



# Web Scrapping With Twitter API



Subject: **Data Mining**



Lecturers: **PHAUK Sökkhey & CHAN Sophal**



Presenter: **CHOENG Veyseng**

## What is Web Scrapping?

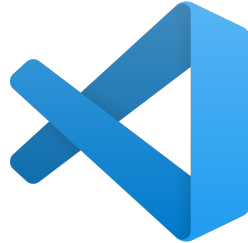
- Web scrapping is the **automatic way** of getting data from websites or online platforms and most of the data is **unstructured** which is required data cleaning.
- There are many ways to do Web Scrapping from very basic methods to advanced methods based on the purpose and how big and complex data we are expecting to do the scrapping
- Python programming language is a power language to support this task and have many supported libraries such as **Scrapy** and **Beautiful Soup**.
- There are many large online platforms such as Facebook, Twitter, StackOverflow and more have their own APIs for developers or data miners or other stakeholders to easily access the data within their platforms

## Introduction to the Project

- The purpose of this project is to get the tweets by hastage **#Cryptocurrencies** and **#Blockchain** from twitter platform
- The potential of this project is to help supporting data analysis to understand the trends within these two topics and general perspectives of the tweets toward these topics

# Project Preparation

## Microsoft Visual Studio Code



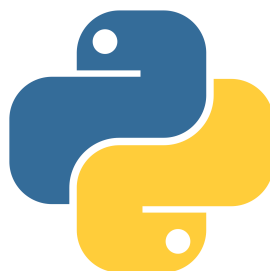
- IDE to access and interact with coding files and other related files

## Anaconda



- The **distribution of Python and R language** for data analysis and more to easily install related libraries and interact with different virtual environments
- It has many commands to access and use Python and R ecosystem

## Python Programming Language



- One of the most popular programming language for multi-purposes
- In the project, we use python code for access the Twitter platform with access tokens having Tweepy library to reduce the code implementation

## Twitter API



- Twitter has enabled the feasibility for businesses, researchers or students with their API to access the data within the platform
- By having the access token or API key, we need to register for Twitter Developer Platform to be eligible for working with Twitter API

## Code and Command to run the code

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b61fb0db-f714-4fdc-b428-04135d5a6a99/main.py>

```
conda run -n web_scrapping python main.py
```

## References

- <https://www.anaconda.com>
- <https://www.python.org>
- <https://developer.twitter.com>
- <https://code.visualstudio.com>

- <https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>
- <https://youtu.be/Lu1nskBkPJU>