



Info Hacker

RESUMO QUINZENAL OFICIAL DA ACADEMIA HACKER



01: Web Scraping

02: Desafio de Programação



Web Scraping

por Gabriel Nicácio

O ato de copiarmos e colarmos informações de um website é, em uma escala bem pequena e manual, o que qualquer *web scraper* pode fazer com sua funcionalidade. *Web scraping* (também conhecido como *web data extraction* ou *data scraping*) é o processo de extrair dados estruturados da web de forma automatizada.

O método de *web scraping* é bastante utilizado por empresas, pessoas, negócios e organizações que buscam fazer uso da enorme quantidade de dados públicos disponíveis na web, seja almejando o monitoramento de preços, insights de notícias, geração de leads, pesquisas de marketing, entre muitos outros usos e objetivos para acelerar consultas e coleta de informações, tomando decisões mais inteligentes, com maior chance de acerto e sucesso.

Funciona por meio de duas partes, o *web crawler* e o *web scraper*. O *crawler*, também chamado comumente de "*spider*", é uma inteligência artificial que navega pela internet indexando, procurando e explorando conteúdos. Após isso, o *scraper*, sendo uma ferramenta especializada, recebe URLs passadas pela "*spider*" e extrai os dados com acurácia e rapidez, utilizando localizadores (ou seletores) que encontram os dados a serem extraídos do arquivo HTML. Uma ferramenta de *web scraping* faz requests HTTP para o website alvo e extrai os dados da página. Normalmente, ela analisa o conteúdo que está publicamente disponível para usuários acessarem e renderizado como HTML. É possível também fazer solicitações às APIs internas para alguns dados associados, que são armazenados em um banco de dados e entregues a um navegador por meio de solicitações HTTP.

Existem vários tipos de ferramentas de *web scraping*, elas podem ser personalizadas para servir a diferentes tipos de projetos de extração. *Web scraping* não é ilegal, porém, torna-se ilegal a partir do momento que ocorre a extração de dados que não estão disponíveis publicamente.



Desafios de Programação



Em Python qual a saída dos trechos de códigos abaixo?

```
>>>numeros = [1, 2, 3, 5, 7, 11]  
>>>print(numeros[-1])
```

- a) 2
- b) undefined
- c) 1
- d) 11
- e) SyntaxError

```
>>>print((2).__add__(4))
```

- a) SyntaxError
- b) 6
- c) [2, 4]
- d) 2
- e) 4

Info Hacker - Edição 011

12/07/2022



Autores:

Gabriel Nicácio

Revisão

Everton Borges

Bruno Severo



acha.ufal

Edições Anteriores

<https://github.com/Academia-Hacker/Info-Hacker>