

Youtube Crawler

O Web Crawler criado extrai o título e o número de visualizações da primeira carga de resultados de vídeos do youtube ao buscar a frase "Chitãozinho e Xororó", mostra o resultado no console e ainda coloca em uma planilha chamada "chitaozinhoexororo.csv"

Para implementar o Web Crawler, foi utilizada a biblioteca Selenium para acessar o site, simulando o navegador Chrome e buscar os elementos desejados.

Na pasta "Crawler" está o arquivo csv gerado, o driver do navegador utilizado e o arquivo com a implementação do Crawler em Python. Basta executar o comando "python3 yt_crawler.py" no terminal para vê-lo funcionando.

Tempo de Execução

Para calcular o tempo de execução, foi adicionada a variável "execution_time", mas uma opção é adicionar o parâmetro time antes do comando de execução. O tempo médio foi de 8 segundos. Com o objetivo de diminuir esse tempo, foi mudado o driver do chrome para executar "headless", ou seja, sem precisar abrir a página. Assim, o tempo médio de execução passou a ser 6 segundos.

Uma alternativa é utilizar o Pypy, que é um compilador "just-in-time" que reduz o tempo de execução em relação ao Python ou o interpretador CPython.

Conteúdo gerado por Javascript vs Conteúdo independente

Quando a página do site é renderizada via Javascript, ou seja, o conteúdo é carregado dinamicamente, é necessário utilizar uma biblioteca ou framework que consiga executar esses scripts para assim localizar os elementos necessários da página desejada.

Já quando o site não utiliza Javascript, o código não precisa executar os scripts para localizar os elementos HTML, podendo percorrer diretamente as tags.