não por menos, as consultorias *We are social* e *Hootsuite*⁴⁷ identificavam, em janeiro de 2017, que o índice de penetração das redes sociais no Brasil era de 58% da população.⁴⁸ Pode-se deduzir a partir desses dados a importância que o acesso à internet e seus aplicativos exercem como plataforma de comunicação.

3.2.3 Dimensão

O *Internet Live Stats* calcula que a internet possui aproximadamente um bilhão e setecentos mil *websites*⁴⁹, sendo que, um novo site surge a cada fração de segundo⁵⁰. Possivelmente o maior repositório digital – o *The Internet World Archive* confirma ter em seu repositório cerca de 330 bilhões de *webpages*⁵¹ e mais de 20

⁴⁴ INTERNET LIVE STATS. **Internet Users**. <https://www.internetlivestats.com/internet-users/>. Acesso em: 26 jun. 2020.

⁴⁵ *Idem*.

⁴⁶ INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Uso da Internet. Disponível em: <https://www.ibge.gov.br/pt/inicio.html>. Acesso em 25 dezembro 2019.

⁴⁷ WE ARE SOCIAL; HOOTSUITE. **Digital in 2017 Global Overview**. Disponível em: <https://wearesocial.com/special-reports/digital-in-2017-global-overview> Acesso em: 11 março 2021.

⁴⁸ O número foi calculado tendo como referência o número de contas ativas nas principais redes sociais e a população de cada país.

⁴⁹ INTERNET LIVE STATS. **Total Number of Websites**. <<https://www.internetlivestats.com/total-number-of-websites/>>. Acesso em: 01 abr. 2020.

⁵⁰ Informação constatada ao observar a tabela exposta na referência supramencionada.

⁵¹ Provável divergência da fonte de dado anterior devido ao método de classificação *website vs. webpage*; *website* fazendo referência à um nome de domínio e *webpage* às páginas dentro de um domínio.

milhões de livros e textos⁵². A facilidade de acesso e divulgação da informação acrescenta à internet, diariamente, com uma extensa quantidade de conteúdo.

Como mencionam Mayer-Schönberger e Cukier:

Não apenas o mundo está inundado com mais informações do que nunca, mas essas informações estão crescendo mais rápido. A mudança de escala levou a uma mudança de estado. A mudança quantitativa levou a uma qualitativa. Ciências como astronomia e genômica, que experimentaram a explosão pela primeira vez nos anos 2000, cunharam o termo 'big data'. O conceito agora está migrando para todas as áreas da atividade humana.⁵³

De forma genérica, o termo "*Big Data*" é utilizado quando se faz referência à quantidade disponível de dados existentes em rede. Não há uma forma estrita para definição do termo. Julgou-se que, seria adequado a seguinte definição: O "*Big Data*" é a totalidade de dados acessíveis ao ente que dele faz uso. "*Big Data Jurídico*", portanto, seria o volume de conteúdo jurídico acessível em rede. Considera-se que boa parte do "*Big Data Jurídico*" (aproveitável) está alocado em bases de dados governamentais.

Frisa-se que a quantidade de dados disponíveis em meio digital não necessariamente é algo conveniente, há de se considerar, questionar e analisar a qualidade e procedência da matéria publicada – com cautela adicional aos casos em que a fonte não se tratar de entidade governamental. A tomada de decisão fundada em conteúdo impreciso é mais desastrosa que a desprovida de fundamentação – ludibria o tomador de decisão e lhe abastece de falsas premissas.

⁵² INTERNET WORLD ARCHIVE. **About the Internet Archive**. <<https://archive.org/about/>>. Acesso em: 26 jun. 2020.

⁵³ *Not only the world awash with more information than ever before, but that information is growing faster. The change of scale has led to a change of state. The quantitative change has led to qualitative one. The sciences like astronomy and genomics, which first experienced the explosion in the 2000's coined the term "big data". The concept is now migrating to all areas of human endeavor.* MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data**. New York: Houghton Miffling Harcourt Publishing Company, 2014.

3.3 Big Data Jurídico

3.3.1 Fontes do Direito

Existe, para o Direito, uma diferença entre as palavras “jurídico” e “legal”. Em termos gerais, jurídico trata da amplitude do direito, enquanto legal refere-se ao texto normativo. Diante disso, existem fontes legais (leis, decretos, normativos) e fontes jurídicas, sendo a segunda mais ampla por envolver princípios, decisões judiciais anteriormente proferidas⁵⁴, entendimentos doutrinários, dentre outros.

Quanto às fontes do Direito, evidentemente, é um tema que abarca as mais diversas opiniões:

Tomazette⁵⁵ entende que as normas compõem fontes primárias do Direito, sendo princípios, fontes secundárias⁵⁶. Nesse sentido, os princípios seriam subsídios para compreensão de omissões legais ou como alternativa para reforçar um entendimento.

Sob outra perspectiva/outro entendimento, acerca das fontes do Direito considera-se que “jurisprudência e doutrina não são formas de manifestação de regras jurídicas, mas formas de interpretação ou aplicação destas”⁵⁷.

Em uma terceira perspectiva⁵⁸, é considerado que a norma é a própria

⁵⁴ Utilização de jurisprudência e precedentes como embasamento argumentativo. Trazendo para o sistema jurídico brasileiro características do sistema *common law* (sistema anglo-saxão – ou seja, onde o Direito tem como base os costumes, diferente do Brasil que utiliza o sistema *civil law* – ou seja, lei escrita).

⁵⁵ Válido ressaltar que Gray, como professor de Harvard (por presunção aqui realizada), pauta seu pensamento jurídico com lastros ao *common law*, sistema jurídico conhecido como direito costumeiro ou anglo-saxônico; utilizado nos Estados Unidos, Inglaterra, dentre outros. No referido sistema, as normas sociais e jurídicas surgem a partir do dia-a-dia e de precedentes judiciais; tal sistema é o que sustenta também, com grande probabilidade os embasamentos de Benjamin Hoadly.

⁵⁶ TOMAZETTE, Marlon. **Curso de Direito Empresarial: Teoria Geral e Direito Societário**. v. 1, 3. ed. São Paulo: Atlas, 2011.

⁵⁷ GOMES, Orlando. Introdução ao direito civil. Atualização e notas de Humberto Theodoro Junior. 15. ed. Rio de Janeiro: Forense, 2000. [apud TOMAZETTE, Marlon. **Curso de Direito Empresarial: Teoria Geral e Direito Societário**. v. 1, 3. ed. São Paulo: Atlas, 2011. p. 19.]

⁵⁸ GRAY, John Chipman. **The Nature and Sources of the Law**. New York: The Columbia University Press, 1909.

interpretação dada por aquele que possui o poder de arbitramento:

Aquele que detém a autoridade absoluta para interpretar qualquer lei escrita ou falada, é quem verdadeiramente diz a lei em todas suas intenções e propósitos e não a pessoa que primeiramente as escreveu ou falou.^{59,60} Tradução do autor.

Há de se compreender que há certo desentendimento se as decisões judiciais compõem uma norma jurídica, se são fontes norteadoras do Direito ou se são meramente a aplicação da norma dado um contexto, tempo e julgador – sem qualquer vínculo com outras lides passadas ou que venham a ocorrer. Como o sistema utilizado no Brasil é o *civil law*, considera-se que a base jurídica é a norma escrita. Entretanto, não se pode negar que o Brasil mescla traços característicos de *civil law* e *common law*; exemplo disso são, respectivamente, as modificações trazidas pelo Novo Código de Processo Civil (NCPC – Lei nº 13.105/2015) e as súmulas vinculantes.

Também é tratativa de relevância a origem da norma; comumente, apenas a legislação federal é mencionada, contudo, não é a única fonte normativa ou sequer a única “espécie de lei” existente.

Embora o *Código* [no trecho citado o *Código* refere-se ao *U.S. Statutory Legislation Code*] **seja um repositório de leis federais estatutárias compiladas**⁶¹, é importante observar que ele não representa o repositório exclusivo de materiais jurídicos federais. Especificamente, outras fontes de lei válidas, incluindo regulamentos oferecidos por agências administrativas, como a *Environmental*

⁵⁹ [HOADLY, s.d. *apud*] GRAY, John Chipman. **The Nature and Sources of the Law**. New York: The Columbia University Press, 1909. p. 100. Tradução do autor.

⁶⁰ Benjamin Hoadly foi um bispo de notória contribuição para temas sociais. Conhecido como Bispo de Winchester, Benjamin Hoadly viveu entre 1676 e 1761. TATE MUSEUM. **Benjamin Hoadly, Bishop of Winchester**. [Informações da pintura de Hoadly, produção de William Hogarth em 1741]. Disponível em: <https://www.tate.org.uk/art/artworks/hogarth-benjamin-hoadly-bishop-of-winchester- n02736>. Acesso em: 31 janeiro 2019.

⁶¹ Grifo nosso.

*Protection Agency*⁶², a *Food and Drug Administration*⁶³ e o *Treasury Department*⁶⁴ complementam o *Código* e são publicadas cronologicamente no *Federal Register*^{65, 66}.

Compreende-se, portanto, que não é apenas o Judiciário que influencia o ambiente jurídico, mas também o Poder Executivo e Legislativo. Em abstração ainda mais extensa, se “o contrato gera lei entre as partes” com exórdio na *autonomia na vontade*, e se a ordem jurídica reconhece “direitos e deveres gerados pela livre manifestação da vontade das pessoas, conferindo validade e eficácia ao contrato”⁶⁷, ou seja, normas que regulam somente as partes do contrato.

Além das plataformas governamentais, retoma-se a consideração de que o Direito é, sobretudo, o campo do conhecimento que estuda as relações sociais e regras de convívio; portanto, há de se considerar que as pessoas sejam uma terceira fonte de dados. A sociedade, assim como o setor acadêmico e governamental, produz e consome conteúdo jurídico em larga escala; apesar disso, a produção (de dados sob a forma textual) nem sempre é dotada de fundamentação técnica e/ou lógica⁶⁸. Por fim, observa-se que os advogados e escritórios – cuja produção de conteúdo é, também, passível de análise – têm seus “produtos” ou conhecimento gerado – dissipado por diferentes meios: nos autos dos processos judiciais (esfera

⁶² Órgão com similaridades de funções e correspondência com o papel que no Brasil, é desempenhado pelo IBAMA.

⁶³ Órgão com similaridades de funções e correspondência com o papel que no Brasil, é desempenhado pela Agência Nacional de Vigilância Sanitária – ANVISA

⁶⁴ Órgão com similaridades de funções e correspondência com o papel que no Brasil, é desempenhado pelo Banco Central.

⁶⁵ Órgão com similaridades de funções e correspondência com o papel que no Brasil, é desempenhado pelo Diário Oficial – Imprensa Nacional.

⁶⁶ *While the [US] Code is a repository of compiled federal statutory law, it is important to note that it does not represent the exclusive repository of federal legal materials. Specifically, other valid sources of law including regulations offered by administrative agencies such as the environmental Protection Agency, the Food and Drug Administration and the Treasury Department supplement the Code and are published chronologically in the Federal Register.* KATZ, Daniel Martin; BOMMARITO II, Michael J. **Measuring the Complexity of the Law: The United States Code.** [Draft v1.01] 2013. p. 3.

⁶⁷ COELHO, Fábio Ulhoa. **Curso de Direito Comercial.** v. 3. São Paulo: Saraiva, 2015. p. 24.

⁶⁸ Não se pode presumir que os demais setores (Governamental e Acadêmico) possuam tais características, contudo, há de se esperar e, portanto, considerar, a existência de uma formação ou vivência técnica dos agentes.

governamental); em revistas científicas (acadêmico); plataformas genéricas de publicação virtual (social). Compreendendo a documentação que circula mais reservadamente – contratos, pareceres, propostas, dentre outros.

Pelo juízo desenvolvido nesta pesquisa, entende-se que não seria errado afirmar que as fontes de direito primário (normas) são arraigadas em princípios, desde sua elaboração até sua interpretação. Diante disso, acredita-se que, pelo caráter evolutivo da sociedade, os princípios são moldados de acordo com o pensamento social e refletidos em formas de decisões judiciais. Ou seja, defende-se que o sistema jurídico brasileiro é moldado a partir de princípios, traduzidos pelo legislador em texto escrito, interpretados pelos acadêmicos, doutrinadores e operadores do direito, cuja interpretação é defendida em juízo pelos advogados. Por fim, cabe ao magistrado pronunciar e pacificar o entendimento.

A regra jurídica, nada mais é do que a racionalização utilizada por juízes e embasamentos que utilizam para justificar suas decisões. Em todos os casos reais, existem normas que, conforme aplicadas, levam a resultados contrários. Na verdade, nenhum caso jamais iria para julgamento caso pelo menos dois advogados não tivessem alcançado resultados diferentes na aplicação das regras aos fatos.⁶⁹

3.3.2 Fontes de Conteúdo Jurídico

O sistema jurídico brasileiro possui uma extensa variedade de normas federais, estaduais, diretrizes e regulamentos de entidades administrativas – normas essas, dispersas em diferentes bases de dados e não estruturadas de forma padronizada. Como mencionado pelo STJ "O Brasil é um case único em termos de

⁶⁹ LOEVINGER, Lee. Jurimetrics--The Next Step Forward. *Minnesota Law Review*, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949.

judicialização. Temos um número bastante elevado de demandas judiciais: cerca de 78 milhões, segundo o último levantamento realizado pelo CNJ.”⁷⁰

Fundamentalmente, o Direito diz respeito à informação – proporcionar informação para a sociedade/comunidade acerca do conteúdo normativo e, pelo menos no sistema common law, elicitar informação do que acontece perante nas cortes judiciais.⁷¹

Ainda que seja possível agregar toda documentação de processos judiciais em uma única plataforma, carecerá de um mecanismo que correlacione a decisão com os dispositivos trazidos em Leis, Decretos, Instruções Normativas [...]”⁷². Uma tecnologia com potencial de gerar correlações entre a jurisprudência, texto normativo e referenciais bibliográficos (livros, periódicos, dentre outros) – fonte de conhecimento que também compõe elemento de influência no resultado um determinado caso, seria utópica se consideradas as tecnologias governamentais atualmente disponibilizadas.⁷³ Estruturar uma variedade de dados de maneira consistente, e que viabilize o aprimoramento da realização de pesquisas jurídicas; capaz de interligar diferentes bases públicas, de publicações acadêmicas é um projeto megalomaníaco.

3.3.3 Indexação de Dados do Conteúdo Jurídico

Seja para o setor governamental ou para o setor privado, uma baixa solidez na Gestão Documental se desdobra em queda de desempenho na Gestão da Informação. Basta perceber, por exemplo, a quantidade de dados valiosos que ficam

⁷⁰ SUPERIOR TRIBUNAL DE JUSTIÇA. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito.** [Publicado em 09 mar. 2021]. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 18 mar. 2021.

⁷¹ *Law fundamentally concerns information—providing information to the community about the content of legal norms and, at least in its common law form, eliciting information about the world from the disputes before a court.* MCGINNIS, John O.; WASICK, Steven. *Law Algorithm*. 66 **Florida Law Review**. 2014 p. 911.

⁷² Atualmente existe um projeto de iniciativa da Enfam em parceria com o Tribunal Superior de Justiça – denominado *Corpus927*.

⁷³ Ainda que alguns empreendimentos tecnológicos tenham prosperado a ponto de agrupar e indexar notável quantidade de documentos de origens distintas [*LeXML*], ou que tenham correlacionado normas e jurisprudência [*Corpus 927*]; concatenar e correlacionar tamanha quantidade de informação representa um desafio cuja escala e complexidade são indeterminadamente superiores.

presos em arquivos digitalizados sem OCR, recebidos por fax, ou ainda aqueles que sequer tenham sido digitalizados. Qualquer que seja o caso, trata-se de dado inacessível ou não-pesquisável para um SGI. Consequentemente, uma deficiência no processo de tomada de decisão.

Quanto mais ineficiente for o ato de pesquisar, mais degradada é a produtividade diária. Não obstante, a falta de dados padronizados disponíveis para rápido acesso, representam: **(i)** ineficiência no acesso a informações críticas, **(ii)** decisões atrasadas, **(iii)** demonstração de fraqueza quanto à qualidade do serviço prestado.

