



Universidade Federal do Ceará
Centro de Ciências
Departamento de Computação

CKP9011 e CK0268 – Introdução à Ciência de Dados 2024.1

Lista 3

Exercício: Extração de Dados

Objetivos: Exercitar os conceitos referente à extração de dados.

Data da Entrega: 01/04/2024

OBS 1: Exercício Individual.

OBS 2: A entrega da lista deverá ser executada utilizando-se o SIGAA.

Questão 1

Crie um arquivo Jupyter Notebook e realize as seguintes operações:

- a) Para estudantes de graduação: extrair pelo menos 2.000 notícias (incluindo data, título, texto, link, autor etc.) de um dos seguintes portais:
 1. Globo.com
 2. UOL
 3. Folha
 4. Estadão
 5. CNN Brasil
 6. Veja
 7. Metrôpoles
- b) Para estudantes de mestrado: extrair pelo menos 1.000 notícias (incluindo data, título, texto, link, autor, avaliação etc.) de uma das seguintes agências de checagem de notícias:
 1. Abraji (projeto Comprova) – <https://projetocomprova.com.br/>
 2. AFP – <https://checamos.afp.com/>
 3. Agência Lupa – <https://lupa.uol.com.br/>
 4. Aos Fatos – <https://www.aosfatos.org/>
 5. Boatos.org – <https://www.boatos.org/>
- c) Para estudantes de doutorado: extrair pelo menos 1.000 notícias (incluindo data, título, texto, link, autor, avaliação etc.) de uma das seguintes agências de checagem de notícias:
 1. E-Farsas – <https://www.e-farsas.com/>
 2. Estadão Verifica – <https://politica.estadao.com.br/blogs/estadao-verifica/>
 3. Globo (Fato ou Fake) – <https://g1.globo.com/fato-ou-fake/>
 4. Universo Online (UOL Confere) – <https://noticias.uol.com.br/confere/>

“A Educação, qualquer que seja ela, é sempre
uma teoria do conhecimento posta em prática”.

Paulo Freire