

Relatório Técnico: Reconhecimento e Normalização de Entidades em Boletins de Ocorrência com BERTimbau

1. Objetivo

Este relatório apresenta um experimento de processamento de linguagem natural (PLN) com o objetivo de identificar e normalizar entidades nomeadas em textos informais, especificamente em boletins de ocorrência simulados. A tarefa utiliza modelos baseados em transformers, com foco no BERTimbau, adaptado para NER (Named Entity Recognition).

2. Metodologia

O experimento foi conduzido a partir de um corpus simulado contendo 10 boletins de ocorrência, com anotações de entidades do tipo Pessoa, Organização e Local. Utilizou-se o modelo 'pucpr/bert-base-portuguese-cased-ner' via Hugging Face Transformers. Foi implementada uma heurística simples para normalização de nomes baseada em similaridade textual.

3. Ferramentas Utilizadas

- Python 3.10
- Hugging Face Transformers
- pandas
- difflib (para normalização)
- CSV como formato de entrada do corpus

4. Exemplo de Saída do Modelo

Exemplo de texto analisado:

Texto: manifestantes foram dispersados pela pm no centro de fortaleza.

Entidades detectadas:

- pm → Organização → Polícia Militar
- centro de fortaleza → Local → Centro, Fortaleza-CE

5. Resultados

O modelo demonstrou boa capacidade de reconhecimento em textos curtos e informais, sendo capaz de extrair entidades-chave mesmo com grafia simplificada. A heurística de normalização funcionou bem para casos com variações previsíveis, como siglas ou nomes abreviados. A integração com um dicionário ampliado pode aumentar a precisão.

6. Considerações Finais

Este experimento demonstrou o potencial de uso de modelos baseados em transformers no domínio de segurança pública. A próxima etapa inclui o aumento do corpus, integração com bases relacionais da SSP e expansão da normalização via embeddings contextuais.