

Close X

You may also right-click on the following link to download and view the file: [click here](#)

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

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

1. **BeautifulSoup:** BeautifulSoup is a Python library used for web scraping 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 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")
```

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 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()}
```

