

CURSO DE EXTENSÃO PYTHON PARA PLN

LISTA DE EXERCÍCIOS 2

1. Crie um dicionário com 5 entradas e suas respectivas chaves e valores.
Faça:
 - a. Imprima todas as chaves do dicionário
 - b. Imprima todos os valores do dicionário
 - c. Imprima todos os itens do dicionário
 - d. Imprima o 2º item do dicionário
 - e. Imprima o dicionário completo
 - f. Percorra o dicionário, imprimindo para cada entrada o modelo “(chave) tem como valor (valor)”
2. Crie um arquivo e:
 - a. Escreva nele os números de 1 a 10
 - b. Depois de escrito, imprima na tela todos os números do arquivo
 - c. Escreva no arquivo os números de 11 a 20, substituindo os números que estavam antes no arquivo
 - d. Escreva no arquivo os números de 21 a 30, adicionando os números no final do arquivo (sem apagar o que já estavam lá)
 - e. Imprima na tela todos os números do arquivo novamente (de 11 a 30)
 - f. Imprima na tela todos os números do arquivo novamente, mas agora linha por linha
3. 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)
4. 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
 - i. Extraia todos os bigramas do corpus
 - j. Extraia todos os trigramas do corpus
 - k. Extraia todos os 4-gramas do corpus
 - l. Retorne as entidades nomeadas do corpus, usando os bigramas e trigramas
 - m. Escolha 3 palavras do seu corpus e faça o *stemming* delas
 - n. Separe uma sentença do seu corpus e retorne todas as classes gramaticais das palavras da sentença. Analise se o etiquetador acertou todas as classes gramaticais
 - o. Retorne as classes gramaticais de todas as palavras do seu corpus
 - p. Retorne as entidades nomeadas do seu corpus, usando a técnica de *chunking*

Após esses exercícios, façam os **exercícios de fixação** presentes nos **slides da aula**, presentes no Drive do curso: parte 1 (slides 26 a 29) e parte 2 (slide 33)