

Trabalho Prático II

Nosso objetivo é estudar a tarefa de POS Tagging para a língua Portuguesa. POS Tagging (do inglês *Part-of-Speech Tagging*) consiste em rotular cada palavra de um texto com sua respectiva classe gramatical (por exemplo, substantivo, verbo, adjetivo, advérbio etc.).

Em termos práticos, o objetivo do POS Tagging é atribuir a cada palavra a categoria que melhor descreva seu papel sintático na frase, **considerando o contexto**.

Por exemplo, na frase em português:

“Ela corre todos os dias.”

- “Ela” pode receber a tag de pronome.
- “corre” pode receber a tag de verbo.
- “todos” pode receber a tag de pronome indefinido.
- “os” é um artigo definido.
- “dias” é um substantivo.

Essa anotação ajuda sistemas de PLN a “entenderem” melhor a estrutura de cada sentença.

Você deverá implementar uma solução de POS Tagging, e para isso utilizará o corpus Mac-Morpho, produzido pelo grupo NILC da ICMC USP. O Mac-Morpho oferece dados para treinamento, validação e teste de modelos preditivos, capazes de classificar a classe gramatical de palavras em Português. Para acessar o conjunto de classes, acesse <http://nilc.icmc.usp.br/macmorpho/macmorpho-manual.pdf>.

O corpus está disponível em <http://nilc.icmc.usp.br/macmorpho/macmorpho-v3.tgz>. Você deverá implementar e avaliar a precisão de um modelo de POS Tagging, a sua escolha. Você pode utilizar pacotes que facilitem a implementação, como gensim, transformers, ou até mesmo LLMs com zero-shot. Você deverá entregar documentação embutida em um notebook, detalhando a tarefa, a metodologia e os resultados. Sua análise deverá discutir quais as classes gramaticais para as quais a precisão é maior/menor.