

Relatório 7 - Prática: Web Scraping com Python p/ Ciência de Dados (II)

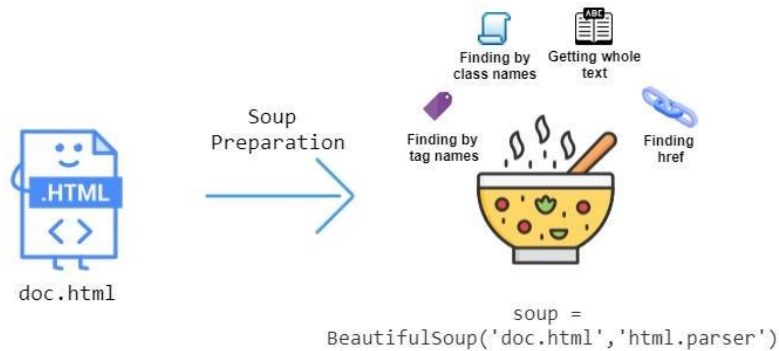
Beatriz Almeida Felício

Descrição da atividade

Web scraping é uma técnica de extração automatizada de dados da web que se tornou fundamental para análise de dados, pesquisa de mercado e automação de processos. O vídeo sugerido aborda a implementação de web scraping utilizando Python, com foco nas bibliotecas BeautifulSoup4 e Requests, apresentando suas funcionalidades, métodos de utilização e melhores práticas.

Fundamentos do Web Scraping

- **Definição e Aplicações:** O web scraping consiste no processo automatizado de extração de dados de páginas web, transformando informações não estruturadas em formato HTML em dados estruturados que podem ser armazenados e analisados. Suas aplicações incluem:
 - Monitoramento de preços em e-commerce
 - Coleta de dados para pesquisa de mercado
 - Agregação de notícias
 - Análise de sentimentos em mídias sociais
 - Geração de leads para marketing
- **Aspectos Legais e Éticos:** É fundamental considerar aspectos legais e éticos ao realizar web scraping. Ou seja, respeitar os termos de serviço dos sites, implementar delays entre requisições para não sobrecarregar servidores e obter autorização quando necessário.
- **Biblioteca Requests:** A biblioteca Requests é responsável por realizar requisições HTTP, sendo fundamental para o processo de web scraping. O módulo oferece suporte completo para diferentes tipos de requisições HTTP, incluindo GET para recuperar dados, POST para enviar dados, PUT para atualizar recursos e DELETE para remover recursos.
- **Biblioteca BeautifulSoup4:** BeautifulSoup4 é uma biblioteca de parsing HTML/XML que facilita a navegação e busca em documentos web. Suas principais funcionalidades incluem parsing de documentos HTML e XML, busca por tags, classes e Ids, extração de texto e atributos, modificação de conteúdo e etc.



Conclusões

O web scraping com Python, utilizando BeautifulSoup4 e Requests, é uma solução robusta e versátil para extração de dados da web. A combinação dessas bibliotecas oferece um conjunto completo de ferramentas para desenvolvimento de scrapers eficientes e confiáveis. A biblioteca Requests simplifica o processo de requisições HTTP, enquanto BeautifulSoup4 fornece uma interface intuitiva para parsing e navegação em documentos HTML. Juntas elas formam uma base sólida para projetos de web scraping em diferentes escalas e complexidades. O domínio dessas ferramentas e técnicas permite o desenvolvimento de soluções profissionais para coleta e análise de dados web, contribuindo significativamente para processos de tomada de decisão baseados em dados e automação de processos de negócio.

Referências

FREECODECAMP.ORG. **Web scraping with python - beautiful soup crash course**. Disponível em: <<https://www.youtube.com/watch?v=XVv6mJpFOb0>>. Acesso em: 26 dez. 2024.