Projeto de Formatura – 2025 – Press Release
PCS - Departamento de Engenharia de Computação e Sistemas Digitais
Engenharia Elétrica – Ênfase Computação

Tema: Desenvolvimento de um Modelo de Segmentação Automática para Documentos Jurídicos

A digitalização acelerada de processos jurídicos aumentou de forma significativa o volume de contratos eletrônicos, tornando sua leitura e análise manual lentas e suscetíveis a falhas. Para enfrentar esse cenário, o trabalho de formatura de Rafael Sandrini Guaracho apresenta um modelo automático capaz de segmentar e rotular contratos jurídicos diretamente a partir do texto bruto.

O projeto combina duas frentes complementares de pesquisa. A primeira é um modelo bi-encoder leve, que analisa contratos linha a linha para detectar fronteiras e atribuir categorias como Título do Contrato, Preâmbulo, Artigo, Assinatura e Anexo. A segunda é um pipeline de aumento de dados baseado em grandes modelos de linguagem (LLMs), responsável por gerar anotações "silver" em contratos não rotulados, ampliando o conjunto de treino e reforçando a capacidade de generalização do modelo.

Os experimentos foram conduzidos sobre um corpus bilíngue (inglês e francês) contendo contratos anotados manualmente e um conjunto maior de documentos não anotados. Utilizando métricas como F2-score, precision e recall, observou-se uma melhora expressiva após a inclusão das anotações geradas pelo LLM, com aumento do desempenho e redução dos erros de fronteira, evidenciando a eficácia do método de aumento de dados, em que LLMs atuam como anotadores automáticos e seus rótulos são transferidos para modelos menores, eficientes e de baixo custo.

A abordagem representa um avanço para o Processamento de Linguagem Natural aplicado ao Direito, permitindo estruturar contratos extensos de forma automática e consistente. Isso abre caminho para sistemas mais ágeis de análise contratual, revisão documental e extração de informações jurídicas complexas, diretamente a partir de texto simples.

Integrante: Rafael Sandrini Guaracho

Professor(a) Orientador(a): Profª Tereza Cristina Melo de Brito Carvalho