Scraping Tools

1. Beautiful Soup: Beautiful Soup is a Python library that focuses on parsing and navigating HTML or XML documents. It provides a convenient interface for extracting data from web pages by traversing the parsed document using methods like searching by tag name, CSS selectors, or XPath expressions. Beautiful Soup supports various parsers, including the built-in Python parsers and external libraries like lxml and html5lib.

1. Scrapy: Scrapy is a powerful Python framework for building web spiders or crawlers. It provides a comprehensive solution for handling the entire web scraping process, including sending HTTP requests, managing cookies and sessions, following links, parsing responses, and storing extracted data. Scrapy uses its own selector system called Selectors, which allows you to extract data using CSS or XPath expressions.

1. Selenium: Selenium is primarily known as a web automation and testing tool, but it can also be used for web scraping. It enables you to control web browsers programmatically, including interacting with JavaScript-rendered content. Selenium supports multiple programming languages, including Python, Java, C#, and JavaScript, and provides a rich set of functions to simulate user interactions, fill out forms, and capture data from dynamic websites.

1. Puppeteer: Puppeteer is a Node.js library developed by Google that provides a high-level API for controlling headless Chrome or Chromium browsers. It allows you to automate browser actions, capture screenshots, generate PDFs, and extract data from web pages. Puppeteer can be particularly useful for scraping websites that heavily rely on JavaScript rendering or require user interactions for data retrieval.

1. Octoparse: Octoparse is a Windows-based desktop application that offers a visual approach to web scraping. It allows users to navigate websites using a built-in browser, select the data they want to extract using point-and-click actions, and configure scraping rules. Octoparse handles various complexities, such as pagination, AJAX requests, and login authentication.

1. ParseHub: ParseHub is another visual web scraping tool that allows users to build scraping projects without writing any code. It offers a point-and-click interface to identify and select data on web pages, handle pagination and AJAX, and export the extracted data in various formats like CSV, Excel, or JSON. ParseHub also provides scheduling options for automated scraping tasks.

When using these tools, it's important to be mindful of legal and ethical considerations. Ensure that you comply with the website's terms of service, respect their data usage policies, and avoid excessive scraping that could potentially harm the website's performance or violate their guidelines.

Additionally, it's worth noting that websites may employ various anti-scraping techniques to prevent or detect scraping activities. These techniques can include CAPTCHAs, rate limiting, IP blocking, and obfuscation of data. It's important to consider these measures and be respectful of a website's protection mechanisms when engaging in web scraping activities.

Video: <https://youtu.be/bTySEwaGc2k>

Non-code scrayper:

Apify: [http://bit.ly/40Az0tE](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbDdFanJ6UUVWWGRfSGZQT3EzT0xVcC16a2R1Z3xBQ3Jtc0tteU1qejJFdUVXdFJfM3JBOUxOWjVNSXU1cjF6YzBxRWY0YTlOLVRoRXJQTFFueExQNVZqd1hXSVE0RHVRcXJvbHNvczhvN1BfOTZCUUxBbWlfVi1QajJJTWNrUE94WFlCX09FOGpnemNVdkk0aWtNWQ&q=http%3A%2F%2Fbit.ly%2F40Az0tE&v=bTySEwaGc2k)

ParseHub: [http://bit.ly/3LTFLTa](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqazNwRTFDTmFoNjZLalMwVFplY2JZRDBaallGZ3xBQ3Jtc0tsYmZURUt3OVctdlBiYU9ULThqM2VqNjZlNkswSl9LQ1NXNUMzZjBCMTBLNE5DLVdtTG8zOE1pNlZMQ2V6MzhiOTZGbXhpSFZ2ZDlyZ1BZZ3VVZnRTdDFBRXlWQUw4dFJQcUFpMi0tSnl0WTB5dmE0UQ&q=http%3A%2F%2Fbit.ly%2F3LTFLTa&v=bTySEwaGc2k)

Smartproxy’s No-Code Scraper: [https://bit.ly/3W13vpL](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbnlkUHVCelhYbEQ4c2hVQ2VEUTk3X2I2TmZIQXxBQ3Jtc0tuX1Q4ck01ME5JdDU3cnVZTU9zLUJmblJTc1NpNUZMRF9ESmJtVmRuNHZwMG80VFNIWG81Z2tPUTl1SGNBUi1KUlJSMDZRTDBPU3I3d1p4ektvR09FOWFKNzlfQmhSMjV1blBQX0JvN04zRDc2VHJGUQ&q=https%3A%2F%2Fbit.ly%2F3W13vpL&v=bTySEwaGc2k) [02:51](https://www.youtube.com/watch?v=bTySEwaGc2k&t=171s)

