

Ciência da Computação / Sistemas de Informação

## **Avaliação Experimental de Abordagens e Bases de Dados para Reconhecimento de Entidades Nomeadas na Língua Portuguesa**

Júlia Ribeiro - 5o período de Sistemas de Informação, UFLA, bolsista PIBIC/FAPEMIG

Denilson Alves Pereira - Orientador DCC,UFLA - Orientador(a)

### **Resumo**

O Reconhecimento de Entidades Nomeadas (Named Entity Recognition, NER) é uma tarefa fundamental do Processamento de Linguagem Natural (PLN), que consiste em identificar e classificar automaticamente entidades mencionadas em textos — como pessoas, organizações, locais, datas e valores — de acordo com categorias predefinidas. Essa tarefa é especialmente desafiadora na língua portuguesa devido à escassez de recursos padronizados e de domínios dos dados disponíveis. Este trabalho teve como objetivo realizar uma avaliação experimental de abordagens e bases de dados para o NER em português, considerando tanto modelos tradicionais, como o BERT, quanto Large Language Models (LLMs). Para isso, foram padronizadas e avaliadas bases de dados amplamente utilizadas na literatura: Primeiro HAREM, Segundo HAREM, HAREM-mini, Paramopama, LeNER-Br, UlyssesNER-Br e CachacaNER. O processo de padronização buscou uniformizar os esquemas de anotação e garantir consistência interna entre partições, viabilizando a criação de benchmarks comparáveis e reproduzíveis. Benchmarks são essenciais para a área, permitindo medir o progresso dos modelos de forma justa e objetiva, além de fomentarem boas práticas experimentais e facilitarem a adoção de novas abordagens em diferentes contextos. O particionamento dos dados foi realizado com um algoritmo modificado de Estratificação Iterativa, garantindo equilíbrio entre as classes de entidades e a integridade documental dos conjuntos. Essa etapa é essencial para assegurar que os modelos sejam avaliados de forma justa e robusta, evitando vieses decorrentes de distribuições desbalanceadas ou inconsistentes. Em seguida, conduziram-se experimentos com fine-tuning do modelo BERT e few-shot learning com LLMs, especificamente LLaMA 3.1 8B e Qwen 3 14B. Os resultados demonstraram que o BERT ajustado apresentou desempenho significativamente superior aos LLMs sem fine-tuning, com F1-scores entre 0,69 e 0,90, enquanto os LLMs obtiveram valores inferiores a 0,30 em todas as bases de dados avaliadas. Como trabalho futuro, pretende-se realizar fine-tuning dos próprios LLMs, buscando explorar seu potencial em tarefas de NER após o ajuste fino, especialmente em contextos de baixa supervisão. Todos os dados, particionamentos e códigos serão disponibilizados publicamente para apoiar pesquisas futuras e fomentar benchmarks padronizados em língua portuguesa, contribuindo para o avanço da área e a consolidação de boas práticas experimentais.

Palavras-Chave: LLMs, Processamento de linguagem natural, Bert.

Instituição de Fomento: FAPEMIG

Link do pitch: <https://www.youtube.com/watch?v=kUSBv1CULa8>