

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

Willian Augusto Soder de Souza

Este relatório visa abordar os principais conhecimentos adquiridos com o vídeo 'Web Scraping with Python - BeautifulSoup Crash Course'. É um vídeo bem didático, onde o apresentador explica de forma clara como realizar web scraping em Python. O vídeo começa com a explicação de alguns conceitos básicos de HTML, essenciais para a compreensão dos comandos em Python. Em seguida, são apresentados alguns comandos e conceitos de maneira didática. Abaixo, apresento um resumo dos principais comandos e bibliotecas, juntamente com suas respectivas funções:

- **Requests:** Biblioteca do Python usada para fazer requisições HTTP e obter o conteúdo das páginas web.
- **requests.get(url).text:** Retorna o conteúdo da página web em formato de texto.
- **BeautifulSoup:** Módulo da biblioteca bs4 usado para analisar e extrair dados de arquivos HTML e XML, facilitando a navegação e busca de informações específicas em páginas web.
- **.find_all e .find:** Métodos usados para buscar todos ou o primeiro elemento respectivamente, com base em critérios especificados em um documento HTML analisado pelo BeautifulSoup, retornando uma lista.
- **.text:** Método usado para obter o texto contido dentro de um elemento HTML, removendo as tags HTML e retornando apenas o texto visível.
- **.replace:** Método usado para substituir uma substring por outra em uma string.

Conclusão

O web scraping desempenha um papel fundamental no ecossistema do machine learning e data science ao possibilitar a coleta eficiente de dados relevantes diretamente da web. A habilidade de realizar web scraping não apenas amplia o acesso a fontes de dados variadas e em tempo real, mas também permite a criação de conjuntos de dados personalizados e de alta qualidade, essenciais para o treinamento de modelos preditivos precisos e escaláveis. Além de facilitar a automação da coleta de dados, aprender a realizar web scraping capacita profissionais a explorar e integrar informações de forma estratégica. Portanto, dominar as técnicas de web scraping não apenas otimiza o trabalho diário de um cientista de dados, mas também se revela crucial para manter-se competitivo e atualizado no dinâmico campo do machine learning e data science.