

Lista 03 - Entrega: 21/12 - wellington@crateus.ufc.br

- 1) Para o corpus “shakespeare-caesar.txt”, usando expressões regulares responda os seguintes itens:
 - a) Quantidades de palavras que terminam com “r”;
 - b) Quantidade de palavras com 5 letras;
 - c) Quantidade de vezes que “err” ocorre no corpus;
 - d) Quantidade de vezes que “are” ocorre no corpus para as palavras com 5 ou mais caracteres.

- 2) Em relação do corpus “shakespeare-hamlet.txt”, faça as seguintes atividades:
 - a) Normalize o corpus (Retire os números e deixe todas as palavras minúsculas);
 - b) Aplique o lematizador em todas as palavras do corpus;
 - c) Aplique o tokenizador em todas as sentenças do corpus;
 - d) Aplique os pos tagger e responda a quantidade de adjetivos existem no corpus; Dica a tag de adjetivo é “JJ”