

Relatório LAMIA 8

Prática: Web Scraping com Python p/ Ciência de Dados (II)

Kaique Medeiros Lima

Introdução

Esse relatório tende a abordar os principais conteúdos do CARD 8, “Prática: Web Scraping com Python p/ Ciência de Dados (II)”, o qual se consiste em assistir a uma vídeo aula do canal freecodecamp.org sobre Web Scraping em Python, fazendo uso da biblioteca BeautifulSoup4.

Descrição da atividade

Os principais conteúdos apresentados pelo vídeo foram os seguintes:

- **BeautifulSoup4:** É um módulo da biblioteca BS4, serve para a análise e extração de conteúdo de arquivos HTML.
- **Biblioteca Requests:** Faz o *request* de sites para a manipulação do conteúdo em Python.
- **find_all('tag', class_='class_name'):** Retorna uma array contendo as ocorrências da tag com o nome da classe, “filtrando” o arquivo HTML.
- **find('tag', class_='class_name'):** Retorna um array com a primeira ocorrência da tag + class_name.
- **.text:** Descarta as tags e sintaxe HTML e retorna apenas o texto.
- **requests.get('https://example.com').text:** Retorna o conteúdo da página em texto. Caso retorne 200, conectou com sucesso.

Conclusões

A importância do Web Scraping na área de IA e Machine Learning é a sua extração de dados de páginas da internet, os quais são usados para o aprendizado da máquina. Com a crescente quantidade de dados disponíveis online, a coleta e análise de grandes volumes de informações em tempo real são automatizadas pela biblioteca Requests, combinada com BeautifulSoup4 (BS4), que oferecem ferramentas robustas para a manipulação dos dados. Essa automação não só economiza tempo, mas também permite a adaptação e resposta rápida às mudanças extremamente rápidas dos dias atuais.