WHAT IS WEB SCRAPING

BASICS & PRACTICAL USES

Web scraping is a way to download data from web pages.

WEB SCRAPING 101

Web scraping works as a dedicated data collector helping fetch data (like emails, phone numbers, articles, etc.) from web pages and organize it into certain formats like Excel, CSV, or HTML, etc. for your download and further use.

WHAT'S THE POINT OF WEB SCRAPING?

In the context of big data collection, web scraping is the rescue. Its nature of automation extremely improves people's working efficiency and eliminates human errors.

HOW DOES WEB SCRAPING WORK?

Web scraping finds the right data according to where it locates in the HTML source code and takes a series of actions just like a human. Once the data is ready, you will be able to download it from the cloud or to the local file.

WHO IS USING WEB SCRAPING?

Web scraping is widely used by professionals in industries like:

- jobs & recruitment
- consultancy
- hotel & travel
- eCommerce & retailing
- finance and more
- marketing

WHAT ARE THE MOST SCRAPED WEBSITES?

Websites like Amazon, eBay, Walmart, Yelp, Yellowpages, Craigslist, social media platforms like Facebook, Twitter, and LinkedIn are among the popular. The data people are getting from these sites includes:

- Online product details;
- Business/leads information;
- Discussions on social media or comments on the review pages.

Octoparse

THE PROS AND CONS OF WEB SCRAPING

THE ADVANTAGES OF WEB SCRAPING

HIGH SPEED

A web scraper will run for you relentlessly, getting all kinds of web data fast from different websites.

COST-EFFECTIVENESS

Web scraping is widely accepted by not only big companies but also SMBs. It just saves money.

COMPATIBILITY AND FLEXIBILITY

These advantages enables you to get data exactly in the form of how you need it.

CAPACITY TO GET DATA THAT API CAN'T

A web scraper can ignore the limit of API and extract what data you can see on the browser.

THE DISADVANTAGES OF WEB SCRAPING

START WITH A LEARNING CURVE

A web scraping tool takes time to get familiar with because you have to understand the basic idea of how it works so that you can command the tool.

YOUR IP MAY BE BLOCKED

IP blocking is likely to happen when you scrape websites that are protected with strict anti-scraping techniques.

WEB SCRAPER REQUIRES MAINTENANCE

If the website you are scraping has changed its structure, you then have to amend your scraper to adjust to the change.