

# Jobs Database Web Scraper & Twitter Bot Notes

- The web scraper scrapes job ads 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 localities (such as Boston).
- Searches for unvisited job urls, gathers employment data from a particular job listing, and enters that data into a SQL database.

## What is a Twitter Bot?

- A Twitter bot is a sort of social bot software that utilizes the Twitter API to manage a Twitter account and may independently carry out tasks including tweeting, retweeting, like, following, unfollowing, and direct messaging other accounts.
- A set of automation rules that define appropriate and impermissible uses of automation are applied to Twitter accounts.
- Appropriate uses of automation include broadcasting useful information, automatically producing interesting or creative content, and automatically replying to users via direct message.
- Inappropriate usage includes spamming, abusing user privacy, and getting around API rate constraints. A larger botnet may contain Twitter bots. They may be employed in disinformation operations and to sway elections.

## Web Scraping:

- The process of extracting data from a webpage is known as web scraping.
- This data is gathered and then exported in a way that the user will find more valuable. For example, a spreadsheet or an API.

## Twitter Scraping using a Bot:

- Using the automatic web scraping application Octoparse, you may extract data from Twitter without writing any code.
- It is a web scraper that mimics how people might interact with websites. You may use it to copy all of the data you see on any website, including Twitter.
- With its simple point-and-click interface, you can quickly create a unique crawler to retrieve Tweets from a given account, Tweets with a specific hashtag, Tweets sent during a certain period of time, etc.
- The extracted data can then be exported into Excel sheets, CSV, HTML, SQL, or streamed live into your database using Octoparse APIs.

