НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Кафедра «ОБЧИСЛЮВАЛЬНА ТЕХНІКА ТА ПРОГРАМУВАННЯ»

Звiт з лабораторної роботи 13

Тема: «Збір даних з веб-документів за допомогою Python. Робота з API.

Бібліотека Pandas»

Виконав:   
ст. гр. КІТ-118в  
Міщенко Д.С.

Перевірив:  
Поворознюк О.А.

Харків – 2021

1 ВИМОГИ

**1.1 Розробник**

- Міщенко Дмитро Сергійович;

- Студент групи КІТ-118в;

**1.2 Загальне завдання**

Завдання:

А) Використовуючи API: <https://newsapi.org/> Отримати усі новини, що містять задане слово пошуку за заданими датами. Додати можливість вказувати джерела пошуку. Зберегти назву новини, опис, дату публікації, посилання, автора. Відсортувати за датою публікації – від найсвіжіших новин.

Б) Використовуючи Beautiful Soup: <https://finance.yahoo.com/> Для заданих компаній та дат (наприклад AAPL, MSFT, NFLX, AMZN) отримати «Book Value Per Share» «Gross Profit» «Return on Assets», компанії. Приклад посилання: <https://finance.yahoo.com/quote/AAPL/key-statistics?p=AAPL>

2 ОПИС ПРОГРАМИ

**2.1 Фрагменти коду**

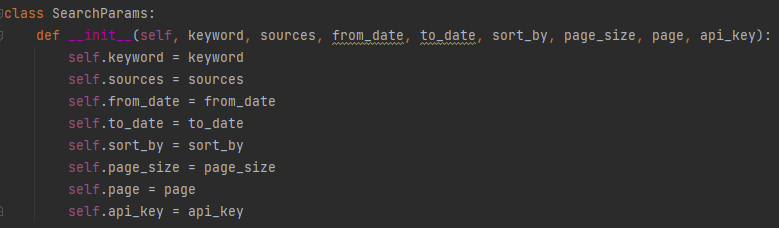


Рисунок 1 – Створений клас з параметрами пошуку



Рисунок 2 – Клас з необхідними даними новин

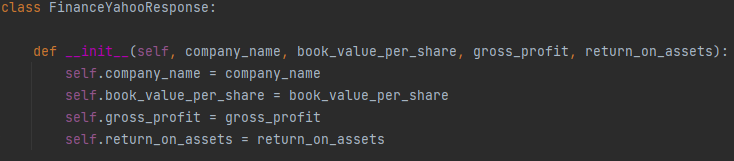
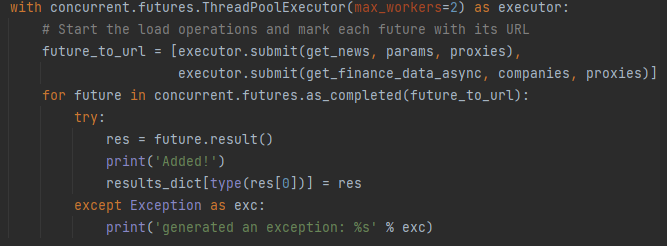


Рисунок 3 - Клас з необхідними даними про компанії



Рисунок 4 - Метод, що створює та посилає запит

  
Рисунок 5 - Використання багатопоточності для паралельного виконання завдань

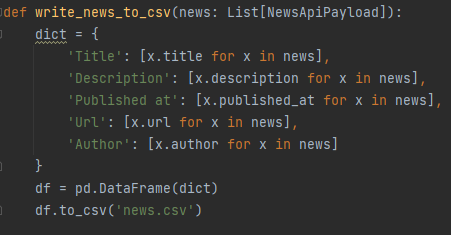
\

Рисунок 6 - Запис у форматі .csv за допомогою бібліотеки Pandas

3 ВАРІАНТИ ВИКОРИСТАННЯ

**3.1 Ілюстрація роботи програми**



Рисунок 7 – Виконуваний код

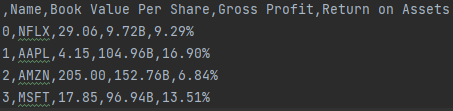


Рисунок 8 – Результат виконання

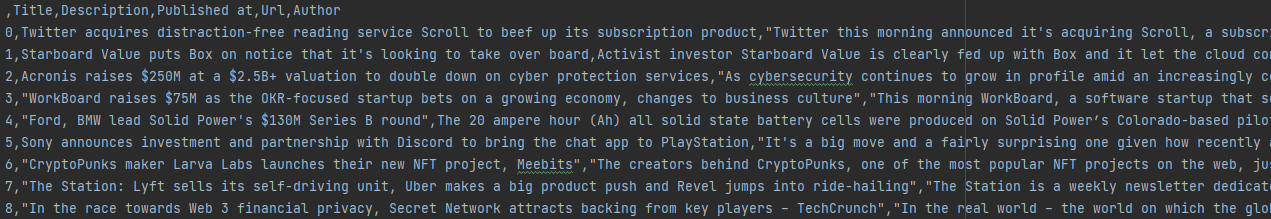


Рисунок 9 – Зміст файлу default.txt

ВИСНОВКИ

У ході виконання лабораторної роботи були набуті практичні навички з роботи з запитами, API деяких сайтів та засобами скраппінгу даних. Був розреблений простіший механізм уникнення блокування зі сторони сервера, а також використана багатопоточність для виконання двох паралельних завдань.