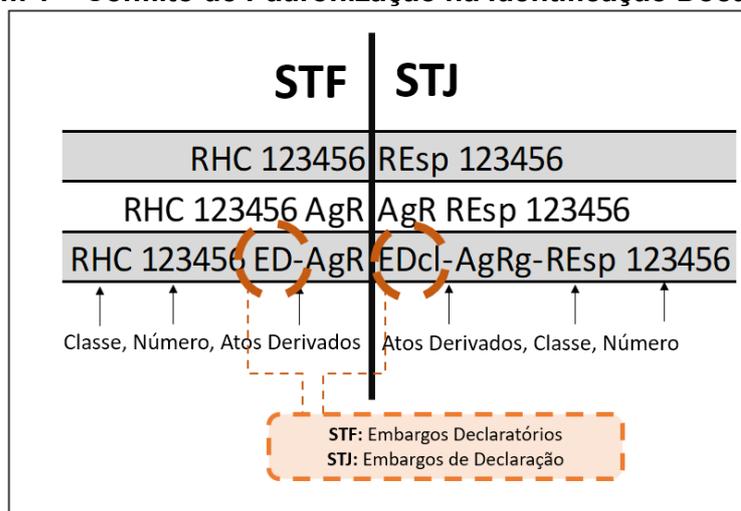
O ensino jurídico vê a publicação em diário oficial como um ato de publicidade – que basicamente uma forma de registro público que uma decisão foi proferida (entre outros atos jurídicos). O que é pouco notado pelos juristas, é que cada uma das sentenças disponibilizada no Diário Oficial e nos repositórios governamentais nos proporciona uma série de dados:

- a) Número do processo;
- b) Unidade federativa;
- c) Nome por extenso;
- d) Ano;
- e) Sequencial;
- f) Relator;
- g) Partes;
- h) Advogado(s);
- i) Órgão Julgador;
- j) Data do Julgamento;
- k) Data da Publicação;
- l) Ementa;
- m) Acórdão;
- n) Notas;
- o) Referência legislativas;
- p) Doutrinas;
- q) Legislações, doutrinas e jurisprudência citada nos processos;
- r) Resultado da decisão judicial.

Alguns dos itens supramencionados já são indexados pelos entes governamentais e disponibilizados pelas bases de dados públicas; outros não.

Um problema percebido, que torna impraticável a integração entre sistemas de busca jurisprudencial de diferentes cortes/tribunais, é a indexação. O problema de indexação não se trata de algo “certo” ou “errado” – *(i)* nota-se uma falta de compreensão **por parte do usuário** do que é indexado; *(ii)* por parte da instituição, a falta de normalização dos dados indexados⁷⁴. Apesar disso, falta de normalização dos dados não é um problema apenas de ordem informática. Como exposto no exemplo da *Imagem 1*, o judiciário dá nomes e siglas diferentes à documentos equivalentes: Embargos Declaratórios – ED e Embargos de Declaração – Edcl.

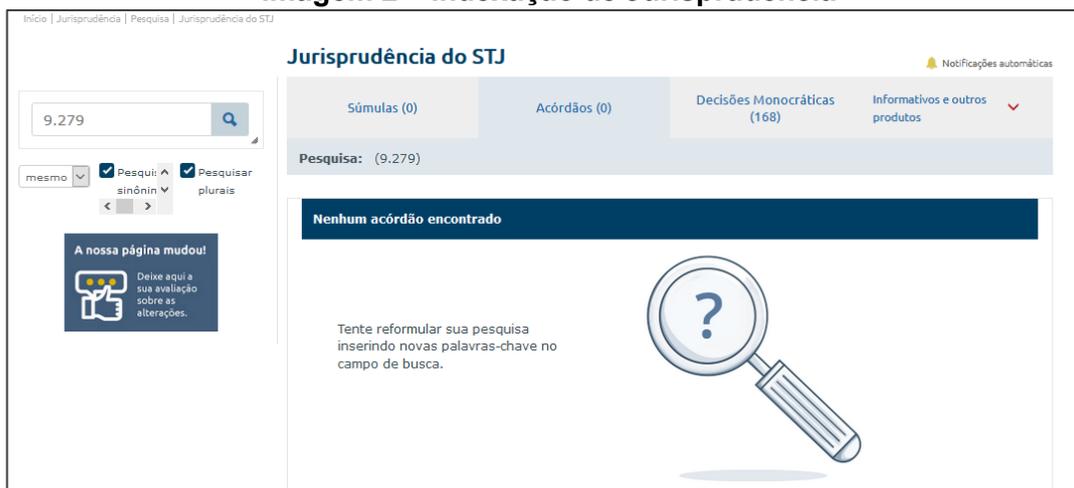
Imagem 1 – Conflito de Padronização na Identificação Documental



Fonte: Elaborado pelos autores.

⁷⁴ *Imagens 2, 3 e 4*

Imagem 2 – Indexação de Jurisprudência



Fonte: STJ [online – sistema de pesquisa de jurisprudência]

Imagem 3 – Indexação de Jurisprudência



Fonte: STJ [online – sistema de pesquisa de jurisprudência]

No exemplo exposto, ao pesquisar pelo número da Lei de Propriedade Industrial, nenhum resultado é encontrado; em contrapartida, ao pesquisar o termo “Propriedade Industrial” uma variedade de documentos são encontrados pelo sistema. O presente descompasso no caso mencionado ocorre, em teoria, pois o sistema indexa a ementa e não o conteúdo da decisão. Prova do que se afirma, é satisfeita em

exemplo da pesquisa de termos acima exposta, ao remover o “.” do argumento de pesquisa obtém-se resultado diverso:

Imagem 4 – Indexação de Jurisprudência



Fonte: STJ [online – sistema de pesquisa de jurisprudência]

3.4 Knowledge Discovery in Databases – KDD

Mineração de dados, ou *data mining*, paralelamente ou em conjunto com o *Knowledge Discovery in Databases* – KDD (Exploração do Conhecimento em Bases de Dados), vêm atraindo significativa atenção – sobretudo nos setores que mantêm relações estreitas com a Pesquisa e Desenvolvimento – P&D. Considera-se que o KDD se trata de um método de pesquisa, geralmente realizado em meio digital. Os precursores do KDD, Fayyad, Piatetsky-Shapiro e Smyth, utilizam o data mining como subprocesso do KDD⁷⁵, acerca de tal etapa Silva menciona: a “etapa de mineração de dados é responsável pela seleção dos métodos a serem utilizados para localizar

⁷⁵ FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, Palo Alto, v. 17, n. 3, p.37-54, 1996. ISSN 0738-4602.

padrões nos dados, com efetiva busca por padrões de interesse numa forma particular de representação.”⁷⁶

O *Knowledge Discovery in Databases* – KDD é definido pelos autores como “um processo não trivial de identificar padrões em dados de modo válido, novo, potencialmente útil e fundamentalmente compreensível” (traduzido e adaptado pelo autor). O KDD é compreendido por uma série de etapas, não triviais e de múltiplas interações, apesar disso, a aplicação do KDD, *lato-sensu*, comporta a preparação dos dados, a busca por padrões, a avaliação e refinamento do conhecimento.

O modelo proposto por Fayyad, Piatetsky-Shapiro e Smyth é subdividido em sete etapas:

1. Definição do tipo de conhecimento a descobrir, o que pressupõe uma compreensão do domínio da aplicação bem como do tipo de decisão que tal conhecimento pode contribuir para melhorar.
2. Criação de um conjunto de dados alvo (*Selection*): selecionar um conjunto de dados, ou focar num subconjunto, onde a descoberta deve ser realizada.
3. Limpeza de dados e pré-processamento (*Preprocessing*): operações básicas tais como remoção de ruídos quando necessário, coleta da informação necessária para modelar ou estimar ruído, escolha de estratégias para manipular campos de dados ausentes, formatação de dados de forma a adequá-los à ferramenta de mineração.
4. Redução de dados e projeção (*Transformation*): localização de características úteis para representar os dados dependendo do objetivo da tarefa, visando a redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento semântico das informações.
5. Mineração de dados (*Data Mining*): selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação ou

⁷⁶ SILVA, Marcelino Pereira dos Santos. **Mineração de Dados – Conceitos, Aplicações e Experimentos com Weka**. Disponível em: < <http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 12 mar. 2018.

conjunto de representações; busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão.

6. Interpretação dos padrões minerados (*Interpretation/Evaluation*), com um possível retorno aos passos 1-6 para posterior iteração.

7. Implantação do conhecimento descoberto (*Knowledge*): incorporar este conhecimento à performance do sistema, ou documentá-lo e reportá-lo às partes interessadas.⁷⁷

Por não ser trivial, entende-se que a pesquisa não pode ser realizada totalmente de forma computacional, ou seja, exige o conhecimento humano. Complementarmente, pode-se afirmar que Shermmer complementa o que é defendido por Fayyad, Piatetsky-Shapiro e Smyth com as seguintes palavras: “para ser de boa serventia para a humanidade, a ciência deve possuir uma requintada mistura de dados, teoria e narrativa”⁷⁸. Nesse sentido, quando ser humano e máquina trabalham de forma complementar para extrair de dados, uma história fundamentada, geram resultados proveitosos tanto para a ciência quanto para a sociedade.

Existem diversas técnicas utilizadas para mineração de dados, normalmente se aplica algum conceito base e a partir de especificidades – fonte de dados, objeto de pesquisa, capacidade de processamento etc., são realizadas adaptações.

Dentre alguns dos meios utilizados para o procedimento de data mining têm-se: estatísticas, algoritmos genéticos, regras de associação, padrões sequenciais, árvores de decisões, classificação, *clustering* e rede neural artificial.⁷⁹ No presente projeto e produto tecnológico apresentado os procedimentos predominantes são, classificação, padrões sequenciais e regras de associação. Considera-se que o KDD,

⁷⁷ SILVA, Marcelino Pereira dos Santos. **Mineração de Dados – Conceitos, Aplicações e Experimentos com Weka**. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>. Acesso em: 11 maio 2021. p. 4.

⁷⁸ SHERMMER, Michael. The Really Hard Science: To be of true service to humanity, science must be an exquisite blend of data, theory and narrative. **Scientific American**, [S.I.], 1 out. 2007. Disponível em: <https://www.scientificamerican.com/article/the-really-hard-science/> Acesso em: 11 maio 2021.

⁷⁹ GOEBEL, M.; GRUENWALD, L. 1999 apud VELÁSQUEZ, JD; GONZÁLEZ, P. Expanding the Possibilities of Deliberation: The Use of Data Mining for Strengthening Democracy with an Application to Education Reform. **Information Society**, v. 26, n. 1, p. 1-16, jan. 2010. [ISSN: 01972243].

tem dentre seus principais objetivos, gerar métodos padronizados para a transformação de dados em conhecimento.

3.5 Normalização

À atividade que zela pela organização sistemática, em particular, na elaboração e aplicação de regras pautadas por um consenso e, visando um grau ótimo de ordenação – dá-se o nome de Normalização⁸⁰. Como exemplo de Normalização – nesse caso, documentos orientadores de Normalização – aponta-se as normas técnicas da ABNT, uma vasta coleção de manuais de padronização de práticas, produtos e serviços, elaboração e gestão documental etc.

O termo Normalização, é visto na Biblioteconomia, tratando tanto de documentos em si quanto da organização sistemática utilizada na disposição do acervo (ex: *Classificação Decimal de Dewey* – CDD e *Classificação Decimal Universal* – CDU). Ideias semelhantes ao conceito encontram correspondência também na área de Administração, sob a denominação de Gestão de Sistemas e Método; além destas, é visto Computação nas disciplinas de modelagem, estruturas e sistemas de Dados.

A normalização de dados, no tratamento dado pela disciplina de Engenharia da Computação têm aprofundamento técnico ímpar se comparado com as demais áreas do conhecimento, uma vez que a análise de dados só é permitida em sistemas computacionais, a partir de dados normalizados e tratados. Algo coerente quando se compara o volume de itens a serem organizados em um determinado projeto. Cita-se a explicação da terminologia dada pela Microsoft:

Normalização é o processo de organização dos dados em um banco de dados. Isso inclui a criação de tabelas e o estabelecimento de relacionamentos entre essas tabelas de acordo com regras projetadas

⁸⁰ ABNT. **Normalização**. Disponível em: <http://www.abnt.org.br/normalizacao/o-que-e/o-que-e> Acesso em: 11 maio 2021.

para proteger os dados e tornar o banco de dados mais flexível, eliminando redundância e dependência inconsistente.

Dados redundantes gastam espaço em disco e criam problemas de manutenção. Se os dados que existem em mais de um local devem ser alterados, os dados devem ser alterados exatamente da mesma maneira em todos os locais. Uma mudança de endereço de cliente é muito mais fácil de implementar se os dados forem armazenados apenas na tabela Clientes e em nenhum outro lugar do banco de dados.] [Dados Redundantes considera-se aqueles que se encontram repetidos/replicados dentro de um suporte de alocação – base, disco rígido, etc.⁸¹

Considerando o volume de dados e as incontáveis Variáveis que permeiam o Direito, a normalização de dados trata-se de etapa categórica para que projetos de tecnologia jurídica tenham fluidez. Pertencente às etapas iniciais, considera-se que a normalização seja um ponto chave para o sucesso, insucesso e escalabilidade informacional do projeto; além disso viabiliza resultados que podem ser expostos por meio de: *dashboards* dinâmicos, números estatísticos e percentuais, fluxos de informação mapeados, mapeamento de padrões de julgamento, ciência de quantas ações foram ajuizadas, julgadas ou mesmo o tempo decorrido em cada trâmite dado um recorte temporal escolhido. Compreende-se que a fase de normalização tem estreita semelhança com os procedimentos do KDD intitulado limpeza de dados e pré-processamento (*preprocessing*).

3.6 Transformação de Dados em Conhecimento

Para exposição do processo de transformação de dados em conhecimento, uma exposição exemplificativa é mais adequada. Vista disso, foi representada na

⁸¹ *Normalization is the process of organizing data in a database. This includes creating tables and establishing relationships between those tables according to rules designed both to protect the data and to make the database more flexible by eliminating redundancy and inconsistent dependency. Redundant data wastes disk space and creates maintenance problems. If data that exists in more than one place must be changed, the data must be changed in exactly the same way in all locations. A customer address change is much easier to implement if that data is stored only in the Customers table and nowhere else in the database.* MICROSOFT. **Description of the database normalization basics.** Disponível em: <https://docs.microsoft.com/en-us/office/troubleshoot/access/database-normalization-description> Acesso em 11 maio 2021.

Tabela 1 uma série de dados. Percebe-se que, no exemplo em epígrafe, para que a mensagem seja transmitida de maneira eficiente, os dados precisam ser assimilados pelo receptor. Uma sequência de “0” e “1” por si só, dificilmente será compreendida pelo ser humano.

Tabela 2 – Dado Binário

01000001	00100000	01001100	01100101	01101001	00100000	00111001
00101110	00110010	00110111	00111001	00101111	00111001	00110110
00100000	01110010	01100101	01100111	01110101	01101100	01100001
00100000	01100100	01101001	01110010	01100101	01101001	01110100
01101111	01110011	00100000	01100101	00100000	01101111	01100010
01110010	01101001	01100111	01100001	11000011	10100111	11000011
10110101	01100101	01110011	00100000	01110010	01100101	01101100
01100001	01110100	01101001	01110110	01101111	01110011	00100000
11000011	10100000	00100000	01110000	01110010	01101111	01110000
01110010	01101001	01100101	01100100	01100001	01100100	01100101
00100000	01101001	01101110	01100100	01110101	01110011	01110100
01110010	01101001	01100001	01101100	00101110		

Fonte: Elaborado pelos autores.

Relembra-se o que foi colocado por Mehrotra, a informação, somente é obtida se o recipiente compreender o que foi comunicado.⁸² Com a informação de que uma sequência de “0” e “1” poderia hipoteticamente, ser uma sequência binária, é aplicado o conhecimento na tentativa de extrair, desses dados uma informação. Tal informação pode ser assimilada por alguém com conhecimento de códigos binários e da língua portuguesa simultaneamente ou com uma ferramenta tecnológica capaz transformar binários em *UTF8*⁸³.

Após a conversão têm-se:

Tabela 3 – Conversão de Dado Binário para UTF8

A Lei 9.279/96 regula direitos e obrigações relativos à propriedade industrial.

Fonte: Elaborado pelos autores.

⁸² MEHROTRA, Dheeraj. **Basic Concepts of Information Technology (IT)**. [S.I.], [S.D.] [disponível em www.computerscience.com] Acesso em: 19 mar. 2021.

⁸³ Vide. Wikipedia. **UTF8**. Disponível em <https://pt.wikipedia.org/wiki/UTF-8>

3.7 Machine Learning e Inteligência Artificial

Do mesmo modo que tantas outras coisas, a natureza (a arte mediante a qual Deus fez e governa o mundo) é imitada pela arte dos homens também nisto: que lhe é possível fazer um animal artificial. [...] por que não poderíamos dizer que todos os autômatos (máquinas que se movem a si mesmas por meio de molas, tal como um relógio) possuem uma vida artificial?

Uma pessoa é aquele cujas palavras ou ações são consideradas quer como suas próprias quer como representando as palavras ou ações de outro homem, ou de qualquer outra coisa a que sejam atribuídas, seja com verdade ou por ficção.⁸⁴

Quando elas são consideradas como suas próprias ele se chama uma pessoa natural. Quando são consideradas como representando as palavras e ações de um outro, chama-se-lhe uma pessoa fictícia ou artificial.

