

Laboratório de Inteligência Artificial Aplicada da 3.^a Região - LIAA-3R

SIGMA - Ranqueamento de modelos de atos judiciais e administrativos a partir de informações identificadas nas peças processuais.

São Paulo, fevereiro de 2020

SIGMA - Ranqueamento de modelos de atos judiciais e administrativos a partir de informações identificadas nas peças processuais.

Projeto de pesquisa apresentado ao Centro de
Inteligência Artificial aplicada
ao sistema Processo Judicial Eletrônico – PJe.

São Paulo, fevereiro de 2020

Sumário

1. Identificação do projeto.....	04
2. Introdução.....	07
3. Problema de Pesquisa.....	08
4. Justificativa.....	08
5. Objetivos.....	09
6. Metodologia.....	09
7. Cronograma.....	10
8. Referências preliminares	10

1. IDENTIFICAÇÃO DO PROJETO

Título: SIGMA - Ranqueamento de modelos de minutas a partir de informações identificadas nas peças processuais.

Grupo de pesquisa:

NOME	LOTAÇÃO	PERFIL
AKI ANDO KOJIMA	SJSP	Validação Ética e Jurídica de Modelos, Programadora Pré-Processamento e Data Augmentation
AUGUSTO CÉSAR DE CASTRO	TRF3R	Cientista de IA
CAIO MOYSÉS DE LIMA	SJSP	Coordenador, Programador Pré-Processamento e Data Augmentation e Testes de Caixa Preta e Validação do Modelo
CLÁUDIO ROBERTO NÓBREGA MARTINS	TRF3R	Documentação e Apoio Operacional
DAVID PANESSA BACCELLI	TRF3R	Testes de Caixa Preta e Validação do Modelo e Apoio Operacional
ELISA EMIKO TANAKA DELLI PAOLI	SJSP	Testes de Caixa Preta e Validação do Modelo
FÁBIO AKAHOSHI	TRF3R	Gestor Técnico,

COLLADO		Cientista de IA e Testes de Caixa Preta e Validação do Modelo
GIULIA YURIKO TANAKA	SJSP	Programadora Pré-Processamento e Data Augmentation, Testes de Caixa Preta e Validação do Modelo, Documentação e Apoio Operacional
JOÃO PAULO TIVERON	SJSP	Cientista de Dados e Programador Pré-Processamento e Data Augmentation
LUCIANA ORTIZ TAVARES COSTA ZANONI	SJSP	Validação Ética e Jurídica dos Modelos e Testes de Caixa Preta e Validação do Modelo
LUIZ GUILHERME MARTINS	SJSP	Testes de Caixa Preta e Validação do Modelo
MAÍRA ZAU SERPA SPINA D'EVA	TRF3R	Validação Ética e Jurídica dos Modelos, Documentação e Apoio Operacional
MARIA ALICE LEIS OLIVARES	TRF3R	Documentação e Apoio Operacional
MATHEUS HENRIQUE DE	TRF3R	Validação Ética e Jurídica dos Modelos

PAIVA CARVALHO		e Testes de Caixa Preta e Validação do Modelo
NATÁLIA TAVARES AMATO	SJSP	Validação Ética e Jurídica dos Modelos
PAULO CEZAR NEVES JUNIOR	SJSP	Validação Ética e Jurídica dos Modelos e Testes de Caixa Preta e Validação do Modelo
PEDRO HENRIQUE LOPES GUERRA	SJSP	Cientista de Dados
RAFAEL ARRUTI ARAGÃO VIEIRA	TRF3R	Cientista de Dados
RENATA DE SOUZA PLENS	SJSP	Validação Ética e Jurídica de Modelos, Programadora Pré-Processamento e Data Augmentation
RENATO ARRUDA ROCHA MONTEIRO	TRF3R	Validação Ética e Jurídica de Modelos e Testes de Caixa Preta e Validação do Modelo
ROBERTO NONATO BARRETO COELHO E SILVA	TRF3R	Cientista de Dados e Programador Full-Stack
RODRIGO GONÇALVES YUNOGUTHI	SJSP	Programador Full-Stack
RODRIGO VIEIRA DA SILVA	TRF3R	Engenheiro de IA

