**JOBS DATABASE NOTES**

**-Mansi Umakant Dandgaval 002956418**

**Script for Web Scraping**

The web scraper scrapes job postings from Indeed.com and stores the links in a SQL database using a list of job keywords (such as finance or machine learning) and a list of cities (such as Boston).

**Version 2**

searches for unvisited job urls, gathers employment data from a particular job listing, and enters that data into a SQL database.

Script 1 for the Twitter Bot The twitter bot retrieves the json for a list of job keywords from a SQL database (for example, finance or machine learning) and saves the json files.

**Script 2 of Bot**

Removes duplicates, extracts pertinent information, and adds longitude and latitude before storing the resulting twitter data in a SQL database.

**3. Bot Script**

Subsets of the Twitter data are produced as csv files (e.g., machine learning tweets from June 1, 2022 to July 1, 2022)

**What does a database crawler do?**

A web crawler, also known as a crawler or web spider, is a computer program used to automatically search and index the content of websites and other online data. The most typical use of these software programs, or bots, is to add items to a search engine index.

**Web Scraping Datasets, Databases & APIs**

[**What is Web Scraping Data?**](https://datarade.ai/data-categories/web-scraping-data)

Web scraping data is data that is collected online with the help of web crawling tools. It's used by business owners to collect online data and website content, for example product price details. Web scraping can also work alongside NLP data to gauge social media sentiment.

**What is twitter bot in database?**

Twitter web scraping is the automated act of data collection. Data that might otherwise go unseen because of the sheer volume of information to be found on Twitter. The scraping tool itself parsing through HTML, or hypertext markup language, gathers its findings and compiles the data into one neat document.

## **What is** [**Scraping Bot and Web Scraping API**](https://www.octoparse.com/blog/basic-introduction-to-web-scraping-bot-and-web-scraping-api#:~:text=Scraper%20bots%20are%20tools%20or%20pieces%20of%20code,with%20a%20scraper%20bot%20is%20called%20web%20scraping.)**?**

Scraper bots are tools or pieces of code used to extract data from web pages. These bots are like tiny spiders that run through different web pages in a website to extract the specific data they were created to get. The process of extracting data with a scraper bot is called web scraping.

**Can you scrape content from every website?**

Scraping increases website traffic and increases the risk of a website server failure. As a result, not all websites permit scraping. How do you determine whether a website is permitted or not? You can view the website's "robots.txt" file. You can see information on whether the website host permits you to scrape the website by adding robots.txt after the URL that you want to collect data from.

**What is the process of web scraping?**

Web scraping simply involves a bot person browsing various website pages and copying and pasting all the contents. The data is contained in the response you receive from the server after the code has been run. The following step is to parse the response data and extract the relevant information.

**Various web scraping tools and libraries:**

There are numerous options for scraping tools that don't require coding. Because it is simple to use and has a large user base, the Python library is still the most popular choice for web scraping.

* Python web scraping libraries such as Beautiful Soup, Requests, and Selenium are the most frequently used ones.
* You can parse HTML or XML documents using Beautiful Soup to create a readable format. It enables you to search different areas of the documents and makes it easier for you to quickly find the information you need.
* Requests: You can send HTTP requests to retrieve contents using this Python module. It enables you to send Get or Post requests to access the HTML of websites or APIs.
* Selenium: It is frequently used for website testing and enables you to automate various website events (such as clicking, scrolling, etc.) to obtain the desired outcomes.
* Selenium or Requests + Beautiful Soup are both options for web scraping. Selenium is recommended if you need to interact with the website (JavaScript events), but Requests + Beautiful Soup is faster and simpler if you don't.