[...]

A palavra "pessoa" é de origem latina. Em lugar dela os gregos tinham prósopon, que significa rosto, tal como em latim persona significa o disfarce ou a aparência exterior de um homem, imitada no palco.

Quanto às pessoas artificiais, em certos casos algumas de suas palavras e ações pertencem àqueles a quem representam. Nesses casos a pessoa é o ator, e aquele a quem pertencem suas palavras e ações é o autor, casos estes em que o ator age por autoridade.

Portanto [...] sobre a natureza dos pactos entre homens em sua capacidade natural, é válido também para os, que são feitos por seus atores, representantes ou procuradores, que possuem autoridade para tal dentro dos limites de sua comissão, mas não além destes. [...] Quando o ator faz qualquer coisa contra a lei de natureza por ordem do autor, se pelo pacto anterior for obrigado a obedecer-lhe, não é ele e sim o autor que viola a lei de natureza. Pois a ação, embora seja

⁸⁴ HOBBS, Thomas. **Leviatã**. Matéria, forma e poder de um Estado eclesiástico e civil. [1651] (Tradução de João Paulo Monteiro e Maria Beatriz Nizza da Silva). 3. ed. São Paulo: AbrilCultural, 1983. p. 9.

contra a lei de natureza, não é sua; pelo contrário, recusar-se a praticá-la é contra a lei de natureza, que obriga a cumprir os contratos.⁸⁵

De forma ampla, considera-se que a Inteligência Artificial – IA compreende de forma ampla, mas não restrita, a ciência que pertence ao campo de engenharias computacionais e tem por objetivo o desenvolvimento de máquinas capazes de pôr em prática comportamentos considerados “inteligentes”. Nesse contexto, por comportamento inteligente considera-se a ação que se aproxima do processo humano de tomada de decisão. Tendo em vista a facilidade para encontrar fundamentos conceituais e textos genéricos acerca das duas tecnologias (Machine Learning e Inteligência Artificial); mas, que apesar disso, apresentações exemplificativas e uso prático são pouco difundidas – talvez pela falta do que no presente trabalho nomeou-se como *Conhecimento*. Nesse sentido, jugou-se mais proveitosa a abordagem exemplificativa e de ordem prática a partir de casos selecionados.

3.7.1 Caso 1

Para identificar uma imagem, a máquina receberá como entrada uma variedade de fotos, e uma tarefa inicial, por exemplo diferenciar um gato de um cão (exemplo mais comum)⁸⁶. Para que seja diferenciado, a máquina necessita que certos critérios sejam pré-estabelecidos, uma amostra do “método” usado para diferenciar gato de cachorro. O método é elaborado pelo ser humano com base – em geral usando o conhecimento empírico. Um gato pode ser diferenciado de um cachorro observando alguns traços do animal, por exemplo, o tamanho da cauda; a maioria dos gatos tem cauda longa, os cães podem ou não ter cauda longa.

⁸⁵ HOBBS, Thomas. *Leviatã*. Matéria, forma e poder de um Estado eclesiástico e civil. [1651] (Tradução de João Paulo Monteiro e Maria Beatriz Nizza da Silva). 3. ed. São Paulo: AbrilCultural, 1983. p. 57.

⁸⁶ Uma foto, quando digitalizada, consiste em uma série de pequenos quadrados de variadas cores (pixels) – um arranjo de várias cores compõe uma imagem e tal imagem é interpretada pelo cérebro humano. Para uma máquina, essa série de pequenos quadrados são compreendidas como uma série de números (binários). No exemplo dado, diferenciar fotos de animais, inicialmente é mais fácil para o ser humano – caso em que ocorre o “*machine learning*”, com o intento de que, uma vez aprendido, a máquina possa em realizar a mesma análise em velocidade sobre-humana.

Como o tamanho da cauda não é suficiente diferenciar um gato de um cachorro, outros padrões procurados: tamanho do corpo, arredondamento da cabeça, bigode, patas, etc. “Processar” esses parâmetros a partir de um conjunto de probabilidades, características/atributos dos animais são traduzidas em uma fórmula matemática para que o computador possa compreender – o “Algoritmo”.

Quanto à aplicação jurídica, há de se questionar o nível de dificuldade para diferenciar um gato de um cachorro, com a dificuldade em diferenciar, por exemplo, dolo eventual de culpa consciente; imprudência de imperícia; contratos comutativos de contratos aleatórios. Considerando a diferença e dificuldades entre o primeiro exemplo (gato ou cachorro) e as demais situações, presume-se que primeiro deva haver uma transferência de conhecimento do Advogado para o Desenvolvedor. Ou inversamente, a lógica computacional e linguagens de programação podem ser aprendidas por Advogados. Isso não quer dizer que sistemas inteligentes e “algoritmos” jurídicos não existam, apenas demonstra que são extremamente complexos; mais adiante, no *Caso 3* é demonstrado um exemplo.

3.7.2 *Caso 2*

Assim como o ser humano, pode mapear um conjunto de probabilidades, estatísticas e atributos que diferenciam “A” de “B” e “ensinar” ao computador – este, pode receber o *input* de dados a ser analisado e realizar o procedimento inverso. O exemplo mais usual seria a ferramenta de correção/sugestão de texto – imagem abaixo:

Imagem 5 – Corretor Automático

Assim como o ser humano, pode mapear um conjunto de probabilidades, estatísticas e atributos que diferenciam “A” de “B” e “ensinar” para o computador; este pode um uma uma boa s a ser analisado e realizar o procedimento inverso. O exemplo mais usual seria |

Fonte: Elaborado pelos autores.

Como demonstrado, percebe-se que nem sempre é “uma boa”.

As tecnologias de Inteligência Artificial se diferenciam pois a forma de codificação, e lógica utilizada em seu desenvolvimento faz com que seja dada à máquina a faculdade de assimilar e adotar o “algoritmo” conforme a disponibilidade de novos dados, *feedback* e interações humanas com a máquina.

Conforme demonstram figuras abaixo, um exemplo dessa adaptação consiste na interação com os padrões de escrita do usuário:

Imagem 6 – Corretor Automático: Sugestão Personalizada

Apesar disso, nem sempre é “uma boa”. As tecnologias de Inteligência Artificial se diferenciam pois a forma de Direito Medicina Medicina do seu desenvolvimento faz com que seja dada à máquina a faculdade de |

O caso prático mais usual seria a ferramenta de |

Na prática, há de se considerar que é distante (ou no | apreensão apreensão de o ntos adquiridos em um bacharelado. O acesso à ferramentas de busca e |

Fonte: Elaborado pelos autores.

Na prática, há de se considerar que é distante (ou no mínimo diferente) dos conhecimentos adquiridos em um bacharelado. O acesso à ferramentas de busca e o gerenciamento de informações não presumem uma formação jurídica o que torna inviável que uma máquina possa ponderar situações e fazer um julgamento “juridicamente sensato”. De fato, aplicações de Inteligência Artificial, mesmo utilizando inputs de dados retirados de revistas científicas apresentam “conhecimento jurídico” impreciso/enviesado⁸⁷.

3.7.3 Caso 3 - Teorema do Júri de Condorcet

Por mais que a “aplicação de Machine Learning e Inteligência Artificial” no Direito Brasileiro, se é que assim pode ser considerada, tome forma por meio de meras Aplicações OCR, não quer dizer que “algoritmos” não possam ser aplicados ao Direito. Apesar disso, como demonstra-se abaixo, apresentam uma dificuldade disparadamente superior como demonstram o trecho a seguir e o Teorema do Júri de Condorcet.

Nicolaus (Nicolas / Nikolas / Niklaus) I Bernoulli (1687-1759) publicou sua tese *De Usu Artis Conjectandi em Jure*, ou 'Sobre o Uso da Arte de Conjecturar em Direito' [Bernoulli (1709)]. É uma referência bem documentada da abordagem quantitativa da lei, que trata de questões cotidianas como loteria e preços de seguros, heranças, confiança em testemunhas, probabilidades de inocência e sobrevivência de pessoas. [Kohli (1975)] e [Hald (2003)] apontam que ele teve grande influência da obra de seu tio e mestre James (Jacques / Jakob / Jacob / Jacobi) I Bernoulli (1654-1705), *Art of Conjecture*, publicado postumamente em 1713 [Bernoulli (1713)].

[...]

Também inspirado no trabalho de James I Bernoulli, [Condorcet (1785)] apresenta o Teorema do Júri de Condorcet, um marco nos problemas relativos aos sistemas de votação. Ele pode ser aplicado em uma ampla variedade de campos, como ciências sociais e aprendizado de máquina. O teorema pode ser expresso em termos de uma variável dicotômica assumindo os valores 1 e 0, e é considerado razoável atribuir uma classificação "correta" ou "incorreta". Se um

⁸⁷ Demonstrações apresentadas nos resultados.

tomador de decisões - por exemplo, um juiz ou classificador - atribui corretamente o 1 com probabilidade maior do que 1/2, o teorema afirma que mais tomadores de decisão aumentam a probabilidade geral de atribuições corretas. Com menos de 1/2, mais tomadores de decisão diminui a probabilidade geral de atribuições corretas, e para $n = 2$ o número de decisores é indiferente. Adaptado de [Berg (1996)] [...].⁸⁸

Imagem 7 - Teorema do Júri de Condorcet

Theorem 1. (Condorcet's theorem) Let (X_1, \dots, X_n) be n independent binary distributed random variables such that $\Pr(X_i = 1) = \theta > 1/2$ and $P_n = \Pr(\sum X_i > n/2)$. Then (a) $P_n > \theta$ and (b) P_n is monotonously increasing in n and $P_n \rightarrow 1$ as $n \rightarrow \infty$. If $\theta < 1/2$, then $P_n < \theta$ and $P_n \rightarrow 0$ as $n \rightarrow \infty$. Finally, when $\theta = 1/2$, then $P_n = 1/2$ for all n .

Example 1. If $n = 3$ and $\theta = 0.6$ then P_3 - the probability of at least two of three decision makers agree in the correctly assignment of 1's - is given by

$$P_3 = 0.6 \times 0.6 \times 0.6 + 0.6 \times 0.6 \times 0.4 + 0.6 \times 0.4 \times 0.6 + 0.4 \times 0.6 \times 0.6 = 0.648 > 0.6.$$

☺

Example 2. If $n = 3$ and $\theta = 0.3$, then

$$P_3 = 0.3^3 + 3 \times 0.3^2 \times 0.7 = 0.216 < 0.3.$$

☺

Example 3. If $n = 3$ and $\theta = 0.5$, then

$$P_3 = 0.5^3 + 3 \times 0.5^3 = 0.5.$$

Fonte: ZABALA; SILVEIRA, 2019, p. 3.

⁸⁸ ZABALA, Filipe J.; SILVEIRA, Fabiano F. *Decades of Jurimetrics*. [arXiv:2001.00476v131]?, 31 dez. 2019. p. 3.

4 MÉTODO

O escopo da ciência é tão amplo e diversificado que, mesmo sem muita pesquisa filosófica, já é de se desconfiar que é quimérica a idéia de um procedimento único, aplicável a todas as áreas. Além disso, está claro para os especialistas que mesmo em domínios mais restritos a investigação científica não é amoldável a nenhum procedimento fixo e explicitável em termos de regras de aplicação automática. A percepção aguda desse ponto levou alguns filósofos contemporâneos a defender a posição extrema de que simplesmente não há nenhum método científico.⁸⁹

O método aplicado na presente pesquisa e projeto foram *(i) Knowledge Discovery in Databases*; *(ii)* empírico e observacional; *(iii)* pesquisa documental em larga escala.

Por larga escala entende-se mais de 6 *Terabytes* (6000 *gigabytes*) de tráfego de internet conforme demonstra extrato do consumo de internet – emitido pelo modem utilizado no decorrer da pesquisa com o objetivo de obter o parâmetro demonstrado.

Incluindo no consumo supracitado estão os downloads da jurisprudência/diários oficiais e documentação de acesso público dos seguintes tribunais/órgãos: Supremo Tribunal Federal – STF, Superior Tribunal de Justiça – STJ, Instituto Nacional de Propriedade Industrial – INPI e Tribunal de Justiça do Distrito Federal e Entornos – TJDFTE – que atualmente compõem o acervo do produto tecnológico elaborado (aproximadamente 700 *Gigabytes*). Não foi incluído no projeto a documentação do TJDFTE dada as restrições temporais e dificuldade dos autores quanto a etapa de normalização dos dados desta fonte. Na documentação extraída foram aplicadas variações do *Método KDD*; adaptadas conforme necessidades singulares de cada fonte de dados alvo.

⁸⁹ CHIBENI. Silvio Seno. *Algumas observações sobre o “método científico*. Disponível em <https://www.unicamp.br/~chibeni/textosdidaticos/metodocientifico.pdf> Acesso em 17 mar. 2021.

Por fim, considera-se que foi aplicado o método empírico. Tal método trata-se daquele desenvolvido diariamente, através do contato com a realidade⁹⁰. De forma subsidiária, tanto os métodos empírico quanto observacional, foram utilizados para identificar necessidades cotidianas do Direito que poderiam ser solucionadas com o auxílio da tecnologia computacional.

⁹⁰ MATTAR, João. **Metodologia científica na Era Digital**. 4.ed. São Paulo: Saraiva, 2017.

5 RESULTADOS

Sabe-se que no contexto internacional existem sistemas que ao nosso ver podem ser chamados de verdadeiras tecnologias jurídicas, dentre as quais mencionam-se o *LexisNexis*, *WestLaw*, *RavelLaw*, *ParallelSearch*, dentre outros. O *Lex Machina*, por exemplo, possibilita a comparação de multa/valor indenizatório de sentenças de acordo com a pretensão do autor, comparar firmas e percentual de ganho de causas – inclusive de acordo com o tipo de causa e cliente, identificar litigantes recorrentes, dentre outros⁹¹. Com o objetivo de verificar o quais produtos e/ou serviços são oferecidos em cenário nacional, procedemos a busca por tecnologias brasileiras capazes de fornecer efetivo amparo à dinamização da prática jurídica.

Os resultados da busca por tecnologias jurídicas foram separados a partir de 3 frentes: Tecnologias Governamentais, Tecnologias Comerciais, Tecnologias Desenvolvidas no Decorrer da Pesquisa.

A primeira, trata-se de ferramentas de iniciativa do setor público, percebe-se que desta frente são ostentadas de forma emocionada, com certa vaidade e overdose todos os seus protótipos – e o extremo apreço pela Inteligência Artificial faz com que toda tecnologia desenvolvida seja anunciada como tal sendo.

A segunda trata-se de ferramentas desenvolvidas por entes privados que se intitulam “*LegalTechs* e *LawTechs*” e comercializam tecnologias, em geral na modalidade de *Software as a Service* – SAAS. O que se percebeu em termos gerais foi a forçosa tentativa de venda de produtos em fase de prototipação, sistemas que funcionam como mero espelhamento/*proxy* de facilidades desenvolvidas e oferecidas

⁹¹ LEX MACHINA. **Legal Analytics Apps**. Disponível em: <https://lexmachina.com/legal-analytics-apps/>. Acesso em: 18 mar. 2021.

pelos tribunais de forma gratuita e a oferta de Sistemas de Gerenciamento de Informação.

Por fim, as Tecnologias Desenvolvidas no Decorrer do Trabalho, agregadas, compõem o produto Tecnológico apresentado ao Programa de Mestrado Profissional. A partir de necessidades percebidas com base no método empírico – do dia a dia profissional, de relatos de necessidades ou ideias/projetos encontrados no decorrer da pesquisa, que se mostraram de desenvolvimento viável considerando os conhecimentos adquiridos, bem como, outros experimentos tecnológicos que se fizeram oportunos durante o decorrer da pesquisa. Uns com potencial, outros nem tanto – pelo menos a curto-médio prazo. Assim como a primeira e segunda iniciativa: protótipos. Se relevantes ou não, cabe a cada leitor uma conclusão – a do autor encontra-se ao final.

5.1 Tecnologias Governamentais

Parece haver certo misticismo social sobre alguns temas tecnológicos – *Blockchain, Machine Learning, Big Data*, e, não menos importante, as duas grandes tendências “O Algoritmo” e Inteligência Artificial.

Cerca de metade dos tribunais brasileiros possui projetos de inteligência artificial operantes ou em desenvolvimento – em sua maioria, a partir do trabalho feito por equipes próprias. É o que aponta o relatório da pesquisa *Tecnologia Aplicada à Gestão dos Conflitos no Âmbito do Poder Judiciário Brasileiro*, produzido pelo Centro de Inovação, Administração e Pesquisa do Judiciário da Fundação Getulio Vargas (CIAPJ/FGV).

