

[Open in app](#)[Get started](#)

Shubham Kshatriya

[Follow](#)Jun 19, 2020 · 3 min read · [Listen](#)

WEB SCRAPING

WHAT | WHY | HOW

The Internet is very Gigantic. There are uncountable websites over the Internet and it contains enormous amount of invaluable data. Stock prices, Product details, Stats, Company contacts, etc whatever you name it. If you ever wanted to retrieve these information, you may have done some copy-paste manually or followed the website format. This is where Web Scraping steps in.

What is Web Scraping?

Web Scraping, also referred as Web Data Extraction, Web Harvesting, is a process of “scraping” or retrieving data from a website. Unlike the manual process, Web Scraping uses automation tools to retrieve huge amount of data within a fraction of time. Though Scraping can be done manually but in most cases, automation tool is always preferred because of less efforts and faster results.



Image Courtesy: Google

Web scraping software may access the World Wide Web directly using the HTTP, or



[Open in app](#)[Get started](#)

Why do we need Web Scraping?

With the overwhelming data available on the internet, web scraping has become the essential approach to aggregating Big Data sets. The application of Web Scraping is increasing drastically day by day. Let me list few of its application: *Price Monitoring, Market Research, Real Estate, Content Monitoring and way more...*

The statistics below are based on the information collected from LinkedIn. The top 10 industries that have the highest demand for web scraping skills are: Computer Software (22%); Information Technology and Services (21%); Financial Services (12%); Internet (11%); Marketing and Advertising (5%); Computer & Network Security (3%); Insurance (2%); Banking (2%); Management Consulting (2%); Online Media(2%).

When you post a link or image on Facebook/ Instagram or any other social platform, the information around it is scraped. Internet without web scraping would be absolutely dead, it would never have become the huge place it is now. See, that's how this trend is getting over every technical aspects that you may know or don't have any idea of. So, what could stop anyone to get into this amazing trend?

How to get started with Web Scraping?

There are plenty of Automation tools better known as Web Scraper which does the task of scraping any website. And for the question, which one to use, I would say *it depends*. There are plenty of Python libraries like **Beautiful Soup**, **Selenium**, **LXML**, etc which helps to develop one's very own Scraper or you may use some existing tools like **Scrapy**, **Octoparse**, **ParseHub** and many more for the same. Choice is yours whether you want to learn Web Scraping and develop our own code and get along with this art or make use of existing tools and serve your purpose.



[Open in app](#)[Get started](#)

variety of laws may apply to unauthorized scraping, including contract, copyright and trespass to chattels laws. Most of the websites state whether you can Scrap them or not in their Terms of Services(ToS). You can find whether you are allowed to scrap a website or not by going through their Terms of Services.

Web scraping of public pages might be legal(Don't try to scrap any of them unless you are allowed to do so). Web scraping by intruding into private pages, without obtaining their prior written permission, or in disregard of their Terms of Service is totally illegal.

While it might sound intimidating, web scraping is way easier than you think. It will take very little effort and you can master this art quite easily.

Happy Scraping!

Source: Wikipedia, Google.

