

**Universidade Federal do Ceará**  
**Inteligência Artificial**  
**Trabalho Prático 4 - 09/02/2022**

Nome: \_\_\_\_\_

1) Calcule os seguintes itens:

- a) Número total de palavras
- a) Número total de sentenças
- b) Número total de palavras não repetidas
- c) Número total de palavras repetidas
- d) Média de palavras por sentenças

Esse procedimento deve ser feito para os seguintes textos:

- shakespeare-caesar.txt,
- shakespeare-hamlet.txt,
- shakespeare-macbeth.txt

2) Encontre as entidades nomeada presentes para o seguintes textos:

- i) apoloxi.txt
- ii) french-revolution.txt
- b) Qual é a quantidade de entidades "GPE" presentes em cada um dos texto?
- c) Qual é a quantidade de entidades "LOCATION" presentes em cada um dos texto?
- d) Qual é a quantidade de entidades "PERSON" presentes em cada um dos texto?

3) Utilizando as técnicas aprendidas sobre análise de sentimentos defina a polaridade dos arquivos presentes na pasta. Utiliza os arquivos positive\_words.txt e negative\_words.txt presentes na pasta.

4) Utilizando as técnicas aprendidas sobre *Term Frequency, Inverse Document Frequency*, listem as 5 palavras mais relevantes de cada texto contido no corpus Gutenberg. A sua análise deve ser em cima dos seguintes textos: 'austen-emma.txt', 'bible-kjv.txt', 'carroll-alice.txt', 'melville-moby\_dick.txt', 'shakespeare-caesar.txt' e 'shakespeare-hamlet.txt'. Lembre-se que antes de aplicar o as técnicas de TF IDF o texto deve está normalizado, com as *stopwords* retiradas e lematizado.