

## **CURSO DE EXTENSÃO PYTHON PARA PLN**

### **LISTA DE EXERCÍCIOS 3**

1. Escolha um corpus qualquer (pode ser o que foram utilizados na aula) e, usando as funções do NLTK, faça:
  - a. Extraia todos os bigramas do corpus
  - b. Extraia todos os trigramas do corpus
  - c. Extraia todos os 4-gramas do corpus
  - d. Retorne as entidades nomeadas do corpus, usando os bigramas e trigramas
  - e. Escolha 3 palavras do seu corpus e faça o *stemming* delas
  - f. 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
  - g. Retorne as classes gramaticais de todas as palavras do seu corpus
  - h. Retorne as entidades nomeadas do seu corpus, usando a técnica de *chunking*
  
2. 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
  
3. 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
4. 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.