Web scraping is used to collect large information from websites.

let’s look at the applications of web scraping:

Price Comparison: Services such as Parse Hub use web scraping to collect data from online shopping websites

and use it to compare the prices of products.

Email address gathering: Many companies that use email as a medium for marketing,

use web scraping to collect email ID and then send bulk emails.

Social Media Scraping: Web scraping is used to collect data from Social Media websites

such as Twitter to find out what’s trending.

Research and Development: Web scraping is used to collect a large set of data

(Statistics, General Information, Temperature, etc.) from websites,

which are analysed and used to carry out Surveys or for R&D.

Job listings: Details regarding job openings, interviews are collected from different websites

and then listed in one place so that it is easily accessible to the user.

How does Web Scraping work?

When you run the code for web scraping, a request is sent to the URL that you have mentioned. As a response to the request, the server sends the data and allows you to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it.

To extract data using web scraping with python, you need to follow these basic steps:

1. Find the URL that you want to scrape
2. Inspecting the Page
3. Find the data you want to extract
4. Write the code
5. Run the code and extract the data
6. Store the data in the required format

Web scraping is an automated method used to extract large amounts of data from websites.

The data on the websites are unstructured.

Web scraping helps collect these unstructured data and store it in a structured form.

There are different ways to scrape websites such as online Services, APIs or writing your own code.

In this article, we’ll see how to implement web scraping with python.

we will be using the following libraries:

* **Selenium**:  Selenium is a web testing library. It is used to automate browser activities.
* **BeautifulSoup**: Beautiful Soup is a Python package for parsing HTML and XML documents. It creates parse trees that is helpful to extract the data easily.
* **Pandas**: Pandas is a library used for data manipulation and analysis. It is used to extract the data and store it in the desired format.