A publicação é a primeira etapa de um estudo inédito sobre o uso da inteligência artificial no Judiciário. A amostra da pesquisa inicial abrangeu o Supremo Tribunal Federal (STF), o STJ, o Tribunal Superior do Trabalho (TST), os Tribunais Regionais Federais, os Tribunais Regionais do Trabalho e os Tribunais de Justiça.

O sistema e-Juris é utilizado pela Secretaria de Jurisprudência da corte superior na extração das referências legislativas e

jurisprudenciais do acórdão, além da indicação dos acórdãos principal e sucessivos sobre um mesmo tema jurídico.

Uma quarta ferramenta de inteligência artificial está em fase de desenvolvimento pelo STJ. Quando for implantado, o sistema da Tabela Unificada de Assuntos (TUA) proporcionará a **identificação automatizada do assunto do processo para fins de distribuição às seções da corte**, conforme o ramo do direito em que atuam.⁹²

Aplicação similar à “ferramenta de inteligência artificial está em fase de desenvolvimento pelo STJ” já foi desenvolvida pelos autores no decorrer da presente pesquisa, diferenciando-se no sentido de que a identificação automatizada de assuntos do processo foi realizada para identificar áreas de atuação do advogado e ministros que julgam seus casos.

Conforme relata publicação realizada pelo Supremo Tribunal Federal, em 2018:

Pelo processo judicial eletrônico (PJe), o recurso extraordinário sobe bruto ao Supremo e era preciso que um servidor separasse e identificasse suas peças, tarefa que demandava em média 30 minutos de serviço. O VICTOR realiza essa tarefa em apenas cinco segundos. Toledo esclareceu que mecanismos de inteligência artificial não irão substituir servidores do Poder Judiciário, apenas permitirão que eles atuem em funções mais complexas.⁹³

Em outra publicação, o STF destaca:

Batizado de VICTOR, a ferramenta de inteligência artificial é resultado da iniciativa do Supremo Tribunal Federal, sob a gestão da Ministra Cármen Lúcia, em conhecer e aprofundar a discussão sobre as aplicações de IA no Judiciário. Cuida-se do maior e mais complexo

⁹² SUPERIOR TRIBUNAL DE JUSTIÇA. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito.** [Publicado em 09 de março de 2021]. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx> Acesso em: 17 de março de 2021.

⁹³ SUPREMO TRIBUNAL FEDERAL. **Projeto VICTOR do STF é apresentado em congresso internacional sobre tecnologia.** [Publicado em 26 de setembro de 2018] Disponível em: <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=390818> Acesso em: 17 de março de 2021.

Projeto de IA do Poder Judiciário e, talvez, de toda a Administração Pública Brasileira.

[...]

A UnB colocou na equipe pesquisadores, professores e alunos de alto nível, muitos com formação acadêmica no exterior, de 3 centros de pesquisa de Direito e de Tecnologias. Dentro de pouco tempo teremos publicações sobre o desenvolvimento de VICTOR e as suas perspectivas. Os artigos científicos, que já estão sendo confeccionados, serão publicados nos mais importantes centros de pesquisa do mundo. Tecnologia brasileira incentivada e destacada no mundo.⁹⁴

No que se extrai do portal de notícias do STF, compreende-se que as atividades executadas pelo VICTOR, são:

- a) Conversão de imagens em textos no processo digital;⁹⁵
- b) Separação e identificação das peças processuais;⁹⁶
- c) Identificação dos temas de repercussão geral de maior incidência;⁹⁷

Também é possível listar necessidades tecnológicas do Tribunal, sendo estas:

- d) Um sistema de busca integrada plena entre os entre os sistemas existentes e os sistemas do STF e STJ;⁹⁸
- e) O desenvolvimento de inteligência artificial para a identificação da similaridade de causas entre os recursos interpostos nacionalmente;⁹⁹

⁹⁴ SUPREMO TRIBUNAL FEDERAL. **Inteligência artificial vai agilizar a tramitação de processos no STF.** [Publicado em 30 de maio de 2018]. Disponível em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=380038&ori=1> Acesso em: 17 de março de 2021.

⁹⁵ Idem.

⁹⁶ Idem.

⁹⁷ Idem.

⁹⁸ SUPREMO TRIBUNAL FEDERAL. **STF vai contratar seu primeiro sistema com fundamento na legislação de incentivo à inovação tecnológica.** [Publicado em 12 de setembro de 2019] Disponível em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=423418&ori=1> Acesso em: 17 de março de 2021.

⁹⁹ Idem.

- f) Extração das referências legislativas e jurisprudenciais do acórdão, além da indicação dos acórdãos principal e sucessivos sobre um mesmo tema jurídico (em andamento – Sistema E-Juris);¹⁰⁰
- g) Identificação automatizada do assunto do processo para fins de distribuição às seções da corte, conforme o ramo do direito em que atuam.¹⁰¹

Quanto ao VICTOR, não foram encontradas informações acerca do custo total do projeto, ou novas notícias da evolução de 2019 em diante. Apesar disso, em 2019 já ressalta o projeto de uma nova tecnologia, com orçamento previsto em “R\$ 3,482 milhões para 12 meses de execução do projeto”¹⁰² sob o regime de “contratação direta, com base no artigo 24, inciso XXXI, da Lei 8.666/1993”¹⁰³.

Em se tratando do orçamento do Projeto VICTOR, Becker e Ferrari comentam:

Outro problema sério envolvendo a VICTOR está relacionado à audiência das partes interessadas e seus custos. Ainda que a IA seja do STF e, portanto, seja público, os custos e seu código não são. Ninguém sabe quanto foi dispendido do erário público nem como o código funciona. Falta a transparência necessária pela legislação brasileira. O primeiro problema é uma questão séria de responsabilidade. O último é uma questão de tecnologia.¹⁰⁴

Quanto ao item “a)” compreendemos que se trata de tecnologia de *Optical Character Recognition* – OCR. O OCR é uma funcionalidade/recurso que reconhece

¹⁰⁰ SUPERIOR TRIBUNAL DE JUSTIÇA. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito.** [Publicado em 09 de março de 2021]. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx> Acesso em: 17 de março de 2021.

¹⁰¹ Idem.

¹⁰² SUPREMO TRIBUNAL FEDERAL. **STF vai contratar seu primeiro sistema com fundamento na legislação de incentivo à inovação tecnológica.** [Publicado em 12 de setembro de 2019] Disponível em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=423418&ori=1> Acesso em: 17 de março de 2021.

¹⁰³ Idem.

¹⁰⁴ *Another serious issue involving VICTOR is related to stakeholders' hearing and its costs. Even though the AI belongs to the STF and, therefore, is public, the costs and its code are not. Nobody knows how much it cost for the public purse nor how the code works. It lacks the transparency required by Brazilian Laws. The first problem is a serious issue of accountability. The last one is a technology concern.* BECKER, Daniel; FERRARI, Isabela. **VICTOR, the Brazilian Supreme Court's Artificial Intelligence: a beauty or a beast?** Disponível em: <https://sifocc.org/app/uploads/2020/06/Victor-Beauty-or-the-Beast.pdf> p.9 Acesso em: 13 de maio de 2021.

caracteres em imagens e os transforma em texto, ou seja, ao escanear um arquivo, ao invés de uma imagem como resultado da digitalização, será produzido um documento de texto (normalmente com a extensão .pdf).

Atualmente, existe uma variedade de impressoras, aplicativos de celular, dentre outros dispositivos que possuem a funcionalidade de OCR, e, caso o arquivo esteja salvo no computador, existe uma variedade de conversores com OCR. Alguns destes produtos possuem a funcionalidade de implementação de I.A., cujo objetivo é permitir que a partir de correções humanas, o reconhecimento de caracteres melhore gradualmente.

A funcionalidade de reconhecer caracteres utilizando I.A. surge quando há baixa qualidade na digitalização de documentos – por falta de zelo ao digitalizar ou pela utilização de maquinário ultrapassado.

Destaca-se que para os itens “a)” e “b)” o Supremo Tribunal Federal parece desconhecer ser detentor, desde 2009, de tecnologia que possibilita a conversão de imagens em textos no processo digital e separação e identificação das peças processuais – o Abby. O *Abbyy FineReader v. 15* é uma ferramenta que, inclusive, possui módulos de treinamento orientados à I.A.

Conforme a licitação nº 124/2009 o STF adquiriu, em igual ano, uma licença *Abbyy* – “Aquisição de licença de uso do *software OCR Abbyy Recognition Server 2.0*”¹⁰⁵. Além desta, outras 2 aquisições que envolviam produtos da *Abbyy*”; Licitação nº 76/2011 – “Aquisição de estações de verificação da solução *Abbyy Recognition Sever 3.0*”¹⁰⁶; e Licitação nº 108/2012 – “Contratação de empresa para prestação de

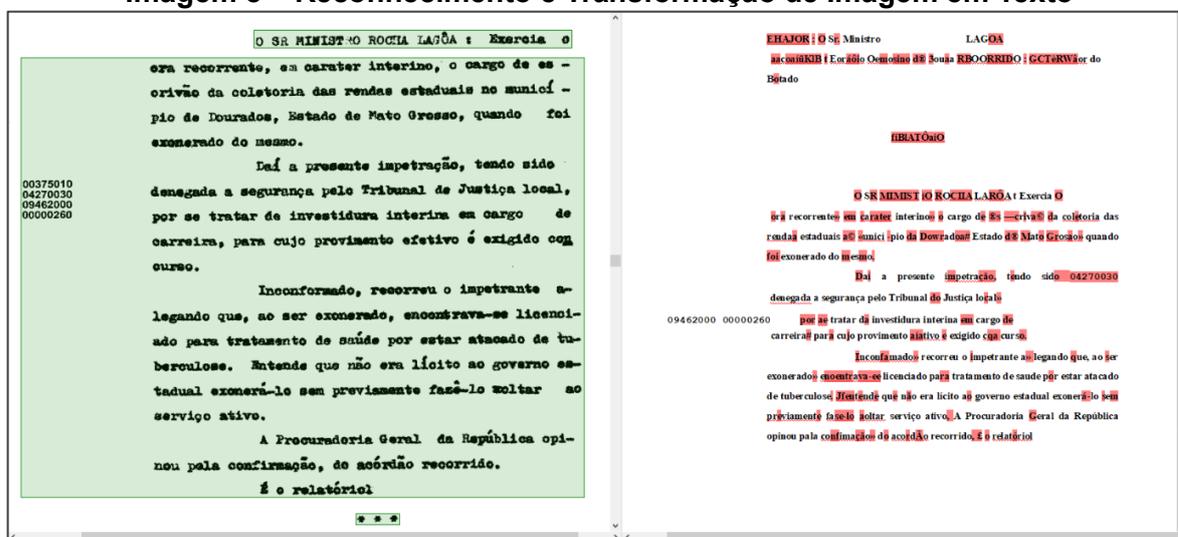
¹⁰⁵ SUPREMO TRIBUNAL FEDERAL. **Licitação 124/2009** [pregão eletrônico]. Processo 336684. Disponível em: <http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=8600>

¹⁰⁶ SUPREMO TRIBUNAL FEDERAL. **Licitação 76/2011** [pregão eletrônico]. Processo 344911. Disponível em: <http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=16201>

serviços de suporte técnico com manutenção corretiva e evolutiva para o *software Abby Recognition Server 3.0*¹⁰⁷.

Em breve análise documental de documentos extraídos do STF verificamos a existência de algumas digitalizações que demandariam a utilização de I.A. para reconhecimento de caracteres, ainda assim, cumpre destacar que o procedimento demanda a interferência humana (*treinamento da máquina*) para a correção dos caracteres. A imagem abaixo, demonstra uma versão diferente (*Finereader Corporate*), contudo do mesmo produto (*Abby*) adquirido pelo STF:

Imagem 8 – Reconhecimento e Transformação de Imagem em Texto

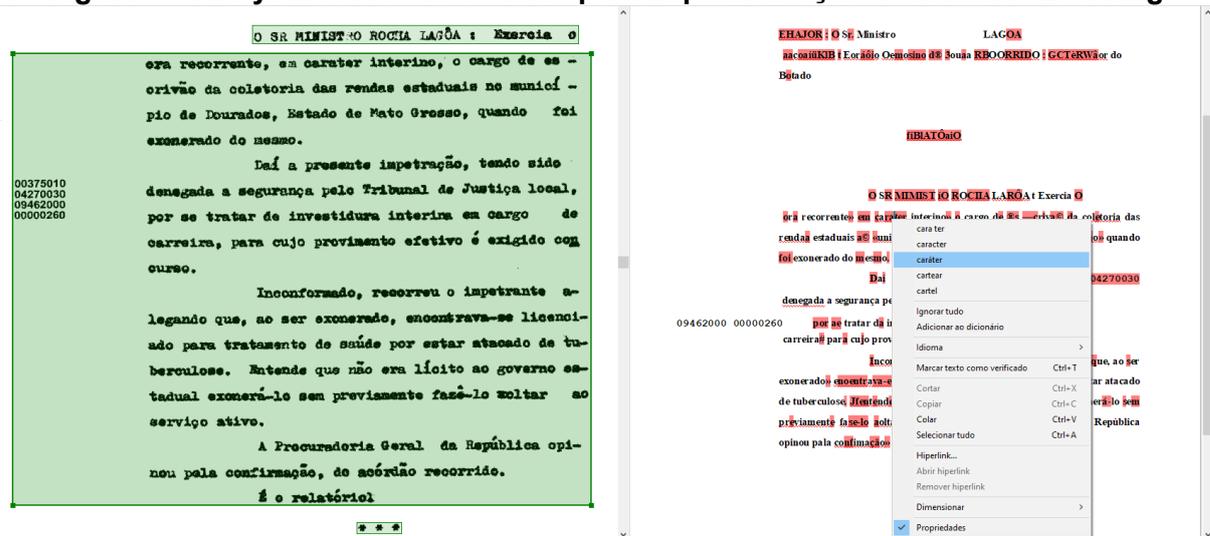


Fonte: Elaborado pelos autores; utilizando o software Abby Fine Reader Corporate¹⁰⁸

¹⁰⁷ SUPREMO TRIBUNAL FEDERAL. **Licitação 108/2012** [pregão eletrônico]. Processo 348788. Disponível em: <http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=21550>

¹⁰⁸ ABBY. **Abby FineReader**. Disponível em: <https://pdf.abbyy.com/finereader-pdf/>. Acesso em: 18 mar. 2021.

Figura 9 – Abby: Interface de Usuário para implementação de Machine Learning



Fonte: Elaborado pelos autores; utilizando o software Abby Fine Reader Corporate¹⁰⁹

Quanto ao Item “c)”, houve dúvidas de qual a tarefa desempenhada pelo VICTOR, pois, não há necessidade de uma inteligência artificial para identificar o que tema de maior incidência em repercussão geral são os de ordem tributária. Há de se presumir que há um erro de compreensão, neste item, acerca do escopo do trabalho realizado pela tecnologia.

A resolução da necessidade descrita no item “d)” foi obstada pela insuficiência de recursos tecnológicos e documentais de tribunais de instâncias inferiores; foi possível extrair, no decorrer da pesquisa, toda (ou grande parte) da documentação disponível em acesso aberto (decisões) dos seguintes tribunais: STF, STJ cerca de 200 mil e 3 milhões de arquivos respectivamente; e TJDF – quantidade não mensurada. Imagens demonstrativas foram anexadas.

¹⁰⁹ *Idem.*

As necessidades apresentadas por ambos os tribunais e indicadas nos itens “e)” e “f)” e “g)” foram solucionadas pelos autores e apresentadas no item 5.3.3 e 5.3.4 respectivamente.

5.2 Tecnologias Comerciais

De modo geral, entende-se que as ferramentas de tecnologia jurídica são oferecidas pelas intituladas “*LawTechs* e *LegalTechs*”. De forma diversa do que é amplamente divulgado compreendemos que o que é oferecido pelas autointituladas “*LawTechs* e *LegalTechs*” não se trata de tecnologia jurídica. Isso pois, tais prestadores de serviço, concentram as soluções oferecidas na classe de Sistemas de Gestão e de Comunicação *Proxy*¹¹⁰.

Em geral ferramentas para acompanhar publicações oficiais que utilizam as credenciais dos advogados para coleta de informação (*Proxy*) e disponibilização dos dados agregados em uma única interface – apesar de serem sistemas simples, facilitam para o advogado o trabalho de acompanhamento processual. Os Sistemas de Gerenciamento de Informação, apesar de serem de suma importância para qualquer que seja o negócio, são, como bem retrata, gerenciais; não se tratando de “tecnologias jurídicas” especificamente.

