Web Crawling:

web crawling refers to the process of discovering the links or URLs on the web.

Python Libraries for WEB Crawling:

Library	GitHub Stars	Key Features	License
Beautiful Soup	84	HTML/XML parsing Easy navigation Modifying parse tree	MIT
Requests	50.9K	Custom headers Session objects SSL/TLS verification	Apache 2.0
Scrapy	49.9K	Web crawling Selectors Built-in support for exporting data	3-Clause BSD
Selenium	28.6K	Web browser automation Supports major browsers, Record and replay actions	Apache 2.0
Playwright	58.3K	Automation for modern web apps Headless mode Network manipulation	Apache 2.0
Lxml	2.5K	High-performance XML processing XPath and XSLT support Full Unicode support	BSD
Urllib3	3.6K	HTTP client library Reusable components SSL/TLS verification	MIT
MechanicalSoup	4.5K	Beautiful Soup integration Browser-like web scraping Form submission	MIT

web crawling tools to scrape websites:

1. Bright Data
Bright Data Web Scraper is designed for developers and consists of ready-made web scraper templates that help to focus on multi-step data collection from the browser.

2.Oxylabs Scraper API
It is designed to collect large volumes of real-time public data from any web page. It helps in providing market research, SEO monitoring, fraud protection, and so on. They provide structured and valuable data to the people and also eliminate the requirement of individual research.

3. Apify
Apify is the most powerful no-code, open-source proxy management web scraping and automation tool which is used for data extraction from social media, mobile apps, web pages, and e-commerce pages, from the API's.

4. Smartproxy
Smartproxy consists of many scraping APIs that are used in e-commerce, social media, and web scraping. They provide client access to any number of exit nodes therefore the users are unlikely to lose access to the required data which they need.

5. ParseHub
ParseHub is a powerful scraping tool that is used for the extraction of online data and is also used to scrape and download images in JSON and CSV files. Parsehub has more useful features than the other scraping tools. They get the data from the tables and maps.

6. Scrape. do
Scrape.do is a web scraping tool that is used to provide fast and scalar web scraper API in an endpoint. They used rotating proxies which allowed them to scrape any websites to extract the data. They are also the super proxy parameter which allows to extract data with protection.They allows the website pages to render javascript.

7. Octoparse
Octoparse is known as the best web crawler which is a client based tool used to get the data into the spreadsheets. It is built for non coders. They have a site parser solution for the users who want to run scarpers in the cloud. There are two types of operation mode in Octoparse such as Wizard mode and advanced mode.

8. Scrapy

Scrapy is an open source free of cost web scraping library, therefore it is a complete web crawling framework which is used by the python developers. Scrapy helps to handle the functions which are used to build web crawlers. They are used for data mining and automated testing

9. Mozenda

Mozenda is a high scalable cloud based self serve web scraping platform which boasts the enterprise of customers all over the world. This tool allows the users to view the report and run it where the data has been collected. It automatically detects the information organised in lists on the website pages and also allows the user to build agents which collect this data.

10. Scraper API

Scraper API is used for handling the web browsers, CAPTCHAs and proxies. It is designed by designers to make web scraping at scale as simple as it can be by rotating proxy pools, solving the CAPTCHAs, detecting bans and managing geotargeting.