Oxylabs: [https://bit.ly/3pC4pN3](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbjFaWGt3djBGczFBTjFRUzI4SXJnd3ZUdDlHZ3xBQ3Jtc0tteFV0dGFSVk1xUG5KNlpBXzZXUW9BdkpDV09OZlRTWVF3bWNudWYtNUpQMkNPeDlycnFNbXA0V2pCTDRCUmlQVUdNLWJmWmJPOVNoM19sanlsbU04c09aamV5aUdSSzlzWUNaNGYtZ251b09Pb29VZw&q=https%3A%2F%2Fbit.ly%2F3pC4pN3&v=bTySEwaGc2k)

Smartproxy: [http://bit.ly/3lQ6uFM](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa19OdGxKNVhDVElUU3NkRUU4OWZPcFI1VkN5Z3xBQ3Jtc0ttQVlZcGVzNW93U0hqVlhXZ1o0R3M0WkE4ZXI1OTFOVzk4NGJTOG1hVjFyTUNnSlNCZmotaTg4Mk5OOWV0UzB2X2lpV0J5UGJOeXpTVnJxVkVRQWFRbS1jM2pyZ1dFNTVYSmdLNGFFWDN0NXlLbTlLaw&q=http%3A%2F%2Fbit.ly%2F3lQ6uFM&v=bTySEwaGc2k)

ScraperAPI: [https://bit.ly/3IzrqYN](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbEZxNzg4SFlZVUNrNUNKUFVpYWpob2NQQ1YwUXxBQ3Jtc0tsdXNuVmthek1EMmF0ZDc0UjBNRDFSVDVtcXBObjNCLVZVdGZ3R0ptbEpDcUs3V1hrMEJRRGs2TGJPSzMtNmtLWWQtMDhkZW0xN2M3ejdmSEpWSlRLc29MNWI3aVVaeG80OWFDd3YyWnUxc3NkY0JYNA&q=https%3A%2F%2Fbit.ly%2F3IzrqYN&v=bTySEwaGc2k)

WEB SCRAPING LIBRARIES:

Requests: [https://pypi.org/project/requests/](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbEZPaUxtS0dCSGF0QlVjXzdsbzhSSVo4aDVmZ3xBQ3Jtc0tuaS03RmxYcUJyNGR4dThoWU9id3BFNG40Tklhai1wR05SQkRKekRUMkFCV0R0c2U4bGUtblVqWTJLR1RlTU5UQ3RpblhtWFhORmQ5OFF6ZnI4V0dCOWVmcDJnZ01JVnRxZG5oTzFPanltRUdtWE5Qbw&q=https%3A%2F%2Fpypi.org%2Fproject%2Frequests%2F&v=bTySEwaGc2k)

Beautiful Soup: [https://pypi.org/project/beautifulsoup4/](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbnJPNFpOc0F4YjNjVmxQdVUzLUdib1VNMWxNd3xBQ3Jtc0ttalRTdU1jQmswTHo5d0hZcGpKVVpockhmeVgyMXVQcTlIdXRWZjE5T1B0Nnh2a01qODhEWEpLd1FJSXJKWDM3QkNqLVo5MHpwOW5lckx3OVlyTm5TQ192em1jRDN0RDczWVlxNVZGUVRCOFhXcm9QMA&q=https%3A%2F%2Fpypi.org%2Fproject%2Fbeautifulsoup4%2F&v=bTySEwaGc2k)

Selenium: [https://www.selenium.dev/](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqa2JzNlo3dEYwVllsckpMMEQtVHNhaGplcS1VZ3xBQ3Jtc0tsMTJjc1VWY2pkTEtqdEw2X3FZVk1NY0tvQk9EYU5UUkJyOFNDS3pkc29HbExPYnhSdFIyZnFIOFpQQ0hpSFc2Q191ckZDeGdCNzJETWRrMWZkZTIyM0g0QVgxQVpYRGNsa2F2cXhmV3k1X0wzQ3c1TQ&q=https%3A%2F%2Fwww.selenium.dev%2F&v=bTySEwaGc2k)