ROGÉRIO ANTÔNIO BATISTA DE ARAÚJO	SJSP	Programador Full-Stack, Documentação e Apoio Operacional
SÉRGIO RICARDO LOZANO	SJSP	Cientista de Dados e Programador Pré- Processamento e Data Augmentation
SINARA MARIA REIS CHAVES	SJSP	Testes de Caixa Preta e Validação do Modelo
VAL EMERSON ARALDI	SJSP	Cientista de Dados, Programador Full- Stack

Linha de pesquisa:

Soluções de apoio à decisão dos magistrados.

2. INTRODUÇÃO

A utilização, no âmbito do Poder Judiciário, do potencial inovador da aplicação da Inteligência Artificial à Linguagem Natural encontra um terreno fértil na Justiça Federal da 3.^a Região em razão da digitalização de seu acervo processual, da implantação do Processo Judicial Eletrônico e da criação dos laboratórios de inovação, bem como em razão da disponibilidade de tecnologias “estado da arte” em código aberto. Faz-se necessária a busca de soluções, principalmente, para o apoio à decisão dos magistrados, conforme preconizado no item 1.3.2 do Edital 2/2019 do CNJ, considerando que é nessa etapa que há a maior quantidade de processos aguardando uma decisão judicial. Nesse contexto, propõe-se o SIGMA, um sistema de centralização dos modelos e ranqueamento com utilização de inteligência artificial e da SINARA, outro projeto em andamento no laboratório. A ideia é que, durante a redação das minutas, sejam identificadas nas peças processuais que compõem o processo informações chaves para a seleção dos

modelos e, dessa forma, sejam sugeridos, ao usuário, modelos que já foram utilizados em casos semelhantes.

O SIGMA é uma ferramenta cujo protótipo já foi implementado e será útil para reduzir o tempo de produção de minutas e evitar decisões contraditórias. Ainda assim, o principal motivo de escolha desse sistema de ranqueamento como segundo projeto do LIAA-3R, na frente de tantos outros, é a possibilidade de utilização de aprendizado por reforço, produzindo um dataset que servirá de base para futuras implementações.

3. PROBLEMA DE PESQUISA

O foco do projeto é a pesquisa das melhores técnicas e desenvolvimento de algoritmos de ranqueamento a partir de informações extraídas das principais peças processuais que influenciam a decisão minutada. As informações extraídas serão fornecidas por outros algoritmos submetidos ao SINAPSES, tais como a SINARA, ou informações mais simples, de fácil extração, já que o foco do trabalho é o ranqueamento.

O SIGMA já teve um protótipo implementado, embora com escopo reduzido, na Vice-Presidência do TRF3.

4. JUSTIFICATIVA

A utilização, no âmbito do Poder Judiciário, do potencial inovador da aplicação da Inteligência Artificial à Linguagem Natural encontra um terreno fértil na Justiça Federal da 3.^a Região em razão da digitalização de seu acervo processual, da implantação do Processo Judicial Eletrônico e da criação dos laboratórios de inovação, bem como em razão da disponibilidade de tecnologias “estado da arte” em código aberto.

Faz-se necessária a busca de soluções para automação dos processos e rotinas de trabalho da atividade judiciária e para o apoio à decisão dos magistrados, conforme preconizado nos itens 1.3.1 e 1.3.2 do Edital 2/2019 do CNJ.

