

Lista 02 (Atividade feita em sala) - Entrega: 16/12 - wellington@crateus.ufc.br

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) Em relação ao corpus “gutenberg” implemente os algoritmos para responder às seguintes questões:

- a) Total de palavras em cada documento do corpus “gutenberg”
- b) Quem é o maior documento do corpus?
- c) Quem é o menor documento do corpus?
- d) Calcular a média da quantidade sentenças por palavras do corpus “gutenberg”.
- e) Calcule a distribuição de frequência das palavras do livro “shakespeare-macbeth.txt”.
- f) Calcule 5 palavras mais frequentes nesse corpus.
- g) Mostre a diferença entre de palavras entre dois livros. (shakespeare-caesar.txt, shakespeare-hamlet.txt,)