Sabe-se que o Brasil possui alguns sistemas mais avançados. Por sistemas avançados, entende-se aqueles que constituem ferramentas de busca integrada, funcionam sem relação de dependência com bases públicas para disponibilidade (possuem sistemas de armazenamento próprios) e possibilitam análises diversificadas

¹¹⁰ *Proxy* é um termo do Inglês Britânico utilizado por volta de 1100 DC a 1500 DC, que significa “procuração/procurado” - cujo entendimento é o mesmo que àquele atribuído juridicamente - ou seja, uma pessoa que está autorizada a agir em nome de outra. Recentemente, o termo *Proxy* ganhou significado na computação, denotando um sistema de computador que facilita a troca de dados entre usuários em uma rede, ou seja, uma máquina que age em nome de outra (diferentemente do sentido jurídico, costuma ser utilizado para anonimizar o requerente original).

e em larga escala – promovendo o que no presente trabalho é denominado de *Gestão de Inteligência Jurídica*; dentre os escassos exemplos há de se mencionar a *Neoway* e *Thompson Reuters – Legal One*.

5.3 Tecnologias Desenvolvidas no Decorrer da Pesquisa

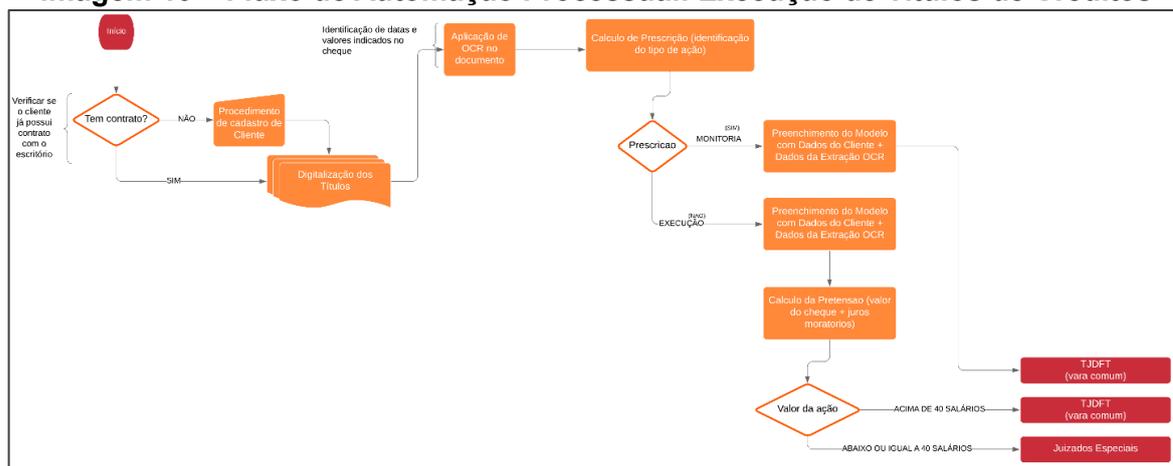
5.3.1 Automação de Peças Processuais

A presente tecnologia surgiu a partir de uma necessidade apresentada no ano de 2020, considera-se que no teste prático realizado, a máquina exerceu em frações de segundos o trabalho que um profissional do Direito possivelmente levaria mais de 30 dias para finalizar.

Após identificarmos a facilidade de automação da petição inicial de um determinado tipo de processo, desenvolvemos um código capaz de produzir cerca de 1.000 petições em 0,85 segundos. A facilidade de automação do documento mencionado se dá pela simplicidade do conteúdo textual, que é repetitivo e raramente necessita de alterações textuais, neste caso, apenas os dados das partes e valor da causa precisam ser substituídos,¹¹¹ portanto, verifica-se que o encargo de tarefas que envolvem procedimentos repetitivos e pouco exercício intelectual são facilmente automatizáveis. Em outras palavras, atividades desvinculadas da necessidade de qualquer tipo de interpretação e análise são ou podem ser facilmente automatizadas.

¹¹¹ Os dados das partes foram extraídos da plataforma de gestão do escritório e o cálculo do valor da causa já inserido no código programado.

Imagem 10 – Fluxo de Automação Processual: Execução de Títulos de Créditos



Fonte: Elaborado pelos Autores.

5.3.2 Análise Preditiva Básica

Por mais que algumas publicações, tais como o Relatório Anual do Conselho Nacional de Justiça, mostrem algumas métricas sob uma perspectiva lato-sensu ou com dados estatísticos, ter a informação em tempo real de quantas demandas similares à de um cliente tramitando no judiciário é uma tarefa mais complexa para os advogados. Tais análises e métricas, por mais que não forneçam a garantia de um resultado, certamente subsidiam resposta para algo que é comumente perguntado aos advogados: “o caso tem chance?”. Quais são as chances de sucesso de uma determinada causa é uma dúvida que em grande parte das vezes não é apenas do cliente. No dia a dia da prática jurídica percebemos que a resposta para tal pergunta é de certo modo, evasiva. De fato, não há como prever o futuro, mas esse tipo de tecnologia pode fornecer uma boa base argumentativa para tal resposta; válido ressaltar que como o próprio nome indica, sugerem uma predileção – ou seja, inclinação/propensão.

Os juristas precisam de uma maneira - uma tecnologia - que o permita fazer perguntas sobre seus dados e receber respostas em um idioma que entendam. Qual advogado nunca foi questionado por um cliente sobre as chances de êxito? Por mais

que seja uma resposta que não deva afirmar a garantia de êxito, o costumeiro "depende" não é uma resposta que satisfaz o cliente; além disso, até mesmo para estabelecer critérios de dificuldade e estimativa de honorários o advogado precisa ter ideia de como foram tratados casos semelhantes e resultados destes casos gerados por decisões anteriores.

Tivemos a oportunidade de apresentar um exemplo prático da utilização de tecnologias preditivas ao sermos questionados acerca das chances de sucesso em um recurso almejando a reforma do despacho de indeferimento do pedido de registro de marca. Uma visão geral de análise extremamente simplória, contudo, relevante diante do contexto – uma apresentação de proposta de serviço.

Para fornecer informação de suporte à tomada de decisão sobre alguns dos caminhos que podem ser tomados – seguir adiante com recurso ou realizar modificações na marca e realizar novo pedido -extraímos cerca de 6000 processos semelhantes ao caso em questão transformamos em um gráfico.

Imagem 11 – Estimativa Recursos contra Indeferimento de Pedidos de Registro de Marcas



Fonte: Elaborado pelos autores.

Forçoso ressaltar que nem todas as bases de dados governamentais proporcionam tal facilidade para extração de informações computacionalmente. Sob nossa percepção cotidiana, os órgãos do judiciário empreendem mais esforços em proteger a sobrecarga do sistema por *bots* do que fortalecer seus sistemas para propiciar acessos como o mencionado. Ao realizar pesquisas em órgão do judiciário não é raro sermos identificados como *bots* – a situação fática, contudo, é que ao invés de um *bot* existe um usuário capaz de realizar ao servidor do órgão múltiplas requisições em pouco tempo. Nesse sentido, os sistemas permitem que um ser humano que queira fazer *download* de 600 acórdãos seja forçado a entrar em página por página, abrir o painel do processo e clicar no botão de *download*.

Em alguns Tribunais o procedimento de *downloads* progressivos deve ser feito lenta e desordenadamente, pois se houver velocidade e/ou um padrão sequencial ao baixar os arquivos¹¹² passa a impor de forma exponencial bloqueadores *captchas* – até que seja impraticável continuar.

Apesar de não ter praticado atos ilícitos, no decorrer da pesquisa fomos identificados como ameaça cibernética algumas vezes, incluindo pelo sistema de proteção da rede e domínio corporativo de nosso escritório.

5.3.3 Identificação de Temas, Partes, Causa de Pedir, Valores, Resultado da Decisão e Outros Elementos

Exemplos e aplicabilidade de *Machine Learning* e Inteligência Artificial ao Direito foi delineado por Sperandio:

Para explicar o conceito do Machine learning, Farina e Buchina apontam como a máquina compreenderia um dado assunto. Utilizando-se de algoritmos de propósito específico, é feita a delimitação dos assuntos relevantes (ou entidades, como descrito em linguagem técnica) que a máquina deverá compreender. Em seguida, tem início o treinamento para a identificação da ocorrência do assunto.

¹¹² Mesmo utilizando a interface gráfica do navegador de internet.

Ao ler a palavra “São Paulo”, por exemplo, como a máquina faz para compreender que se trata de um time de futebol, uma cidade, um estado ou um santo? Se a referência a santo não tiver sido descrita como relevante, a máquina simplesmente não considerará essa possibilidade. Para as outras alternativas, um conjunto de outras situações envolvendo cada uma das entidades (time, cidade, estado) terá sido apresentado anteriormente à máquina que, por comparação obterá a compreensão desejada.

No caso de uma ação judicial, por exemplo, a máquina deverá ser capaz de [i] compreender a comarca, [ii] o juiz e a [iii] causa de pedir. Para identificar cada um dos elementos, não precisará procurar pelo vocábulo específico, mas terá sido treinada a encontrá-lo a partir de um contexto e de um catálogo prévio desses elementos. Assim, cada vez que encontrar uma referência a juiz, aprenderá em que contexto estava tal nome, se estava em conjunto com as expressões meritíssimo, juiz de direito e/ou na abertura do documento etc. A partir de certo ponto, a máquina não precisará nem mesmo do catálogo, porque já terá a compreensão do contexto em que estará a designação do juiz.¹¹³

Pelo que se depreende do texto, o tratamento de dados mencionado pelo autor, seria realizado por ente público – que recebe o pedido ou por ente privado que busca tratar os próprios dados¹¹⁴. Considera-se esforço desnecessário fazer com que a máquina compreenda a comarca – tendo em vista que, (i) tal informação seria indicada pelo sistema do tribunal, ou para arquivos recebidos, selecionada pelo advogado e (ii) o juiz bastaria uma comparação de texto com uma lista de nome de juízes do tribunal; a causa de pedir (iii) trata-se de uma aplicação promissora, apesar de ser de difícil desenvolvimento devido a variedade de formas que pode ser escrita.

¹¹³ SPERANDIO, Henrique Raimundo do Carmo. Desafios da Inteligência Artificial para a Profissão Jurídica. 107 p. Dissertação de Mestrado – Fundação Getulio Vargas (FGV). 2018. P. 28-29.

¹¹⁴ A segunda opção não parece ser o objetivo considerando o que foi descrito no texto.

Imagem 12 – Identificação de Elementos Jurídicos por Linguagem de Marcação

```

36 <court-comprehension>
37 1. "1. <p>Esta Corte e o STF posicionaram-se quanto às fitas de vídeo e aos <object>programas de computadores</object>, diante dos itens 22
38 e 24 da Lista de Serviços</p>.
39 2. <p>Os <object>programas de computador</object> desenvolvidos para clientes, de <description>forma personalizada</description>,
40 <comprehension>geram incidência de tributo do ISS</comprehension></p>.
41 3. <p>Diferentemente, se o <object-syn>programa</object-syn> é criado e <description>vendido de forma impessoal para clientes</description>
42 que os <description>compra como uma mercadoria qualquer</description>, esta venda é <comprehension>gravada com o <area-var-taxlaw>ICMS/</area-var-taxlaw></comprehension>. "Consecutivamente, como no caso sub examine, as operações envolvendo a exploração econômica de
<object>programas de computador</object>, <description>quando feitos em larga escala e de modo uniforme, são consideradas operações de
compra e venda</description>, <description>sujeitando-se, consecutivamente, à tributação pelo ICMS</description> (<ref-null-resp>RESP 123.
022-RS</ref-null-resp>, DJ de 27.10.1997, Rel. Min. José Delgado; <ref-null-resp>RESP 216.967-SP</ref-null-resp>, DJ de 22.04.2002, Rel.
Min. Eliana Calmon; ROMS 5.934-RJ, DJ de 01.04.1996, Rel. Min. Hélio Mosimann)</p>.
42 </court-comprehension>

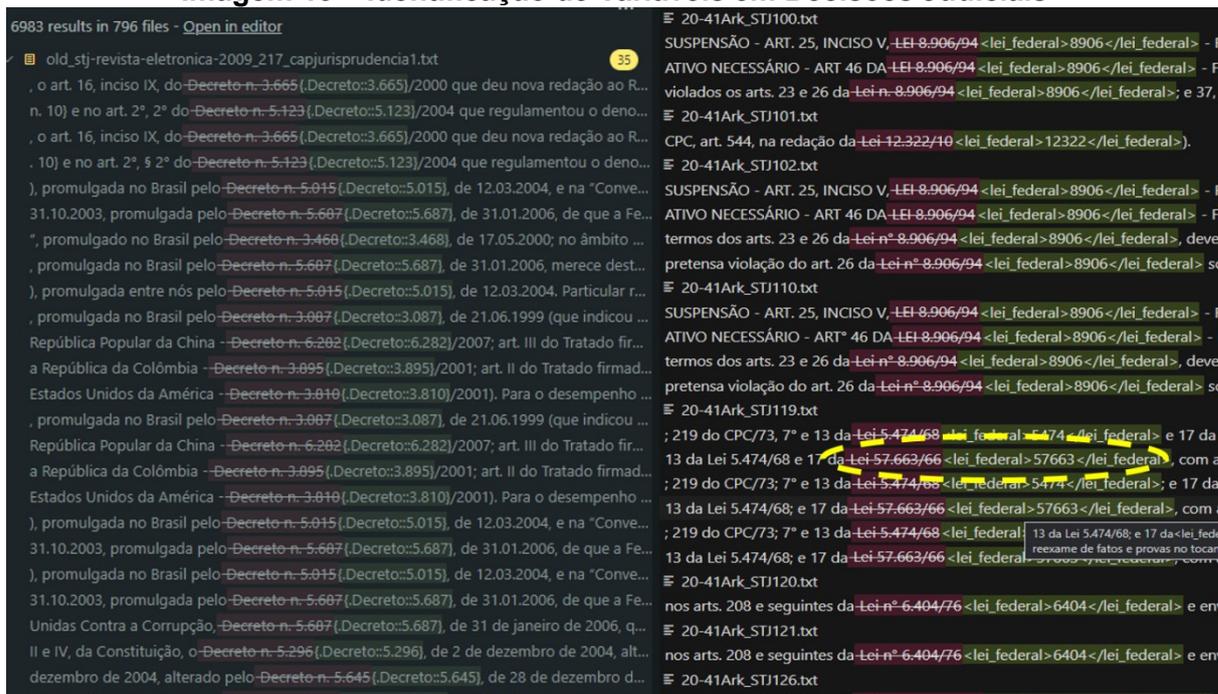
```

Fonte: Elaborado pelos autores.

Conforme demonstra a figura acima, não é necessário a utilização de algoritmos para que se produza a delimitação de assuntos relevantes, sendo necessário tão somente, a indexação de termos desejados e *scripts* de manipulação textual que injetam marcações quando encontram o termo procurado. O mesmo procedimento, foi utilizado para satisfação das necessidades de entes governamentais – item “e)” apresentado no capítulo de Tecnologias Governamentais – “extração das referências legislativas e jurisprudenciais do acórdão”.

As marcações acima demonstradas compõem uma linguagem desenvolvida pelos autores – sendo uma variação das linguagens *XML* e *XHTML* criada e orientada à identificação de elementos do setor jurídico (leis e normas, decisões, referências bibliográficas, objeto da lide, compreensão da corte, entidades governamentais etc.).

Imagem 13 – Identificação de Variáveis em Decisões Judiciais



Fonte: Elaborado pelos autores

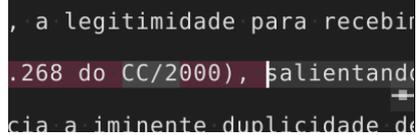
Acerca da Imagem XX, alguns detalhes carecem de atenção: (i) apesar de todos os documentos pertencerem ao mesmo tribunal (Superior Tribunal de Justiça), há falta de padronização¹¹⁵ – ainda que exista um manual de normalização de escrita; exemplo: **(i)** LEI 1.234/56; **(ii)** Lei 1.234/56; **(iii)** Lei n. 1.234/56; **(iv)** Lei nº 1.234/56. Por mais que para o ser humano seja intuitivo representem o mesmo objeto, a máquina não é capaz de tal compreensão (inicialmente), ou seja, caso não seja identificada a variação, leis serão ignoradas.

Acima, foram enumeradas quatro variações; no decorrer do projeto nos deparamos com cerca de 1250 (mil duzentos e cinquenta) formas diferentes de mencionar uma única lei. Considerando a estimativa de 14.000 (quatorze mil) leis federais, são cerca de 17.500.000 (dezessete milhões e quinhentas mil) variações

¹¹⁵ Compreensível e esperado, considerando o volume documental analisado e que a execução desse trabalho seja feita por ser humano.

apenas na forma de mencionar uma lei que, em tese, podem aparecer em uma análise desta ordem. Ressalta-se que não compõem o cálculo apresentado, os casos em que houver algum erro de digitação no número da lei como demonstra o item circulado na Imagem XX e nas imagens abaixo expostas – situação gera um número “incalculável”¹¹⁶ de variações possíveis.

