

# Relatório 8 - Web Scraping com Python p/ Ciência de Dados (II)

Yuri Vacelh Zamulhak Zdebski

## Descrição da atividade

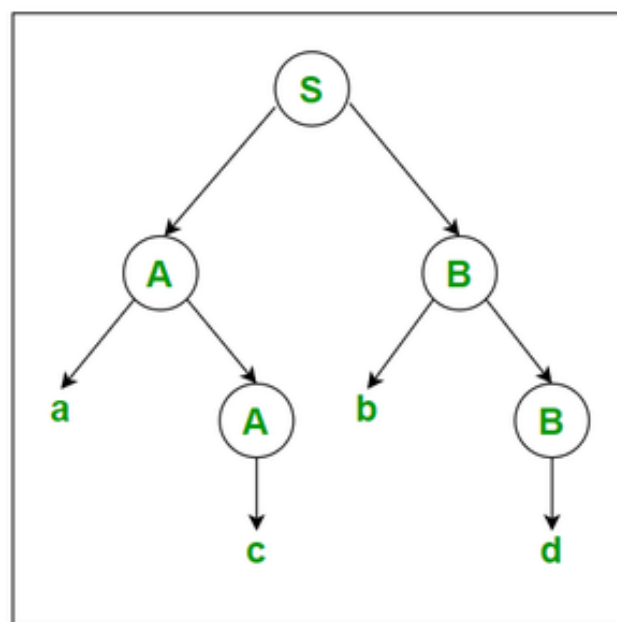
Assistir ao vídeo Web Scraping with Python - BeautifulSoup Crash Course, do canal FreeCodeCamp, sobre Web Scraping usando Python. O objetivo é ter contato com as ferramentas e aplicar os conhecimentos obtidos.

## Web Scraping

Web scraping é uma técnica de extração de dados de sites de forma automatizada. É amplamente utilizada para coletar informações de páginas web para análise de dados, monitoramento de preços, agregação de notícias, e muitas outras finalidades. A prática envolve a criação de scripts ou programas que enviam solicitações HTTP para uma página da web, recuperam o conteúdo da página e, em seguida, extraem as informações desejadas.

A biblioteca utilizada no curso foi a BeautifulSoup4, ela é extremamente popular pois apresenta sintaxe extremamente simples e intuitiva. O que a biblioteca faz por baixo dos panos é criar uma parser tree (analisar a sintaxe do HTML) e realizar buscas na árvore(o que é objetivamente mais eficiente do que buscar linearmente na string). O parser utilizado no curso foi o lxml, conhecido principalmente por sua velocidade, ele tem a capacidade de interpretar tanto HTML quanto XML.

Figura 1: Parser Tree



Fonte: Geeks for Geeks

## Atividade Prática

A atividade seguiu o padrão de dados que venho utilizando nos últimos módulos, o clima. Para ela, realizei um scrapping na página weather.com, de onde retirei dados como a temperatura atual, previsão de chuvas e qualidade do ar, para a capital Brasília. Isso possibilita a criação de um dataset especializado, extrair a temperatura atual sem a utilização de APIs, análise do clima na capital e muito mais.

## Conclusão

Em suma, o web scraping é uma ferramenta poderosa e versátil para extração automatizada de dados de sites, utilizada em diversas aplicações. Essa técnica possibilita a criação de grandes datasets, sem que necessariamente façamos as medições por conta própria, porém é sempre importante lembrar de fazer o scraping de forma ética, pedindo autorizações e creditando os dados, evitando assim futuras complicações jurídicas.

## Referencias

<https://www.geeksforgeeks.org/parse-tree-in-compiler-design/>