IS324 - Spring 2022
Lab – 7
JSON, Multithreading, & Web scraping

LAB OBJECTIVES:
- Parse JSON objects
- Make use of multithreading approach in python programs
- Understand the basic concepts of request & BeautifulSoup modules
- Collecting useful information from the web

Exercise 1:

Go to LMS, and download three files (TweetsDataPart1.txt, TweetsDataPart2.txt, TweetsDataPart3.txt) each of which has 1000 Tweet JSON objects. Each line is one Tweet JSON object. You can check out the Twitter documentation to understand the structure of Tweet JSON here (if needed)

Read each file and parse the Tweet JSON object to extract the following: Date of Tweet (created_at), The User who posted this tweet (user>screen_name), and The Text (text). Then, Save the output to a file (this output file should contain 3000 Tweets).

Report the total time in seconds needed to generate this output file from starting reading the input files until finishing writing the output file.

Exercise 2:

Redo **Exercise 1** by utilizing a multithreading approach. Report the total time in seconds needed to generate this output file from starting reading the input files until finishing writing the output file. and compare it with the total time from **Exercise 1.**

Report how much this approach is faster than the one in **Exercise 1** ?

Optional Exercise 3:

Go to the following website: **https://simplywall.st/stocks/sa/telecom/market-cap-large?page=1**

Write a python program to automate the process of collecting the stock prices of the Telecommunication Services sector.

Whenever the program is executed, it should generate a file named (TSStocks+(current date&time.txt) and the file content should be Short Name and Price.

For example TS-Stocks(09/04/2022-22:33:16).txt:

Saudi Telecom 116.00ر.س

Etihad Etisalat 40.30ر.س

Mobile Telecommunications Company Saudi Arabia 13.32ر.س

Etihad Atheeb Telecommunication 54.10ر.س

You can use the following code to name the generated file each time the program is executed.

```python
import datetime
now = datetime.datetime.now()
now = now.strftime("%d/%m/%Y-%H:%M:%S")
filename = "TS-Stocks("+str(now)+").txt"
```