

Introdução

O que é web scraping?

Por que usar web scraping?

Introdução – O que é web scraping?

Na internet existem muitos dados espalhados em diversos sites.

Muitas vezes

Precisamos

Agrupar estes

Dados para obter

Informações importantes.



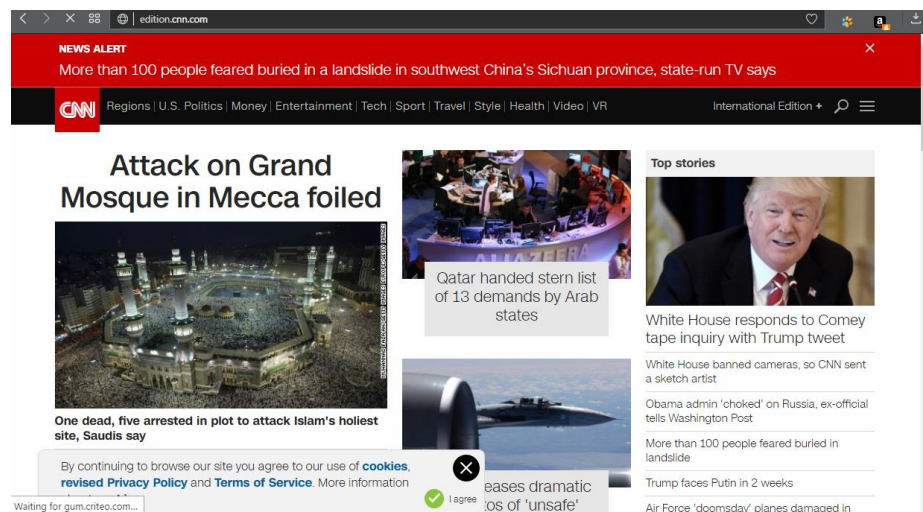
Introdução – O que é web scraping?

Web scraping (raspagem web) é a coleta automatizada de dados na Internet e a conversão destes dados em informação estruturada para posterior análise.



Introdução – O que é web scraping?

É a prática de coletar dados através de qualquer meio que não seja um programa interagindo com uma API, ou óbvio, por intermédio de uma pessoa acessando um navegador web.



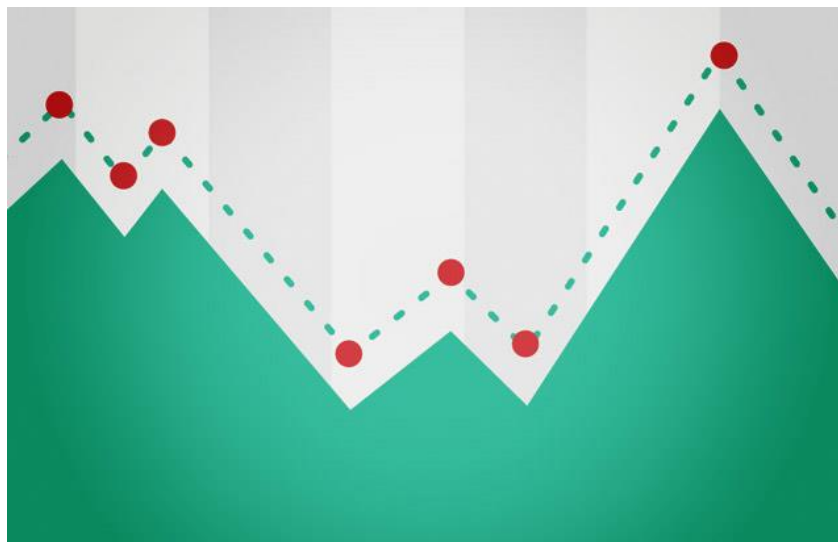
Introdução – O que é web scraping?

Fazemos web scraping normalmente escrevendo um programa que consulte um servidor web, solicite seus dados e analise-os para extrair as informações necessárias.



Introdução – O que é web scraping?

Web scraping engloba uma ampla variedade de técnicas e tecnologias de programação, como a análise de dados e a segurança das informações.



Introdução – O que é web scraping?

Por exemplo, podemos usar a coleta de dados para comparação de preços, pesquisas de mercado, coleta de dados governamentais e monitoramento de dados.



Introdução – Por que usar web scraping?

Em vez de ficar visualizando página por página no seu navegador, você pode visualizar bancos de dados repletos de informações de uma só vez.

Os web scrapers (programas para realização de web scraping) podem ir além dos lugares acessados por mecanismos de buscas como o Google.

O Google só mostra o que os sites dizem sobre suas páginas de conteúdo e não exhibe os resultados exatos das consultas que seriam inseridas em um aplicativo de busca. Um web scraper bem desenvolvido pode mostrar detalhes de informações reunidos de vários sites ao mesmo tempo.

Introdução – Por que usar web scraping?

Existem muitas APIs para coleta de dados, mas nem sempre encontramos uma que atenda às nossas necessidades. Existem APIs para muitos tipos de dados, como posts de Tweeter e páginas do Wikipédia.

Caso exista uma APIs para coleta de dados que esteja precisando, é melhor que use, mas, quando não encontrar, é melhor construir seu bot para obter os dados necessários.

Introdução – Por que usar web scraping?

Se você pode visualizar determinada informação em sites web, normalmente consegue acessar estas mesmas informações via script Python. Se puder acessar via script, pode armazenar estas informações em um banco de dados. Se puder armazenar em banco de dados, poderá fazer praticamente qualquer coisa com estes dados.

Podemos obter informações importantes em sites de notícias e fóruns por exemplo.

FIM