

Web Scraping com Python para Ciência de Dados

O tutorial proposto aborda uma introdução ao web scraping em Python usando a biblioteca BeautifulSoup.

- O vídeo começa com uma explicação sobre web scraping e conceitos básicos de HTML, demonstrando como as tags são usadas para estruturar o conteúdo em uma página da web.
- O tutorial inclui um exemplo prático de web scraping, onde o locutor demonstra como extrair postagens de empregos de um site específico usando o BeautifulSoup, fornecendo insights sobre filtragem e extração de informações relevantes. No projeto, ele demonstra como usar a criatividade para realizar extrações de dados já filtrados, coletar links e escrever os resultados separadamente em arquivos distintos. Ainda, é apresentado como deixar a coleta dos dados em looping, para rodar a cada determinada quantia de tempo.

Os códigos gerados a partir da aula estão disponíveis no GitHub. Tomei a liberdade de acompanhar a aula realizando a coleta de dados de um site diferente do mostrado no vídeo. Escolhi coletar os dados do site Rotten Tomatoes, que disponibiliza informações e avaliações a respeito de filmes. O projeto desenvolvido coleta informações sobre uma lista de 100 filmes melhores avaliados de todos os tempos. As informações são geradas e colocadas em arquivos em uma pasta chamada output.