Tais necessidades dependem da criação de algoritmos de entendimento de um texto jurídico constante de uma peça processual, os quais servirão de apoio a tarefas mais específicas, tais como o agrupamento de processos semelhantes, identificação de assuntos para distribuição, auxílio na produção de minutas através de seleção de modelos, pesquisa de Jurisprudência,

emissão automática de certidões etc. Deve ser considerada a modularização desses algoritmos de entendimento de um texto jurídico, considerando as necessidades de auditorias e manutenções quando ocorrerem alterações legislativas e jurisprudenciais.

Nesse contexto, propõe-se a criação do SIGMA, um sistema de centralização dos modelos e ranqueamento com utilização de inteligência artificial e da SINARA, outro projeto em andamento no laboratório. A ideia é que, durante a redação das minutas, sejam identificados nas peças processuais que compõem o processo informações chaves para a seleção dos modelos e, dessa forma, sejam sugeridos, ao usuário, modelos que já foram utilizados em casos semelhantes.

5. OBJETIVOS

O objetivo do projeto é a criação de um sistema centralizado de armazenamento dos modelos de decisões, com sistemas de buscas avançados e ranqueamento por inteligência artificial. Para o ranqueamento deve-se estudar que informações podem ser extraídas do processo e como trabalha-las para que o sistema preditivo tenha o máximo de eficiência. Ao final, o sistema de ranqueamento deve ser transformada em uma API para possibilitar a utilização em outros projetos.

6. METODOLOGIA

Já existe um dataset produzido por aprendizado por reforço do protótipo do SIGMA, implementado na Vice-Presidência do TRF-3. Diferentemente de todos os demais problemas, esse dataset não será utilizado para entregar uma rede treinada, mas simplesmente para testar nossos diferentes algoritmos. Isso porque o objetivo do SIGMA é o ranqueamento de modelos de um setor da área fim, e os modelos mudam conforme a área em que se encontrem. Nesse sentido, o SIGMA é implantado, inicialmente, sem qualquer treinamento.

Ou seja, inicialmente o SIGMA gera a lista de modelos em ordem aleatória, para que o usuário os selecione com base em buscas comuns. Posteriormente, conforme for aumentando a quantidade de dados produzidos, o SIGMA vai alterando o algoritmo de ranqueamento e treinando a rede com os dados gerados.

A aferição dos resultados será por comparação entre os primeiros colocados do ranking com os modelos selecionados pelo usuário.

7. CRONOGRAMA

ENTREGA	CRONOGRAMA	
	INÍCIO	TÉRMINO
Verificação da possibilidade técnica e orçamentária de implementação do SIGMA em um servidor do Tribunal	01/04/2020	01/05/2020
Reuniões para planejamento estratégico e definição das próximas etapas	01/04/2020	01/05/2020
Identificação dos problemas e oportunidades que o SIGMA ainda não explorou	01/04/2020	01/05/2020
Implementação das soluções	01/05/2020	26/08/2020
Implementação da SINARA no SIGMA	01/05/2020	01/06/2020
Implementação do sistema de ranqueamento por IA	01/05/2020	01/08/2020

8. REFERÊNCIAS PRELIMINARES

A escolha do tema é decorrência da possibilidade de utilização de aprendizado por reforço. Diferentemente da SINARA, as tecnologias e os algoritmos que serão implementados não

envolvem grande complexidade. Ainda assim, sugerimos, por ora, o seguinte referencial teórico:

Rama Kumar Pasumarthi, Sebastian Bruch, Xuanhui Wang, Cheng Li, Michael Bendersky, Marc Najork, Jan Pfeifer, Nadav Golbandi, Rohan Anil, Stephan Wolf (2019). TF-Ranking: Scalable TensorFlow Library for Learning-to-Rank.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. Le. (2019) XLNet: Generalized Autoregressive Pretraining for Language Understanding.

Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, Russell Power. (2019) End-to-End Neural Ad-hoc Ranking with Kernel Pooling.

Deepti Ghadiyaram, Matt Feiszli, Du Tran, Xueting Yan, Heng Wang, Dhruv Mahajan. (2019) Large-scale weakly-supervised pre-training for video action recognition.