

Python libraries and tools used for web crawling:

Web Crawling -

It is a process that allows the search engine to match, with the use of indexed data, relevant search results to a query

Python Libraries for Web Crawling

Library

Beautiful Soup - HTML/XML parsing, Easy navigation, Modifying parse tree

Requests - Custom headers, Session objects, SSL/TLS verification.

Scrapy - Web crawling, Selectors, Built-in support for exporting data

Selenium - Web browser automation, Supports major browsers, Record and replay actions

Playwright - Automation for modern web apps, Headless mode, Network manipulation

Lxml - High-performance XML processing, XPath and XSLT support, Full Unicode support

Urllib3 - HTTP client library, Reusable components, SSL/TLS verification

MechanicalSoup - Beautiful Soup integration, Browser-like web scraping, Form submission

Web Crawling Tools

Tool

Bright Data - Web Scraper designed for developers with ready-made templates for multi-step data collection.

Oxylabs Scraper API - Collects large volumes of real-time public data, useful for market research, SEO monitoring, fraud protection. Provides structured and valuable data.

Apify - No-code, open-source proxy management tool for data extraction from social media, mobile apps, web pages, and e-commerce pages.

Smartproxy - Provides scraping APIs for ecommerce, social media, and web scraping. Access to numerous exit nodes to ensure data availability.

ParseHub - Extracts online data, including images, into JSON and CSV files. Offers more features than other tools, such as scraping data from tables and maps.

Scrape.do - Fast and scalable web scraper API using rotating proxies for data extraction. Supports JavaScript rendering on web pages.

Octoparse - Client-based tool for data extraction into spreadsheets, built for non-coders. Offers Wizard and Advanced modes for scraping operations.

Scrapy - Open-source web scraping library, used for data mining and automated testing.

Mozenda - Cloud-based self-serve web scraping platform, detecting and organising data from lists on web pages. Allows building agents for data collection and report viewing.

Scraper API - Handles web browsers, CAPTCHAs, and proxies, making large-scale web scraping simple with features like rotating proxies and geotargeting.