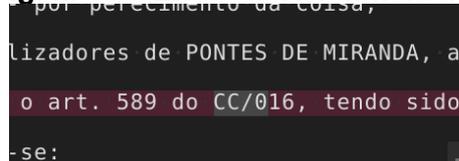
Imagem 14 – Erros de Referência à Lei



, a legitimidade para receber
.268 do CC/2000), salientando
cia a iminente duplicidade da

Fonte: Elaborado pelos autores.

Imagem 15 – Erros de Referência à Lei



por perecimento da coisa,
lizadores de PONTES DE MIRANDA, a
o art. 589 do CC/016, tendo sido
-se:

Fonte: Elaborado pelos autores.

Imagem 16 – Erros de Referência à Lei

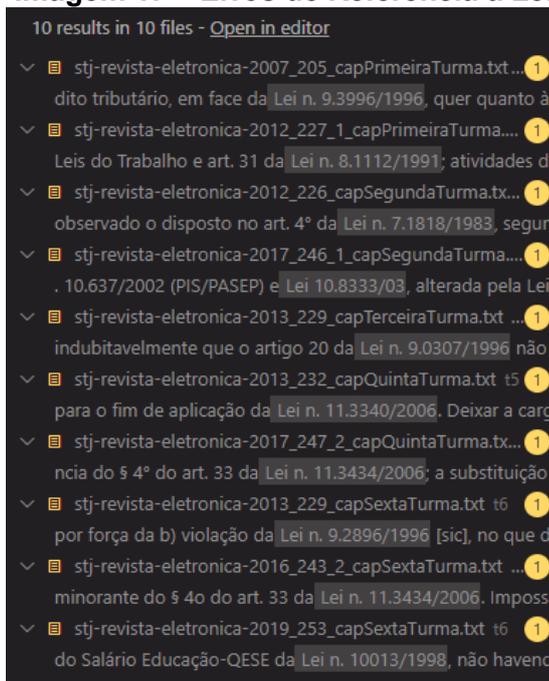


DJ 02.03.1998);
do Código de Processo Civil de 1916. Pr

Fonte: Elaborado pelos autores.

¹¹⁶ Sob a perspectiva de praticidade e necessidade de calcular tal informação.

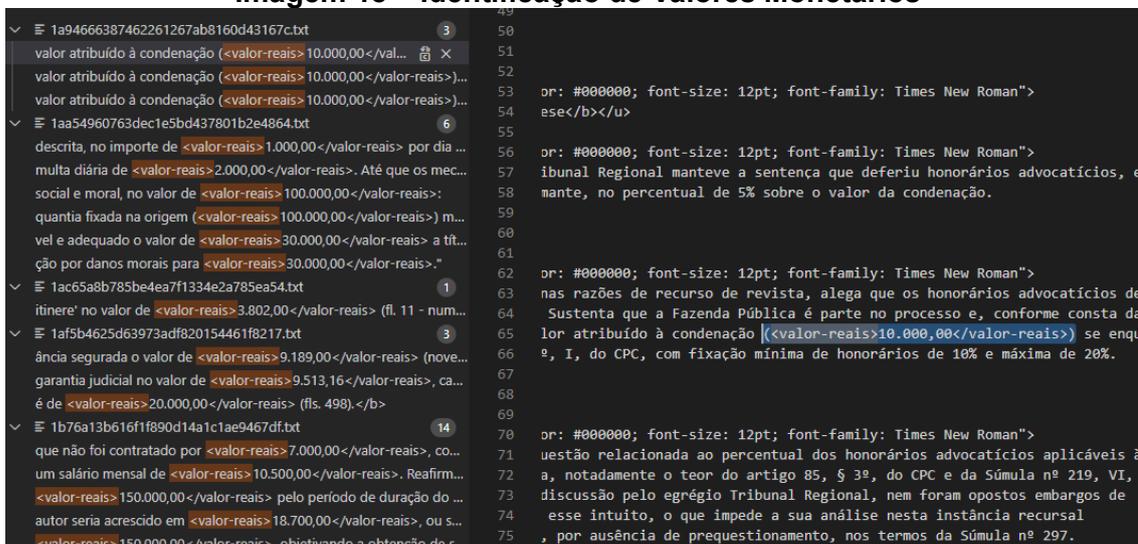
Imagem 17 – Erros de Referência à Lei



Fonte: Elaborado pelos autores.

Dentre outros elementos identificados de forma automatizada:

Imagem 18 – Identificação de Valores Monetários



Fonte: Elaborado pelos autores.

Imagem 19 – Identificação de Data do Julgamento

250 results in 1 file - Open in editor	3570	enriquecimento sem causa vedado pelo ordenamento pátrio.
2020-05-02_T3.txt CAUsers\lagoLora... 250	3571	5. Recurso especial provido.
Brasília (DF), 14 de maio de 2019 (data do jul...	3572	<BODY-AGREEMENT>
Brasília (DF), 10 de setembro de 2019 (data do jul...	3573	Prosseguindo no julgamento, após o voto-vista do Sr. Ministro Marco
Brasília (DF), 04 de junho de 2019 (data do jul...	3574	Aurélio Belizze, vistos, relatados e discutidos os autos em que são partes as
Brasília (DF), 21 de maio de 2019 (data do jul...	3575	acima indicadas, acordam os Senhores Ministros da Terceira Turma do Superior
Brasília (DF), 02 de abril de 2019 (data do jul...	3576	Tribunal de Justiça, por maioria, em dar provimento ao recurso especial, nos
Brasília (DF), 07 de maio de 2019 (data do jul...	3577	termos do voto do Sr. Ministro Moura Ribeiro, que lavrará o acórdão.
Brasília (DF), 27 de agosto de 2019 (data do jul...	3578	Vencidos os Srs. Ministros Ricardo Villas Bôas Cueva e Nancy Andrighi.
Brasília (DF), 14 de maio de 2019 (data do jul...	3579	Votaram com o Sr. Ministro Moura Ribeiro os Srs. Ministros Paulo de Tarso
Brasília (DF), 21 de maio de 2019 (data do jul...	3580	Sanseverino e Marco Aurélio Bellizze.
Brasília (DF), 25 de junho de 2019 (data do jul...	3581	Brasília (DF), 07 de maio de 2019 (data do julgamento).
Brasília (DF), 12 de fevereiro de 2019 (data d...	3582	Ministria Moura Ribeiro, Relator
Brasília (DF), 09 de abril de 2019 (data do jul...	3583	DJe 10.6.2019
Brasília (DF), 05 de fevereiro de 2019 (data d...	3584	RELATÓRIO
Brasília (DF), 23 de abril de 2019 (data do jul...	3585	O Sr. Ministro Ricardo Villas Bôas Cueva: Trata-se de recurso especial
Brasília (DF), 27 de novembro de 2018 (data ...	3586	interposto por Paulo Braganti Camilo e Outros, com fundamento no art. 105,
Brasília (DF), 06 de outubro de 2018 (data d...	3587	III, "a" e "c", da Constituição Federal, contra acórdão proferido pelo Tribunal de
Brasília (DF), 23 de outubro de 2018 (data d...	3588	Justiça do Estado de São Paulo assim ementado:
Brasília (DF), 04 de dezembro de 2... X	3589	Execução por quantia certa. Título extrajudicial (alugueres e encargos de
Brasília (DF), 23 de outubro de 2018 (data d...	3590	locação não residencial escrito). Embargos do devedor. Juízo de procedência.
Brasília (DF), 13 de novembro de 2018 (data ...	3591	Apelo do exequente, embargado. Parcial provimento (fl. 269 e-STJ).
	3592	Os embargos de declaração foram rejeitados (fls. 293-296 e-STJ).
	3593	Nas razões recursais (fls. 299-313 e-STJ), os recorrentes apontam violação
	3594	dos arts. 393 e 567 do <LEG-FED>10406</LEG-FED> e 26 da Lei n. 8.245/1991, além do
	3595	dissídio jurisprudencial.

Fonte: Elaborado pelos autores.

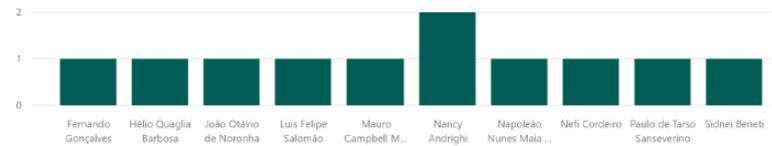
Quanto às causas de pedir, identificamos a descrição de atributos do objeto ou situações fáticas que ensejam a lide – essa etapa em específico, ainda que tenha sido acelerada por alguns métodos desenvolvidos, consiste em procedimento que requer maior empenho humano do que computacional. A partir de uma amostra de decisões marcadas, foram gerados gráficos para demonstração exemplificativa de resultados e informações que podem ser obtidos.

Gráfico 1 – Análise: Decisão Unânime/Parcial

Qtd Processos por Voto_Parcial

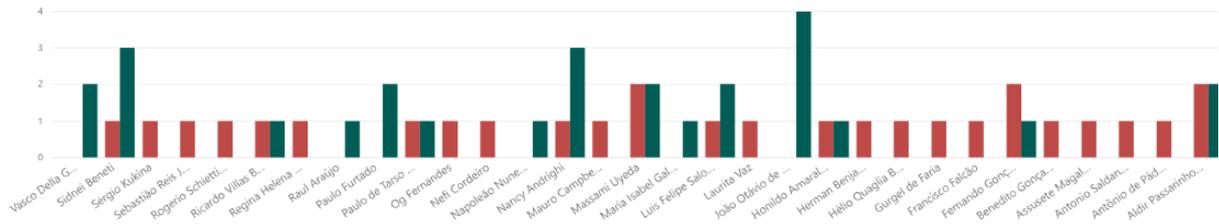


Qtd Processos por Desc Relator



Qtd Processos por Ministros e Voto_Ministro

Voto ● Desprovido ● Provido



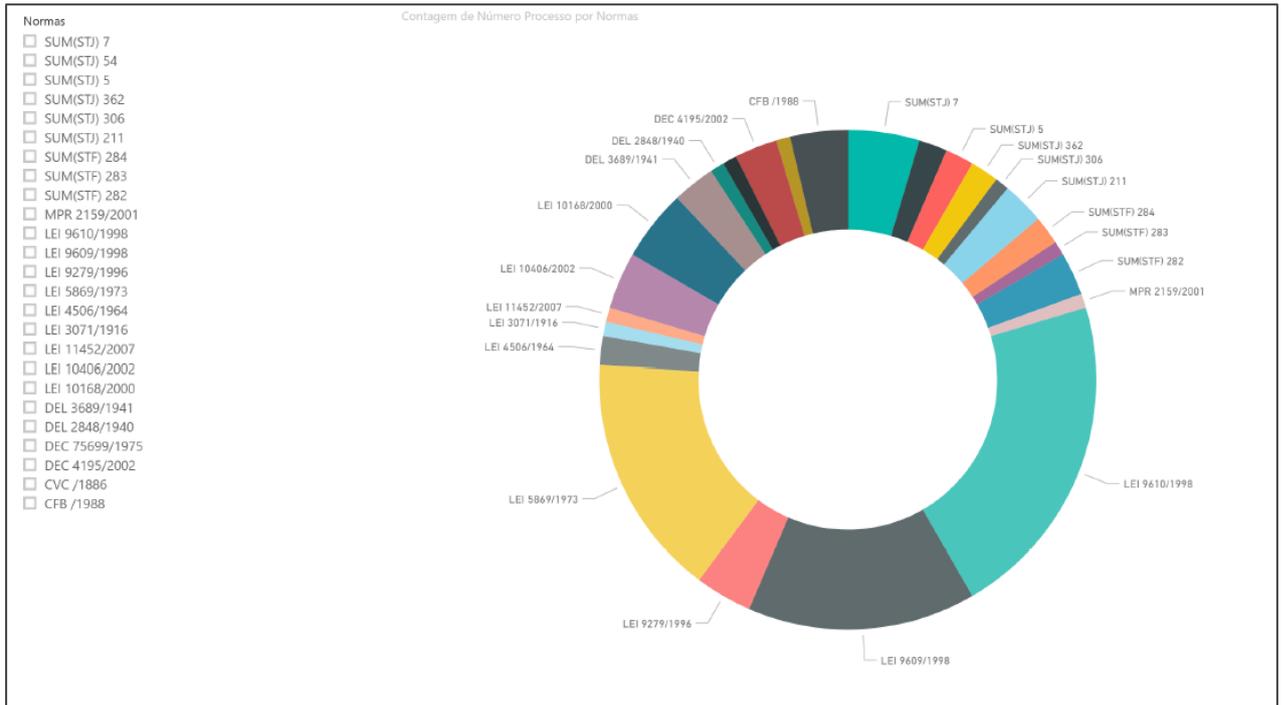
- Número Proc...
 - 1779215
 - 1642249
 - 1641775
 - 1403865
 - 1185943
 - 1136676
 - 1127220
 - 1114889
 - 991721
 - 913008
- Ministros
 - Aldir Passarinho Junior
 - Antônio de Pádua Ribeiro
 - Antonio Saldanha Palhei...
 - Assusete Magalhães
 - Benedito Gonçalves
 - Fernando Gonçalves
 - Francisco Falcão
 - Gurgel de Faria
 - Hélio Quagjlia Barbosa
 - Herman Benjamin
- Turma
 - T1 - PRIMEIRA TURMA
 - T2 - SEGUNDA TURMA
 - T3 - TERCEIRA TURMA
 - T4 - QUARTA TURMA
 - T6 - SEXTA TURMA

Ano	Número Processo	Ministros	Relator	Turma	Decisão
2009	1127220	Vasco Della Giustina	0	T3 - TERCEIRA TURMA	Provido
2009	1136676	Vasco Della Giustina	0	T3 - TERCEIRA TURMA	Provido
2009	1114889	Sidnei Beneti	0	T3 - TERCEIRA TURMA	Desprovido
2009	1127220	Sidnei Beneti	1	T3 - TERCEIRA TURMA	Provido
2009	1136676	Sidnei Beneti	0	T3 - TERCEIRA TURMA	Provido
2013	1403865	Sidnei Beneti	0	T3 - TERCEIRA TURMA	Provido
2016	1641775	Sergio Kukina	0	T1 - PRIMEIRA TURMA	Desprovido
2018	1779215	Sebastião Reis Júnior	0	T6 - SEXTA TURMA	Desprovido

Fonte: Elaborado pelos Autores¹¹⁷

¹¹⁷ Projeto – *Protótipo de Legal Intelligence* criado em conjunto com ARGOLO, Leonardo; ROCHA, Renato. MARSAN IT [empresa em parceria] Disponível em <https://marsanit.com.br/> Acesso em: 15 março 2021.

Gráfico 2 – Análise de Normas Mencionadas



Fonte: Elaborado pelos autores¹¹⁸

Gráfico 3 – Dashboard Geral



Fonte: Elaborado pelos autores¹¹⁹

¹¹⁸ Projeto – *Protótipo de Legal Intelligence* criado em conjunto com ARGOLO, Leonardo; ROCHA, Renato. MARSAN IT [empresa em parceria] Disponível em <https://marsanit.com.br/> Acesso em: 15 marco 2021.

¹¹⁹ Projeto – *Protótipo de Legal Intelligence* criado em conjunto com ARGOLO, Leonardo; ROCHA, Renato. MARSAN IT [empresa em parceria] Disponível em <https://marsanit.com.br/> Acesso em: 15 marco 2021.

Gráfico 4 – Análise de Referência Normativa



Fonte: Elaborado pelos autores¹²⁰

5.3.4 Ferramenta para Identificação de Similaridades de Decisões

Como demonstra-se na imagem a seguir, a ferramenta identifica diferenças entre duas decisões – no presente caso exposto, foram comparadas documentações do Painel Arbitral da Câmara de Disputa em Nome de Domínio na Internet.

¹²⁰ Projeto – *Protótipo de Legal Intelligence* criado em conjunto com ARGOLO, Leonardo; ROCHA, Renato. MARSAN IT [empresa em parceria] Disponível em <https://marsanit.com.br/> Acesso em: 15 março 2021.

Imagem 20 – Verificação de Diferenças Textuais

WIPO

ARBITRATION

AND

MEDIATION CENTER

WORLD

INTELLECTUAL PROPERTY

ORGANIZATION

WIPO

ARBITRATION

AND

MEDIATION CENTER

WORLD

INTELLECTUAL PROPERTY

ORGANIZATION

DECISÃO DO PAINEL ADMINISTRATIVO [Rhodia Services v. Emerson Fortunato Maia](#) Caso No. DBR2011-0001

1. As Partes

O Reclamante é Rhodia Services, Aubervilliers, França, representado por Nameshield, França.

O Reclamado é Emerson Fortunato Maia, Piracicaba, São Paulo, Brasil.

2. O Nome de Domínio

O nome de domínio em disputa é <proban.net.br>, o qual está registrado com a NIC.br.

