

# Web Scrapping: A Key Tool in Data Science

Estimated Effort: 5 mins

## Introduction

Web scrapping, also known as web harvesting or web data extraction, is a technique used to extract large amounts of data from websites. The data on websites is unstructured, and web scrapping enables us to convert it into a structured form.

## Importance of Web Scrapping in Data Science

In the field of data science, web scrapping plays an integral role. It is used for various purposes such as:

1. **Data Collection:** Web scrapping is a primary method of collecting data from the internet. This data can be used for analysis, research, etc.
2. **Real-time Application:** Web scrapping is used for real-time applications like weather updates, price comparison, etc.
3. **Machine Learning:** Web scrapping provides the data needed to train machine learning models.

## Web Scrapping with Python

Python provides several libraries for web scrapping. Here are some of them:

1. **BeautifulSoup:** BeautifulSoup is a Python library used for web scrapping purposes to pull the data out of HTML and XML files. It creates a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner.

```
1. 1
2. 2
3. 3
4. 4
5. 5

1. from bs4 import BeautifulSoup
2. import requests
3. URL = "http://www.example.com"
4. page = requests.get(URL)
5. soup = BeautifulSoup(page.content, "html.parser")
```

Copied!

2. **Scrapy:** Scrapy is an open-source and collaborative web crawling framework for Python. It is used to extract the data from the website.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7

1. import scrapy
2. class QuotesSpider(scrapy.Spider):
3.     name = "quotes"
4.     start_urls = ['http://quotes.toscrape.com/tag/humor/']
5.     def parse(self, response):
6.         for quote in response.css('div.quote'):
7.             yield {'quote': quote.css('span.text::text').get()}
```

Copied!

3. **Selenium:** Selenium is a tool used for controlling web browsers through programs and automating browser tasks.

```
1. 1
2. 2
3. 3

1. from selenium import webdriver
2. driver = webdriver.Firefox()
3. driver.get("http://www.example.com")
```

Copied!

## Applications of Web Scrapping

Web scrapping is used in various fields and has many applications:

1. **Price Comparison:** Services such as ParseHub use web scrapping to collect data from online shopping websites and use it to compare the prices of products.
2. **Email address gathering:** Many companies that use email as a medium for marketing, use web scrapping to collect email ID and then send bulk emails.
3. **Social Media Scrapping:** Web scrapping is used to collect data from Social Media websites such as Twitter to find out what's trending.

## Conclusion

Web scrapping is an essential skill in the fast-growing world of data science. It provides the ability to turn the web into a source of data that can be analyzed, processed, and used for a variety of applications. However, it's important to remember that one should use web scrapping responsibly and ethically, respecting the terms of use or robots.txt files of the websites being scraped.

## Author(s)



**Skills** Network