

Relatório- Card - 8 - Prática: Web Scraping com Python p/ Ciência de Dados (II)

Jefferson korte junior

Web scraping é o processo de extrair dados de sites da web. Pode ser feito manualmente ou usando ferramentas e bibliotecas que automatizam o processo. nesse caso do vídeo foi retirado dados do site 'time jobs'.

Para ler a página usamos **BeautifulSoup** que é uma biblioteca Python utilizada para Web Scraping de arquivos HTML e XML.

Facilitando a navegação na estrutura do documento, buscando e filtrando informações de forma simples. A princípio, utiliza-se a biblioteca BeautifulSoup, responsável por facilitar e simplificar o processo de coletar e filtrar informações.

É necessário coletar dados com base em cada tag e interesse que temos, requisitando o conteúdo HTML, analisando e extraindo as informações desejadas. Usa-se a biblioteca **requests** e BeautifulSoup. Requests faz a requisição HTTP e BeautifulSoup analisa o HTML.

`request.get('https://exemplo.com').text` faz uma requisição GET para o link especificado e retorna o conteúdo da página em formato de texto.

`.text` é usado para extrair o texto contido dentro de uma tag HTML, descartando todas as tags e sintaxe do HTML.

`find_all('tag', class_='class_name')` retornará uma array contendo todas as ocorrências específicas de tag com o nome da classe.

`find('tag', class_='class_name')` muito parecido com o `find_all()`, porém esse retorna apenas a primeira ocorrência detectada.

Adicionar features ao projeto pode melhorar a eficiência, a capacidade de lidar com diferentes tipos de dados e a automação de tarefas repetitivas. Caso o site utilizado tenha várias páginas, é possível automatizar a navegação por todas elas para extrair dados de todas as páginas disponíveis.

Também é possível salvar os dados em um arquivo CSV, JSON ou banco de dados para facilitar a análise posterior ou adicionar filtros para coletar apenas informações específicas.

Conclusão:

No card, aprendemos sobre web scraping, uma ferramenta incrivelmente poderosa e versátil para extrair dados de sites. Esta técnica possibilita a criação de grandes datasets, assim possibilitando análises e pesquisas. Com o web scraping, é possível automatizar a coleta

de informações, facilitando o acesso a dados que podem ser transformados em insights valiosos depois.

Referências:

<https://www.youtube.com/watch?v=XVv6mJpFOb0>