

PYTHON PARA PLN

LISTA DE EXERCÍCIOS 2

1. Crie uma variável com a string “ instituto de ciências matemáticas e de computação” e faça:
 - a. Concatene (adicione) uma outra string chamada “usp”
 - b. Concatene (adicione) uma outra informação: 2021
 - c. Verifique o tamanho da nova string (com as informações adicionadas das questões a e b), com referência a caracteres e espaços
 - d. Transforme a string inteiramente em maiúsculo
 - e. Transforme a string inteiramente em minúsculo
 - f. Retire o espaço que está no início da string e imprima a string
 - g. Substitua todas as letras ‘a’ por ‘x’
 - h. Separe a string em palavras únicas
 - i. Verifique quantas palavras existem na string
 - j. Separe a string por meio da palavra “de”
 - k. Verifique agora quantas palavras/frases foram formadas quando houve a separação pela palavra “de”
 - l. Junte as palavras que foram separadas (pode usar a separação resultante da questão h ou j)
 - m. Junte as palavras que foram separadas, mas agora separadas por uma barra invertida, não por espaços (pode usar a separação resultante da questão h ou j)

2. Escolha um corpus qualquer (pode ser o que foram utilizados na aula) e, usando as funções do NLTK, faça:
 - a. *Tokenize* o corpus inteiro (palavras, números e pontuações)
 - b. Verifique a quantidade de tokens do corpus
 - c. *Tokenize* o corpus apenas por suas palavras
 - d. Verifique a quantidade de palavras do corpus
 - e. Verifique a frequência de palavras no corpus
 - f. Verifique quais são as 5, 10 e 15 palavras mais frequentes do corpus
 - g. Extraia as *stopwords* do NLTK (não do corpus ainda)
 - h. Verifique a frequência dos tokens sem *stopwords* do corpus