3. Histórico do Procedimento

A Reclamação foi apresentada no Centro de Arbitragem e Mediação da OMPi (o "Centro") em 9 de fevereiro de 2011. Em 10 de fevereiro de 2011, o Centro transmitiu por e-mail para o NIC.br o pedido de verificação de registro em conexão com o nome de domínio em disputa. Na mesma data, o NIC.br transmitiu por e-mail para o Centro a sua resposta de verificação, confirmando que o Reclamado está listado como o titular do registro do domínio em disputa e fornecendo os detalhes de contato.

O Centro verificou que a Reclamação satisfaz os requisitos formais do Regulamento do Sistema Administrativo de Conflitos de Internet Relativos a Nomes de Domínios sob ".Br", SACI-Adm (o "Regulamento") e das Regras do Centro de Arbitragem e Mediação da OMPi para o SACI-Adm (as "Regras").

De acordo com as Regras, art. 3, o Centro notificou formalmente o Reclamado da Reclamação, e o processo teve início em 18 de fevereiro de 2011. Com base no art. 7(a) das Regras, a data limite para resposta era 9 de março de 2011.

O Reclamado não apresentou resposta. Assim, de acordo com o Art. 7(e) das Regras, o Centro notificou a revelia do Reclamado em 11 de março de 2011.

O Centro nomeou Rodrigo Azevedo como Especialista em 17 de março de 2011 para decidir o caso. O Especialista considera que o Painel Administrativo foi devidamente constituído. O Especialista apresentou a Declaração de Acolhimento e a Declaração de Imparcialidade e Independência, tal como exigido pelo Centro para assegurar o cumprimento das Regras, Arts. 4 e 5.

DECISÃO DO PAINEL ADMINISTRATIVO

[General Nutrition Investment Company v. Silvania de Souza Silva](#)

Caso No. DBR2011-0002

1. As Partes

A Reclamante é General Nutrition Investment Company, de Phoenix, Estados Unidos da América, representado por Momen, Leonardos & Cia, do Rio de Janeiro, Brasil.

A Reclamada é Silvania de Souza Silva, de Fortaleza, Ceará, Brasil, sem representante autorizado.

2. Os Nomes de Domínio e a Unidade de Registro

Os nomes de domínio em disputa são <gmosuplementos.com.br>, <gmvitaminas.com.br> e <gnovitaminas.com.br>, os quais estão registrados com o NIC.br.

3. Histórico do Procedimento

A Reclamação foi distribuída ao Centro de Arbitragem e Mediação (o "Centro") em 12 de agosto de 2011. Em 15 de agosto de 2011, o Centro transmitiu por e-mail para o NIC.br o pedido de verificação de registro em conexão com os nomes de domínio em disputa. Em 15 de agosto de 2011, o NIC.br transmitiu por e-mail para o Centro a resposta de verificação dos nomes de domínio em disputa, confirmando que a Reclamada é a titular dos registros e fornecendo os respectivos dados de contato. Em 16 de agosto de 2011, o Centro notificou a Reclamante acerca de irregularidade formal diagnosticada na Reclamação. Ato contínuo, em 18 de agosto de 2011, a Reclamante enviou ao Centro a Reclamação retificada.

O Centro verificou que a Reclamação e respectiva retificação satisfazem os requisitos formais do Regulamento do Sistema Administrativo de Conflitos de Internet Relativos a Nomes de Domínios sob ".Br" - Denominado SACI-Adm (o "Regulamento") e das Regras do Centro de Arbitragem e Mediação da OMPi para

o SACI-Adm (as "Regras").

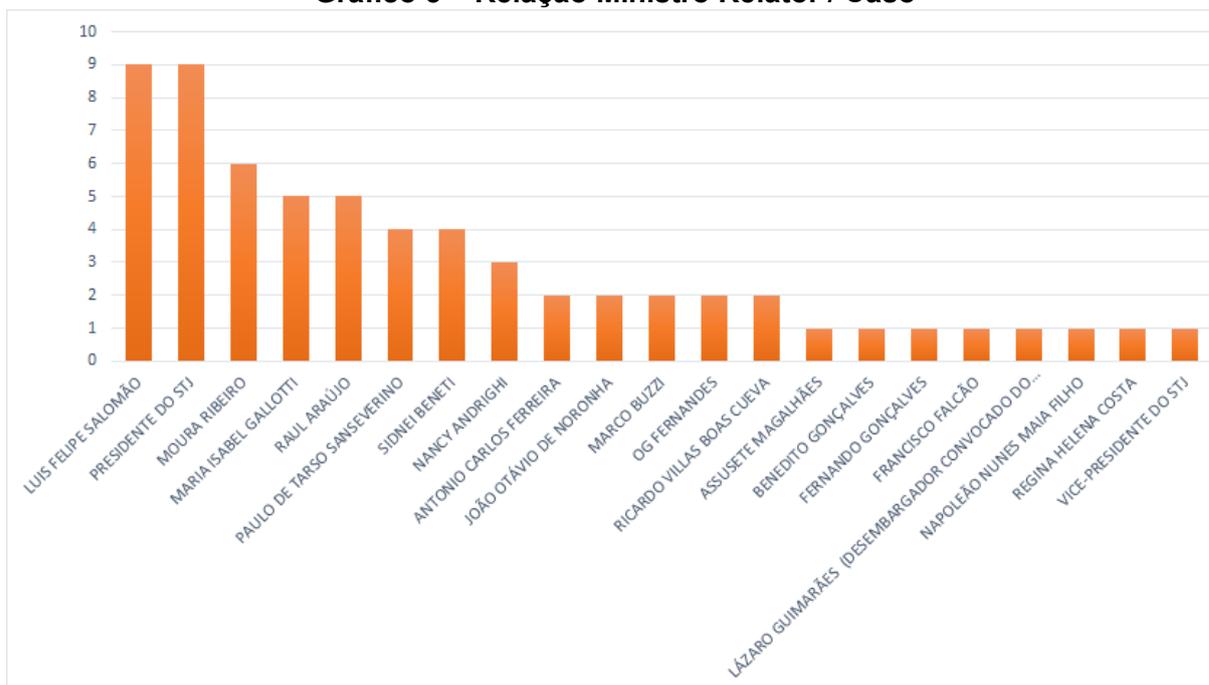
De acordo com as Regras, art. 3, o Centro formalizou a notificação da Reclamação e o procedimento administrativo iniciou em 19 de agosto de 2011. De acordo com as Regras, art. 7(a), a data limite para o envio da defesa findou em 11 de setembro de 2011. O Centro recebeu a Defesa da Reclamada em 10 de setembro de 2011.

Fonte: Elaborado pelos Autores.

5.3.5 Ferramenta de Verificação de Áreas e Ministros

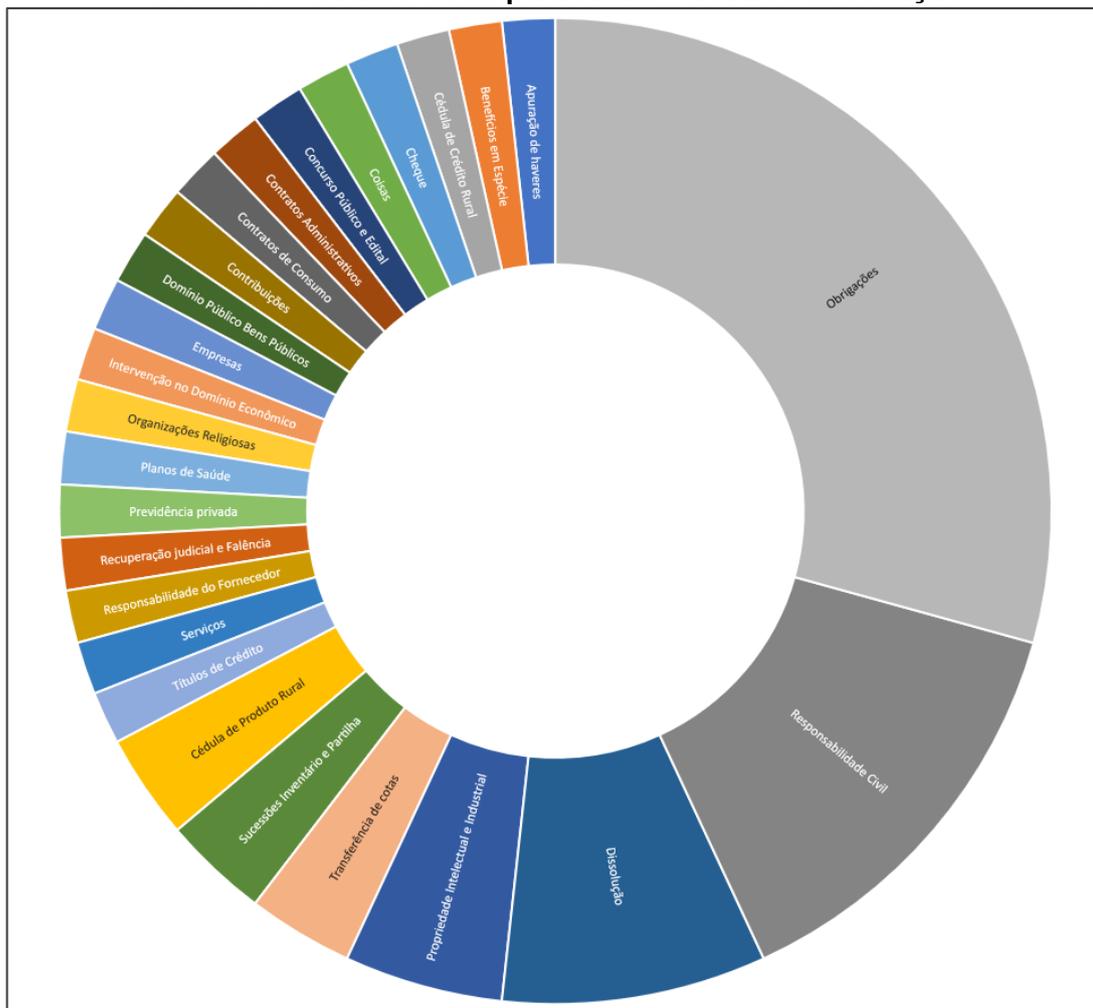
A presente ferramenta, foi desenvolvida com o intuito de identificar temas de atuação do advogado e ministros que mais julgam seus casos. Os gráficos abaixo, demonstram os resultados obtidos.

Gráfico 5 – Relação Ministro Relator / Caso



Fonte: Elaborado pelos autores.

Gráfico 6 – Gráfico Solar para Análise de Área de Atuação

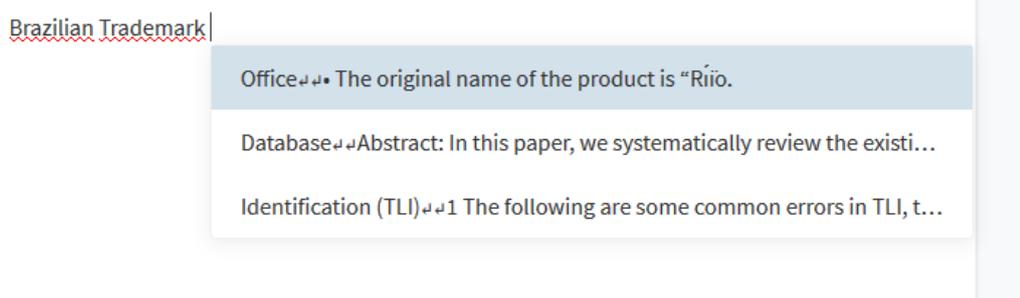


Fonte: Elaborado pelos Autores

5.3.6 Inteligência Artificial para Produção de Textos

Em vários testes realizados pelos autores, utilizando dois modelos diferentes de Inteligência Artificial – o GPT-2¹²¹ e o Arvix-NPL¹²² – com o objetivo de que produzissem conteúdo jurídico, obteve-se os seguintes resultados:

Imagem 21 – Produção de Texto por Inteligência Artificial



Fonte: Elaborado pelo autor com base nas ferramentas GPL-2 e Arvix-NPL de Huggingface Team

¹²¹ *The almighty king of text generation, GPT-2 comes in four available sizes, only three of which have been publicly made available. Feared for its fake news generation capabilities, it currently stands as the most syntactically coherent model.* [O rei todo-poderoso da geração de texto, GPT-2 vem em quatro tamanhos disponíveis, apenas três dos quais foram disponibilizados publicamente. Temido por seus recursos de geração de notícias falsas, ele atualmente se destaca como o modelo mais sintaticamente coerente.] HUGGINFACE TEAM, The. **Transformer**. Disponível em: <https://transformer.huggingface.co>

¹²² *Built on the OpenAI GPT-2 model, the Hugging Face team has fine-tuned the small version on a tiny dataset (60MB of text) of Arxiv papers. The targeted subject is Natural Language Processing, resulting in a very Linguistics/Deep Learning oriented generation.* [Construído no modelo OpenAI GPT-2, a equipe do Hugging Face ajustou a versão pequena em um pequeno conjunto de dados (60 MB de texto) de papéis Arxiv. O assunto alvo é Processamento de Linguagem Natural, resultando em uma geração muito orientada para Linguística / Aprendizado Profundo.] HUGGINFACE TEAM, The. **Transformer**. Disponível em: <https://transformer.huggingface.co>

Imagem 22 – Produção de Texto por Inteligência Artificial

The brazilian intellectual property law |

was first implemented by the European parliament in 1998, after a fiv...

, which was written and organized by the International Law Office, is a...

was founded in the early 20th century.

Fonte: Elaborado pelo autor com base nas ferramentas GPL-2 e Arvix-NPL de Huggingface Team

Imagem 23 – Produção de Texto por Inteligência Artificial

The brazilian trademark law was implemented by the Portuguese government, and the Portuguese patent law was implemented by the Spanish government. This is obviously |

a very complex law, but it is the first law that has been implemented i...

the first time that the Portuguese government has implemented a pat...

the reason why the Portuguese patent law was not implemented.

Fonte: Elaborado pelo autor; [utilizando o GPL-2 e Arvix-NPL de Huggingface Team]

Tradução: “A Lei de Marcas Brasileira foi implementada pelo Governo Português, e, a Lei de Patentes de Portugal foi implementada pelo governo Espanhol. [...]”

Inserida a abertura do parágrafo “*Brazilian Trademark Law*”; incentivou-se que a máquina continuasse a escrever a partir do próprio *output*, ou seja, utilizar o que escrevia como *input* para o próximo trecho ou frase.

Imagem 24 – Produção de Texto por Inteligência Artificial

Brazilian Trademark Law The "Monsanto Law" is often referred to as a "monopolistic law" in Brazil because it regulates the business practices of the seed companies that produce genetically engineered (GE) seed and plant material for industrial and agricultural use. In many countries, this law restricts the commercial activities of commercial seed companies that produce and distribute genetically engineered seeds.

Fonte: Elaborado pelo autor; [utilizando o GPL-2 e Arvix-NPL de Huggingface Team]

Após a remoção de trechos redundantes e adaptações linguísticas, obteve-se: “a Lei de Marcas Brasileira – ou “Lei Monsanto”, regula a prática de comercialização de sementes e plantas geneticamente modificadas, costuma ser referida como uma ‘lei monopolística”.

6 CONCLUSÕES

Quanto à aplicabilidade e replicabilidade podem ser encontradas no método de *Knowledge Discovery in Databases* – KDD. O KDD, por ser um método amplo e adaptável pode ser empregado nas mais diversas áreas. A aplicação do KDD ao Direito já foi utilizada pelos autores em pesquisa anterior, ainda assim, de forma geral nos parece ser algo não usual na área – apesar de útil para pesquisas jurimétricas.

Conforme afirma a origem da informação e detentora da Tecnologia, o “maior a mais complexo Projeto de IA do Poder Judiciário e, talvez, de toda a Administração Pública Brasileira”¹²³ é um OCR. Sendo o OCR uma funcionalidade que pode ser desempenhada por qualquer *smartphone* relativamente moderno, a Administração Pública Brasileira encontra-se em preocupante atraso tecnológico.

Grande parte da tecnologia jurídica que é comercializada por entidades privadas, são espelhamentos de serviços já oferecidos gratuitamente pelos tribunais, “espelhamentos” uns dos outros (alterando-se basicamente a interface de usuário). Tais tecnologias, em geral são comercializadas na modalidade de *Software As A Service* – SAAS, com mensalidades que variam entre 3 e 4 dígitos financeiros, apesar disso, boa parte do que é ofertado pode ser executado por uma planilha de Excel – vendida por R\$ 150,00 (cento e cinquenta reais) em sites de compra e venda.

