**Web Scraping**



**What is Web scraping?**

**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.**

**Detail information :**

Web scraping refers to theprocess of extracting content and data This information is collected and then exported into a format that is more useful for the user. The process of scraping can be done manually but it takes massive efforts which makes it impractical (e.g., costly, slow, human error, etc.).

This is where automated web scraping comes into play, delivering efficiency and accuracy in data extraction.

Let’s say, data is vital for your e-commerce company. You are able to see the data on your competitor’s website.

The question is how will you download it in a usable format?

Most people would be able to only copy and paste it manually. However, it is not feasible to do it for large websites with hundreds of pages.

This is where web scraping comes into play.

Web scraping is a process of automating the extraction of data in an efficient and fast way. With the help of web scraping, you can extract data from any website, no matter how large the data is, on your computer.

Moreover, websites may have data that you cannot copy and paste. Web scraping can help you extract any kind of data that you want.

That’s not enough. Let’s say, you copy and paste some data but how to convert or save it in a format of your choice?

Web scraping takes care of that too. When you extract web data with the help of a web scraping tool, you would be able to save the data in a format such as CSV. You would then be able to retrieve, analyze and use the data the way you want.

So web scraping simplifies the process of extracting data, speeds it up by automating it and creates easy access to the scraped data by providing it in a CSV format.

In simple terms, web scraping saves you the trouble of manually downloading or copying any data and automates the whole process.

Some websites can contain a very large amount of invaluable data

Stock prices, product details, sports stats, company contacts, you name it.

If you wanted to access this information, you’d either have to use whatever format the website uses or copy-paste the information manually into a new document. Here’s where web scraping can help.

**What is the need of data scraping?**

Well, “**data**” is the new differentiator nowadays. You can derive strategies and insights from the data. It is the core of market research and business strategies

Whether you want to start a new project or churn out a new strategy for an existing business, you need to invariably access and analyze a vast amount of data. This is where web scraping comes in.

Following is the list of applications of web scraping in different industries:

* *Comparison Shopping Sites / Competitor price monitoring* :

In this era of e-commerce, price plays a key role. You need to keep track of competitor pricing strategies.

We cannot manually keep track of the price change, because prices keep on changing every now and then .Therefore, it becomes virtually impossible to keep track of prices in a manual way.

Several websites and applications can help you to easily compare pricing between several retailers for the same product.

One way that these websites work is by using web scrapers to scrape product data and pricing from each retailer daily. This way, they can provide their users with the comparison data they need.

This is where web scraping comes in. It automates the process of extracting prices of your competitors and keeps you updated about the new pricing strategies deployed by your competitors. Web scraping is able to do it in no time and as many times as you want.

* *Monitoring Consumer Sentime*nt :

It is necessary to track and analyze consumer sentiment. It is done by reviewing consumer feedback and reviews of different businesses.

However, getting all the reviews from different websites in a manual way is not feasible.

Therefore, web scraping is leveraged to make it stunningly easy. With web scraping, you can get all the reviews on a spreadsheet and even compare different reviews based on keywords.

* *News and Reputation Monitoring :*

In order to keep track of the information regarding individuals, products or company, news scraping is quite useful.

Web scraping is integral to the process because it allows quick and efficient extraction of data in the form of news from different sources.

Such data can then be processed in order to glean insights as required. As a result, it also makes it possible to keep track of the brand and reputation of a company.

**Initial steps of web scraping:**

**Step 1:** setting up the virtual enviornment for that install the following existing libraries of python

1. Pip install requests
2. Pip install html5lib
3. Pip install bs4 (beautiful soup)

**Step 2:** Using the request module we will get the html content of the website as a string and parse the html content and give it a tree like structure so that it can traverse.