

# Naive Parser



Disciplina: MATA04 - Tópicos em Bancos de Dados  
Aluno: Fernando Medeiros do Nascimento

# A idéia

Não ter que fazer toda implementação de regras para descobrir a POS tag de palavras

Porém, sem muito conhecimento destas regras, portanto não tem como inferi-las.

# Execução

Criação de duas tabelas:

A primeira armazena a ocorrência de cada palavra, e depois de normalizada, vira uma tabela de probabilidades.

A segunda armazena sequencia de tags. A quantidade de vezes que uma palavra com tag X é sucedida por uma palavra de tag Y. Depois de normalizada, vira uma nova tabela de probabilidades.

Recuperar, para cada token, a sua POS, baseado ou no seu valor (palavra), ou na derivação da palavra anterior.

# Resultados

Para os diferentes corpora, *dev*, *test* e *train*:

Working with macmorpho-dev corpus  
Testing Naive POS tagger:  
Num. of sentences: 50  
Num of Tokens: 1253  
Results: 779 right guesses, 1253  
total answers. 62.17% of precision

Working with macmorpho-test corpus  
Testing Naive POS tagger:  
Num. of sentences: 50  
Num of Tokens: 793  
Results: 435 right guesses, 793  
total answers. 54.85% of precision

Working with macmorpho-train corpus  
Testing Naive POS tagger:  
Num. of sentences: 50  
Num of Tokens: 1043  
Results: 936 right guesses, 1043  
total answers. 89.74% of precision

Resumo: Método INEFICIENTE para marcações morfosintáticas de palavras.

Apresenta *overfitting* para o corpus de treino. Porém, é ineficaz em qualquer outros corpora.