Vista disso, se aplicados os critérios de avaliação do Mestrado Profissional em Propriedade Intelectual e Transferência de Tecnologia para a Inovação – PROFNIT, entendemos que grande parte do que hoje existe de Tecnologia Jurídica, sob uma análise rigorosa seria classificada como produção de *Baixo Teor Inovativo*. Isso pois, basicamente existe a adaptação de conhecimentos existentes – em geral

¹²³ SUPERIOR TRIBUNAL DE JUSTIÇA. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito.** [Publicado em 09 de março de 2021]. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx> Acesso em: 17 de março de 2021.

de Administração/Engenharia de Produção/Ciência da Computação, sendo de uma ou outra apartadamente. Nos casos em que as tecnologias jurídicas agregam diferentes áreas de conhecimento (Administração, Direito, Computação, Biblioteconomia, etc) eleva-se à classificação para *Médio Teor Inovativo*.

A documentação jurídica obtida durante a pesquisa, resultou em aproximadamente 5.600.000 decisões de tribunais diversos. Fora a importância atribuída quando se trata de normalização dos dados e tomada de decisão, evidenciamos o valor financeiro que é atribuído através de duas rápidas pesquisas de mercado. Dada a cláusula de sigilo e conseqüente inviabilidade de referenciar a fonte, foram adicionados apenas os valores finais – considerando o volume documental supracitado – a título de complementação do presente projeto.

- a) Fornecedor A – Aquisição Documental: 5.600.000 decisões (documentos em acesso aberto) – Orçamento: R\$ 8.400.000,00 (oito milhões e quatrocentos mil reais).
- b) Fornecedor B – Aquisição Documental: Valor por Tribunal – Orçamento: R\$ 3.000.000,00 (três milhões de reais).

O presente projeto manteve um sistema integrado de computadores (*cluster*), que somados resultam em 26 processadores lógicos e 104 *gigabytes* de RAM tratando dados jurídicos incessantemente.¹²⁴ Durante a consolidação presente trabalho consideramos ter atingido patamares corporativos no que tange à aplicação de métodos orientados ao desenvolvimento de soluções de análise jurídica.

¹²⁴ Para atribuir parâmetros comparativos, em rápida coleta de dados identificamos que as máquinas utilizadas por advogados possuem em média 4 processadores lógicos e entre 4 e 8 gigabytes de RAM.

7 REFERÊNCIAS

ABBY. **Abby FineReader**. Disponível em: <https://pdf.abbyy.com/finereader-pdf/>. Acesso em: 18 mar. 2021.

ABNT. **Normalização**. <http://www.abnt.org.br/normalizacao/o-que-e/o-que-e> Acesso em: 17 mar. 2021.

BACON, Francis. **Meditationes sacrae**. [1603/4]. Stuttgart-Bad Cannstatt: Frommann-holzboog, 1998. [in latin]. 1998 [Frommann-Holzboog in Latin].

BECKER, Daniel; FERRARI, Isabela. **VICTOR, the Brazilian Supreme Court's Artificial Intelligence: a beauty or a beast?** Disponível em: <https://sifocc.org/app/uploads/2020/06/Victor-Beauty-or-the-Beast.pdf> p.9 Acesso em: 13 de maio de 2021.

BODEMEINHER, Edgar. The Inherent Conservatism of the Legal Profession. **Indiana Law Journal**, Indiana, v. 23, n. 3, p. 221-235, spring [set.-dez.]. 1948.

BRASIL. **Lei Nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

CHIBENI. Silvio Seno. **Algumas observações sobre o “método científico**. Disponível em <https://www.unicamp.br/~chibeni/textosdidaticos/metodocientifico.pdf> Acesso em 17 mar. 2021.

COELHO, Fábio Ulhoa. **Curso de Direito Comercial**. v. 3. São Paulo: Saraiva, 2015.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, Palo Alto, v. 17, n. 3, p.37-54, 1996. ISSN 0738-4602.

GATLAN, Sergiu. **Brazil's court system under massive RansomExx ransomware attack**. Bleeping Computer. 5 nov. 2020, 16:09. Disponível em:

<https://www.bleepingcomputer.com/news/security/brazils-court-system-under-massive-ransomexx-ransomware-attack/>

GENESERETH, Michael. **Computational Law: The Cop in the Backseat**. [White Paper, CodeX—The Stanford Center for Legal Informatics], 2015.

GOEBEL, M.; GRUENWALD, L. 1999 apud VELÁSQUEZ, JD; GONZÁLEZ, P. Expanding the Possibilities of Deliberation: The Use of Data Mining for Strengthening Democracy with an Application to Education Reform. **Information Society**, v. 26, n. 1, p. 1-16, jan. 2010. ISSN: 01972243.

GOMES, Orlando. **Introdução ao direito civil**. Atualização e notas de Humberto Theodoro Junior. 15. ed. Rio de Janeiro: Forense, 2000. [apud TOMAZETTE, Marlon. Curso de Direito Empresarial: Teoria Geral e Direito Societário. v. 1, 3. ed. São Paulo: Atlas, 2011. p. 19.]

GOOGLE. [Chrome] **The Browser Build by Google**. [software] Disponível em <https://www.google.com/chrome/index.html> Acesso em: 17 mar. 2021

GRAY, John Chipman. **The Nature and Sources of the Law**. New York: The Columbia University Press, 1909.

HOBBS, Thomas. **Leviatã: Matéria, forma e poder de um Estado eclesiástico e civil**. [1651] (Tradução de João Paulo Monteiro e Maria Beatriz Nizza da Silva). 3. ed. São Paulo: AbrilCultural, 1983.

HUGGINFACE TEAM, **The Transformer**. Disponível em: <https://transformer.huggingface.co>

INTERNET LIVE STATS. **Internet Users**. <<https://www.internetlivestats.com/internet-users/>>. Acesso em: 26 jun. 2020.

INTERNET LIVE STATS. **Total Number of Websites**. <<https://www.internetlivestats.com/total-number-of-websites/>>. Acesso em: 01 abr. 2020.

INTERNET WORLD ARCHIVE. **About the Internet Archive**. <<https://archive.org/about/>>. Acesso em: 26 jun. 2020.

KATZ, Daniel Martin; BOMMARITO II, **Michael J. Measuring the Complexity of the Law: The United States Code.** [Draft v1.01] 2013.

LEIBNIZ, Gottfried Wilhelm. **De Conditionibus.** Paris: Librairie Philosophique J. VRIN, 2002.

LEX MACHINA. **Legal Analytics Apps.** Disponível em: <https://lexmachina.com/legal-analytics-apps/>. Acesso em: 18 mar. 2021.

LOEVINGER, Lee. Jurimetrics--The Next Step Forward. **Minnesota Law Review**, [1796] v. 33 n. 5, University of Minnesota Law School, abr. 1949.

MATTAR, João. **Metodologia Científica na Era Digital.** São Paulo: Saraiva, 2017.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data.** New York: Houghton Miffling Harcourt Publishing Company, 2014.

McGINNIS, John O.; WASICK, Steven. Law Algorithm. 66 **Florida Law Review.** 2014

MEHROTRA, Dheeraj. **Basic Concepts of Information Technology (IT).** [S.l.], [S.D.] [disponível em www.computerscience.com] Acesso em: 19 mar. 2021.

MICROSOFT. **Description of the database normalization basics.** Disponível em: <https://docs.microsoft.com/en-us/office/troubleshoot/access/database-normalization-description> Acesso em 19 mar. 2021.

MICROSOFT. **Microsoft Edge.** [software] Disponível em <https://microsoftedgewelcome.microsoft.com/> Acesso em: 17 mar. 2021

MOZILLA. **Firefox Browser.** [software] Disponível em: <https://www.mozilla.org/> Acesso em: 17 mar. 2021

NUNES, Marcelo Guedes. **Jurimetria: Como a Estatística pode reinventar o Direito.** São Paulo: Revista dos Tribunais, 2016.

OBSERVATORY OF ECONOMIC COMPLEXITY – **OECD. Brazil.** Disponível em: https://atlas.media.mit.edu/en/profile/country/bra/#Trade_Balance. Acesso em: 26 out. 2018.

PINTO, Henrique Alves. A utilização da inteligência artificial no processo de tomada de decisões: Por uma necessária accountability. **Revista de Informação Legislativa – RIL**, Brasília a. 57 n. 225 p. 23-42 jan./mar. 2020.

ROSS, Alec. **The Case for Connectivity**: The upward economic mobility needed to sustainably address basic needs in the 21st century requires an Internet connection. Any economy that remains offline will grow less competitive and more isolated as the globe does its business online. HuffPost, online, abr. 2014. Disponível em:

RUSSEL, Jon. **Tencent posts \$21.9 billion in annual revenue, its highest growth since 2012**. Disponível em: <<https://techcrunch.com/2017/03/22/tencent-2016-revenue/>>. Acesso em: 26 out. 2020.

SHERMER, Michael. **The Really Hard Science**: To be of true service to humanity, science must be an exquisite blend of data, theory and narrative. Scientific American, [S.I.], 1 out. 2007. Disponível em: <<https://www.scientificamerican.com/article/the-really-hard-science/#>>.

SILVA, Marcelino Pereira dos Santos. **Mineração de Dados – Conceitos, Aplicações e Experimentos com Weka**. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 12 mar. 2018.

SPERANDIO. Henrique Raimundo do Carmo. **Desafios da Inteligência Artificial para a Profissão Jurídica**. 107 p. Dissertação de Mestrado – Fundação Getulio Vargas (FGV). 2018. P. 28-29.

STANFORD [Stanford Profiles]. **Michael Genesereth**. Disponível em: <https://profiles.stanford.edu/michael-genesereth?tab=bio> Acesso em: 17 março 2021.

SUPERIOR TRIBUNAL DE JUSTIÇA. **Inteligência artificial está presente em metade dos tribunais brasileiros, aponta estudo inédito**. [Publicado em 09 mar. 2021]. Disponível em: <https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/09032021-Inteligencia-artificial-esta-presente-em-metade-dos-tribunais-brasileiros--aponta-estudo-inedito.aspx>. Acesso em: 18 mar. 2021.

SUPREMO TRIBUNAL FEDERAL. **Inteligência artificial vai agilizar a tramitação de processos no STF**. [Publicado em 30 de maio de 2018]. Disponível em: <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=380038&ori=1> Acesso em: 17 de março de 2021.

SUPREMO TRIBUNAL FEDERAL. **Licitação 108/2012** [pregão eletrônico]. Processo 348788. Disponível em:
<http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=21550>

SUPREMO TRIBUNAL FEDERAL. **Licitação 124/2009** [pregão eletrônico]. Processo 336684. Disponível em:
<http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=8600>

SUPREMO TRIBUNAL FEDERAL. **Licitação 76/2011** [pregão eletrônico]. Processo 344911. Disponível em:
<http://www.stf.jus.br/portal/edital/verEditalAndamento.asp?licitacao=16201>

SUPREMO TRIBUNAL FEDERAL. **Projeto VICTOR do STF é apresentado em congresso internacional sobre tecnologia.** [Publicado em 26 de setembro de 2018] Disponível em:
<http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=390818> Acesso em: 17 de mar. 2021.

SUPREMO TRIBUNAL FEDERAL. **STF vai contratar seu primeiro sistema com fundamento na legislação de incentivo à inovação tecnológica.** [Publicado em 12 de setembro de 2019] Disponível em:
<http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=423418&ori=1> Acesso em: 17 de março de 2021.

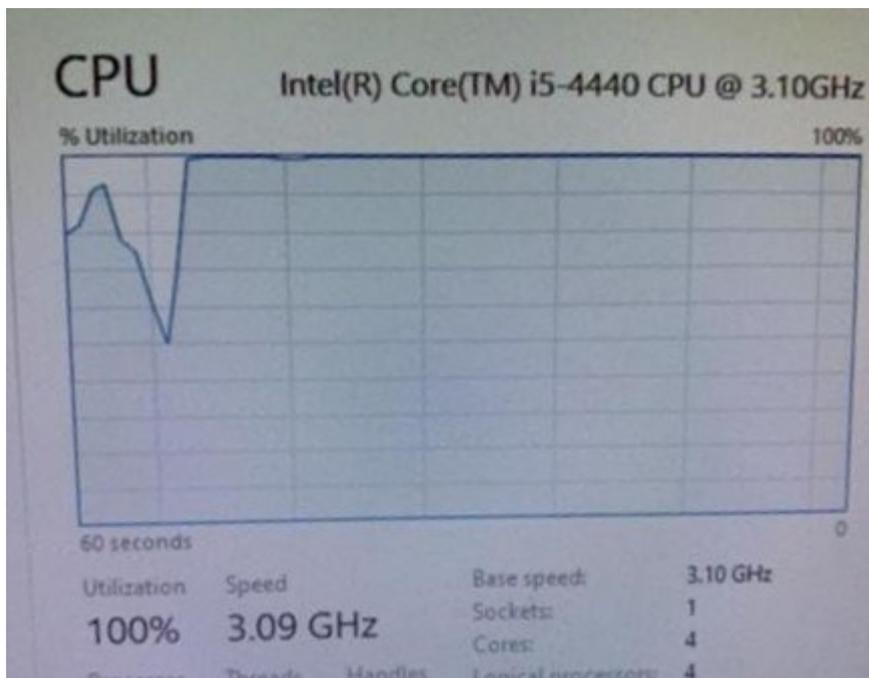
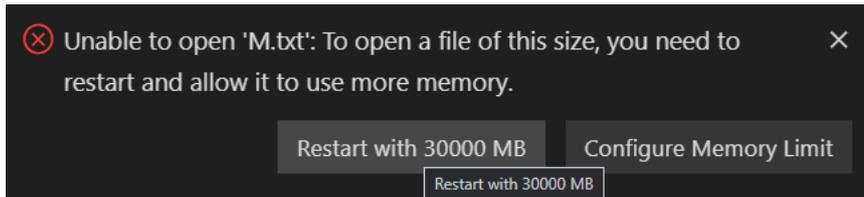
TATE MUSEUM. Benjamin Hoadly, **Bishop of Winchester.** [Informações da pintura de Hoadly, produção de William Hogarth em 1741]. Disponível em:
<<https://www.tate.org.uk/art/artworks/hogarth-benjamin-hoadly-bishop-of-winchester-n02736>>. Acesso em: 31 jan. 2019.

TOMAZETTE, Marlon. **Curso de Direito Empresarial: Teoria Geral e Direito Societário.** v. 1, 3. ed. São Paulo: Atlas, 2011.

WE ARE SOCIAL; HOOTSUITE. **Digital in 2017 Global Overview.** Disponível em:
<https://wearesocial.com/special-reports/digital-in-2017-global-overview> Acesso em: 11 março 2021.

ZABALA, Filipe J.; SILVEIRA, Fabiano F. **Decades of Jurimetrics.** [arXiv:2001.00476v131]?, 31 dez. 2019.

ANEXO A – EVIDÊNCIAS DA UTILIZAÇÃO DE RECURSOS



826,910 Files, 1,188 Folders

Task Manager

File Options View

Processes Performance App history Startup Users Details Services

CPU
100% 3.76 GHz

Memory
22.4/63.8 GB (35%)

Disk 0 (X:)
HDD
0%

Disk 1
SSD
0%

Disk 2
SSD
0%

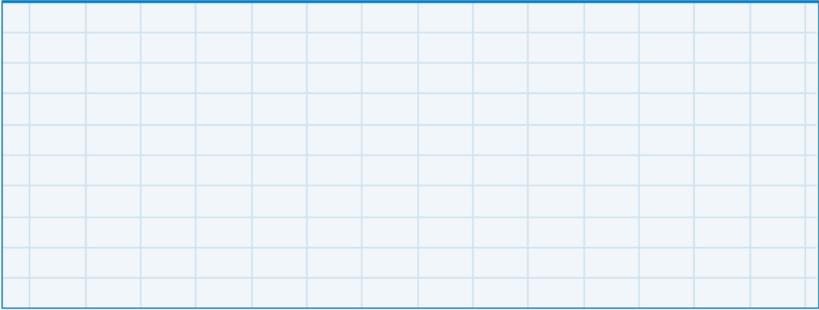
Disk 3 (C:)
SSD
6%

Disk 4 (I:)
Removable
0%

Disk 6 (K:)

Intel(R) Core(TM) i9-9900 CPU @ 3.10GHz

% Utilization



60 seconds

Utilization	Speed	Base speed:	3.10 GHz
100%	3.76 GHz	Sockets:	1
Processes	Threads	Handles	Cores:
244	3753	111450	8
Logical processors:	16		
Virtualization:	Enabled		
L1 cache:	512 KB		
L2 cache:	2.0 MB		
L3 cache:	16.0 MB		
Up time	0:22:12:42		

^ Fewer details | Open Resource Monitor

CrystalDiskMark 7.0.0 x64 (UWP) <REAL>

File Settings Profile Theme Help Language

All 9 16MiB C: 75% (147/195GiB)

	Read [MB/s]	Write [MB/s]
SEQ1M Q1T1	1551.07	1790.09
RND4K Q1T1	51.01	99.69
RND4K (IOPS)	12454.35	24338.38
RND4K (µs)	80.05	40.87