
UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO

Processamento de Linguagem Natural
Professor: Thales Vieira

7a lista de exercícios

26 de agosto de 2020

Instruções:

A lista deve ser respondida por grupos de até 3 pessoas (graduação) e individualmente (mestrado).

Resoluções idênticas de grupos distintos serão desconsideradas.

O código e demais dados devem ser anexados a cada questão.

Data limite para entrega: 11/09/2020.

1. Usando sua base de textos e a biblioteca spaCy, realize as seguintes tarefas:

- a) Extraia as etiquetas gramaticais (POS) de cada token do seu textos.
- b) Calcule e plote um gráfico com as frequências de cada tipo gramatical.
- c) Extraia entidades do tipo pessoa e lugar dos seus textos.
- d) Identifique e liste as pessoas mais frequentes nos seus textos. Você só deve contar cada entidade 1 vez por documento.

- e) Extraia as árvores de dependências de seus documentos.
- f) Selecione as árvores onde aparecem alguma das 30 pessoas mais frequentes.
- g) Pesquise relações de dependência adequadas para extrair informações sobre estas 30 pessoas.
- h) Usando estas relações, extraia informações sobre estas 30 pessoas e organize-as de forma estruturada em triplas, como por exemplo: (pessoa, verbo, informação).