

Universidade Tecnológica Federal do Paraná
Graduate Program in Electrical and Computer Engineering - CPGEI

CAIA003 / DAT003 - Data Mining

Professors: Heitor Lopes and Thiago H Silva

Exercício 1 - Sobre coleta de dados da Web

i e ii – não precisam ser entregues.

i- Instale todas as bibliotecas mencionadas na apresentação teórica sobre coleta de dados.

ii- Crie uma aplicação no Twitter. Em seguida, gere os credenciais necessárias para acessar dados da API do Twitter.

1- Faça uma adaptação no coletor de tweets para coletar qualquer tipo de tweet compartilhado no Brasil. Dica: veja a documentação da biblioteca TwitterApi. Faça uma coleta por alguns minutos, salvando o resultado em um arquivo ou em um banco de dados.

2- Com base no dataset gerado na questão 1, faça:

- a) Separe todas as hashtags usadas nos Tweets.
- b) Descubra a porcentagem de tweets com hashtags.
- c) Identifique a hashtag mais popular.

3- Faça um coletor que capture as reviews da página principal de: https://www.apontador.com.br/local/sp/campinas/faculdades_e_universidades/88Q2E372/unicamp.html.

4- Crie uma aplicação que usa o Selenium para abrir o site <https://www.uol.com.br/> e rolar a página para baixo até o fim do conteúdo.