

Web Scraping using BeautifulSoup

Before we start with Web Scraping with BeautifulSoup, let us understand:

What is mean by Web Scraping and how we can achieve it?

Web scraping is a data extraction method used to exclusively gather data from websites. It is widely used for Data mining or collecting valuable insights from large websites. Web scraping comes in handy for personal use as well. Python contains an amazing library called **BeautifulSoup** to allow web scraping.

Or in other words, Web scraping is an automated way of extracting large chunks of data from websites which can then be saved on a file in your computer or accessed on a spreadsheet.

So basically, Web Scraping is a method that let user to extract/acquire any data from any websites that available on the World Wide Web.

Why we do Web Scraping?

By using this method, user can gain any information that needed from the data! You can do things like Internet of Things Analytics, Business Analyst.

This method is generally used to predicting the future of that item. In Data Analytics language, the term we used is “Prediction of Data.”

How we start or execute Web Scraping?

First of all of everything, we need to import required libraries of Python in our Jupyter Notebook, Google Colab, VS Code or PyCharm. The following libraries and their importance are mentioned below:

1. BeautifulSoup4: Used to parse HTML or XML pages from the website.
2. Pandas: Used for dataframes.
3. NumPy: Used for arrays and scientific calculation, generally numbers.
4. Matplotlib: Used for creating high-quality visualization and graphs.

If the above, libraries are not installed. This is the way to install it.

1. For BeautifulSoup, use script:

!pip install beautifulsoup4

2. For Pandas, use script:

!pip install pandas

3. For NumPy, use script:

!pip install pandas

4. For Matplotlib, use script:

!pip install matplotlib