

CURSO DE EXTENSÃO PYTHON PARA PLN LISTA DE EXERCÍCIOS 3

1. Utilizando um corpus de seu interesse (pode ser o que foi utilizado na aula) e utilizando o spaCy, faça:
 - a. Tokenize todo o corpus
 - b. Tokenize todo o corpus e retorne a lista com strings
 - c. Tokenize todo o corpus e retorne apenas as palavras
 - d. Retorne a quantidade de palavras do corpus
 - e. Tokenize todo o corpus e retorne apenas os números
 - f. Retorne a quantidade de números do corpus
 - g. Tokenize todo o corpus e retorne apenas as pontuações
 - h. Retorne a quantidade de pontuações do corpus
 - i. Retorne a frequência de pontuação do corpus (obs: pode usar o NLTK nesse exercício!)
 - j. Tokenize todo o corpus e retorne a quantidade de espaços presentes no corpus
2. Ainda utilizando o mesmo corpus e o spaCy, faça:
 - a. Analise o 1º parágrafo do corpus, lematizando todos os tokens possíveis. Verifique se foi retornado os lemas corretos
 - b. Ainda com o 1º parágrafo, retorne todas as classes gramaticais dos tokens.
 - c. Retorne todas as classes gramaticais do corpus
 - d. Use o 1º parágrafo para retornar as informações morfológicas dos tokens
 - e. Retorne todas as entidades nomeadas do corpus
 - f. Verifique a classificação de cada entidade nomeada do corpus
 - g. Use o displaCy para mostrar graficamente todas as entidades nomeadas do corpus
3. Usando a sentença “*Ele queria jogar tênis com Janete, mas também queria jantar com Suzana. Sua indecisão o deixou louco.*”, faça, com o spaCy:
 - a. A análise sintática da frase
 - b. Use o displaCy para mostrar graficamente a árvore sintática resultante.

4. Use a função **similarity()** do spaCy para verificar a similaridade entre as palavras:
- a. Menino e menina
 - b. Pequim e China
 - c. Tablet e celular
 - d. Água e refrigerante
 - e. Caneta e teclado
 - f. Livro e computador

Após esses exercícios, façam os **exercícios de fixação** presentes nos **slides da aula**, presentes no Drive do curso, slides 32 e 33.