

Step 1:

Web scraping is **the process of using bots to extract content and data from a website**. Unlike screen scraping, which only copies pixels displayed onscreen, web scraping extracts underlying HTML code and, with it, data stored in a database. The scraper can then replicate entire website content elsewhere.

<https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>

What is Web Scraping and How to Use It?

Last Updated : 16 Jul, 2024

•

Suppose you want some information from a website. Let's say a paragraph on Donald Trump! What do you do? Well, you can copy and paste the information from Wikipedia into your file. But what if you want to get large amounts of information from a website as quickly as possible? Such as large amounts of data from a website to train a **Machine Learning algorithm**? In such a situation, copying and pasting will not work! And that's when you'll need to use **Web Scraping**. Unlike the long and mind-numbing process of manually getting data, Web scraping uses intelligence automation methods to get thousands or even millions of data sets in a smaller amount of time.

Table of Content

- [What is Web Scraping?](#)
- [How Web Scrapers Work?](#)
- [Types of Web Scrapers](#)
- [Why is Python a popular programming language for Web Scraping?](#)
- [What is Web Scraping used for?](#)



If you are coming to a sticky end while trying to collect public data from websites, we have a solution for you. [Smartproxy](#) is a tool that offers a solution to deal with all the hurdles with a single tool. Their formula for scraping any website is: 40M+ pool of residential and data center proxies + powerful web scraper = [Web Scraping API](#). This tool ensures that you get the needed data in raw HTML at a 100% success rate.

With Web Scraping API, you can collect real-time data from any city worldwide. You can rely on this tool even when scraping websites built with JavaScript and won't face any hurdles.

Additionally, Smartproxy offers four other scrapers to fit all your needs – enjoy eCommerce, SERP, Social Media Scraping APIs and a No-Code scraper that makes data gathering possible even for no-coders. Bring your data collection process to the next level from \$50/month + VAT.

But before using Smartproxy or any other tool you must know what web scraping actually is and how it's done. So let's understand what Web scraping is in detail and how to use it to obtain data from other websites.

What is Web Scraping?

Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications. There are many different ways to perform web scraping to obtain data from websites. These include using online services, particular API's or even creating your code for web scraping from scratch. **Many large websites, like Google, Twitter, Facebook, StackOverflow, etc. have API's that allow you to access their data in a structured format.** This is the best option, but there are other sites that don't allow users to access large amounts of data in a structured form or they are simply not that technologically advanced. In that situation, it's best to use Web Scraping to scrape the website for data.

Web scraping requires two parts, namely the **crawler** and the **scraper**. **The crawler is an artificial intelligence algorithm that browses the web to search for the particular data required by following the links across the internet. The scraper, on the other hand, is a specific tool created to extract data from the website.** The design of the scraper can vary greatly according to the complexity and scope of the project so that it can quickly and accurately extract the data.

How Web Scrapers Work?

Web Scrapers can extract all the data on particular sites or the specific data that a user wants. Ideally, it's best if you specify the data you want so that the web scraper only extracts that data quickly. For example, you might want to scrape an Amazon page for the types of juicers available, but you might only want the data about the models of different juicers and not the customer reviews.

So, when a web scraper needs to scrape a site, first the URLs are provided. Then it loads all the HTML code for those sites and a more advanced scraper might even extract all the CSS and Javascript elements as well. Then the scraper obtains the required data from this HTML code and outputs this data in the format specified by the user. Mostly, this is in the form of an Excel spreadsheet or a CSV file, but the data can also be saved in other formats, such as a JSON file.

Types of Web Scrapers

Web Scrapers can be divided on the basis of many different criteria, including Self-built or Pre-built Web Scrapers, Browser extension or Software Web Scrapers, and Cloud or Local Web Scrapers.

You can have **Self-built Web Scrapers** but that requires advanced knowledge of programming. And if you want more features in your Web Scraper, then you need even more knowledge. On the other hand, pre-built **Web Scrapers** are previously created scrapers that you can download and run easily. These also have more advanced options that you can customize.

Browser extensions Web Scrapers are extensions that can be added to your browser. These are easy to run as they are integrated with your browser, but at the same time, they are also limited because of this. Any advanced features that are outside the scope of your browser are impossible to run on Browser extension Web Scrapers. But **Software Web Scrapers** don't have these limitations as they can be downloaded and installed on your computer. These are more complex than Browser web scrapers, but they also have advanced features that are not limited by the scope of your browser.

Cloud Web Scrapers run on the cloud, which is an off-site server mostly provided by the company that you buy the scraper from. These allow your computer to focus on other tasks as the computer resources are not required to scrape data from websites. **Local Web Scrapers**, on the other hand, run on your computer using local resources. So, if the Web scrapers require more CPU or RAM, then your computer will become slow and not be able to perform other tasks.

Why is Python a Popular Programming Language for Web Scraping?

[Python](#) seems to be in fashion these days! It is the most popular language for web scraping as it can handle most of the processes easily. It also has a variety of libraries that were created specifically for Web Scraping. [Scrapy](#) is a very popular open-source web crawling framework that is written in Python. It is ideal for web scraping as well as extracting data using APIs. [Beautiful soup](#) is another Python library that is highly suitable for Web Scraping. It creates a parse tree that can be used to extract data from HTML on a website. Beautiful soup also has multiple features for navigation, searching, and modifying these parse trees.

What is Web Scraping Used for?

Web Scraping has multiple applications across various industries. Let's check out some of these now!

1. Price Monitoring

Web Scraping can be used by companies to scrap the product data for their products and competing products as well to see how it impacts their pricing strategies. Companies can use this data to fix the optimal pricing for their products so that they can obtain maximum revenue.

2. Market Research

Web scraping can be used for market research by companies. High-quality web scraped data obtained in large volumes can be very helpful for companies in analyzing consumer trends and understanding which direction the company should move in the future.

3. News Monitoring

Web scraping news sites can provide detailed reports on the current news to a company. This is even more essential for companies that are frequently in the news or that depend on daily news for their day-to-day functioning. After all, news reports can make or break a company in a single day!

4. Sentiment Analysis

If companies want to understand the general sentiment for their products among their consumers, then Sentiment Analysis is a must. Companies can use web scraping to collect data from social media websites such as Facebook and Twitter as to what the general sentiment about their products is. This will help them in creating products that people desire and moving ahead of their competition.

5. Email Marketing

Companies can also use Web scraping for email marketing. They can collect Email ID's from various sites using web scraping and then send bulk promotional and marketing Emails to all the people owning these Email ID's.