

**CENTRO UNIVERSITÁRIO FEI**

PEL 218 – Processamento de Linguagem Natural

**Classificar Part-Of-Speech**

**JONATHAN KENJI KINOSHITA**

**Matrícula: 120102-9**

**São Bernardo do Campo**

**2020**

## **Tarefa**

- Implementar o HMM
- Objetivo: classificar Part-Of-Speech (V, N, P, DT, etc...)
- Treinar na base Folha 95 (pode ser parcial)
- Testar a classificação na base Folha 94 (pode ser parcial)
- Fazer um relatório com experimentos

[Código disponibilizado no GitHub](#)

Primeiramente foi criado um código que lê os arquivos “cg.Folha.1994” e cg.Folha.1995” e a partir de cada um gera um arquivo .txt, o qual possui *tuples* contendo a palavra e a sua classe gramatical. Depois o código hmm.py recebe esses dois arquivos e realiza a classificação do Part-of-Speech implementado com o HMM.

Pode-se perceber que algumas vezes a mesma palavra apresenta classificação diferente entre as duas bases, possível motivo pelo qual a acurácia girou em torno de 80%. Já dividindo o mesmo dataset (cg.Folha.1994) entre treinamento e teste, obteve-se uma acurácia média de 98%.