

Python Scrapy

- Scrapy is a Python framework for large scale web scraping. It gives you all the tools you need to efficiently extract data from websites, process them as you want, and store them in your preferred structure and format.
- Scrapy uses spiders, which are self-contained crawlers that are given a set of instructions. In Scrapy it is easier to build and scale large crawling projects by allowing developers to reuse their code.

Scrapy Vs. BeautifulSoup

- BeautifulSoup is used for simple scraping jobs with efficiency. It is slower than Scrapy if you do not use multiprocessing.
- Scrapy can get big jobs done very easily. It can crawl a group of URLs in no more than a minute depending on the size of the group and does it very smoothly as it uses Twister which works asynchronously (non-blocking) for concurrency.
- If the project does not require much logic, BeautifulSoup is good for the job, but if you require much customization such as proxies, managing cookies, and data pipelines, Scrapy is the best option.

Google Chrome - Web Scraper Extension

Web Scraper is a simple web scraping tool that allows you to use many advanced features to get the exact information you are looking for. It offers features like:

- Data scraping from multiple pages;
- Multiple data extraction types (text, images, URL's, and more);
- Scraping data from dynamic pages (JavaScript + AJAX, infinite scroll);
- Browsing scraped data;
- Exporting scraped data from a website to Excel