

Aula 5: Desenvolvendo Scrappers

Eduardo Ribeiro Silva de Oliveira

07 de Outubro de 2024

Estratégias de Coleta de Dados

- ▶ Como identificar e selecionar fontes de dados.
- ▶ Importância de escolher fontes confiáveis e relevantes para o objetivo do scraping.

Identificação de Sites Relevantes

- ▶ Identificar os sites que possuem informações relevantes para o objetivo do scraping.
- ▶ Considerar critérios como qualidade da informação, acessibilidade e atualização dos dados.

Critérios para Seleção de Páginas

- ▶ Estabelecer critérios para selecionar páginas ou seções que forneçam dados úteis.
- ▶ Avaliar a estrutura do site e a localização dos dados de interesse.

Cuidados Legais e Éticos

- ▶ Respeitar a privacidade dos usuários e os termos de uso das plataformas.
- ▶ Entender as implicações legais de coletar dados de diferentes sites.

Implementação de Scrappers com Selenium

- ▶ Aplicação das noções de HTML, JSON e programação para acessar sites e extrair dados.
- ▶ Selenium permite simular um navegador para interagir com elementos da página.

Interação com Elementos Dinâmicos

- ▶ Navegação entre páginas, interações com botões, campos de pesquisa e outros elementos dinâmicos.
- ▶ Importância de compreender JavaScript para interagir com elementos carregados dinamicamente.

Extração e Armazenamento dos Dados

- ▶ Extração e armazenamento dos dados coletados em JSON para uma estrutura organizada e acessível.
- ▶ JSON é amplamente utilizado por sua simplicidade e compatibilidade com várias linguagens.

Testes e Validação dos Scrappers

- ▶ Criação de rotinas de teste para verificar se os scrappers estão extraindo os dados esperados.
- ▶ Testes automatizados são importantes para garantir a robustez dos scrappers.

Manutenção dos Scrappers

- ▶ Como lidar com mudanças na estrutura dos sites que possam quebrar o scrapper.
- ▶ Atualização constante dos scrappers para acompanhar mudanças nos sites de interesse.

Monitoramento e Logs

- ▶ Dicas para monitorar o desempenho e registrar logs de execução.
- ▶ Logs ajudam a identificar erros e pontos de falha, facilitando a manutenção.