**Міністерство освіти і науки України**

**Національний технічний університет України «КПІ» імені Ігоря Сікорського Кафедра обчислювальної техніки ФІОТ**

**ЗВІТ**

**з лабораторної роботи №4**

**з навчальної дисципліни «Технології Data Science»**

**Тема:**

**РЕАЛІЗАЦІЯ ПРОЦЕСІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ: МІНІ**

**ПРОЕКТИ В ГАЛУЗЯХ OLAP, Data Mining, Text Mining, Voice Recognition**

**Виконав:**

Студент 4 курсу кафедри ФІОТ,

Навчальної групи ІП-11

Олександр Головня

**Перевірив:**

Професор кафедри ОТ ФІОТ Олексій Писарчук

**Київ 2024**

**І. Мета:**

Виявити, дослідити та узагальнити особливості інтелектуального аналізу даних та

технологій OLAP, Data Mining, Text Mining, Voice Recognition.

**ІІ. Завдання:**

1. Лабораторія провідної ІТ-компанії реалізує проект із розробки ERP системи для

автоматизації процесів управління у сфері завдань електронної комерції. Вам, як Data

Science Engineer поставлене наступне завдання. Розробити програмний скрипт мовою

Python що реалізує обчислювальний алгоритм ERP-системи із впровадженням технології

інтелектуального аналізу даних відповідно до технічних умов:

**ІІ рівень складності 9 балів, викладених у табл.2** – R&D міні проект.

З табл.2:



**ІІІ. Результати виконання лабораторної роботи.**

* 1. Блок схема алгоритму.

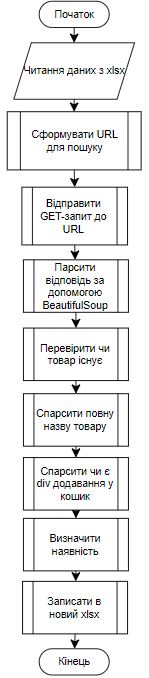


Рис.1 – Блок-схема алгоритму програми

* 1. **Опис структури проекту програми.**

Для реалізації розробленого алгоритму мовою програмування Python з використанням можливостей інтегрованого середовища сформовано проект.

Проект базується на лінійній бізнес-логіці функціонального програмування та має таку структуру.

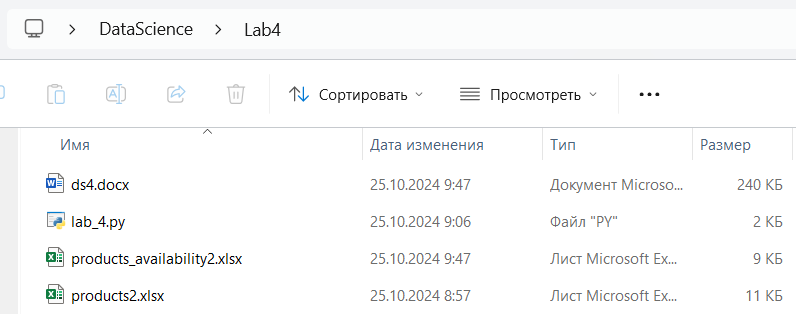


Рис.2 – Структура проєкту

ds4.py – файл програмного коду лабораторної роботи;

ds4.docx – файл звіту лабораторної роботи

products2.xlsx – початковий файл з моделями товару

products\_availability2.xlsx – кінцевий файл з наявністю товару

* 1. **Результати роботи програми відповідно до завдання.**

Спочатку я обрав Інтернет-магазин Bolf – це магазин брендового одягу, також сформував файл .xlsx з моделями десяти різних товарів, для кожного з яких буде виявлено статус наявності на цьому сайті.

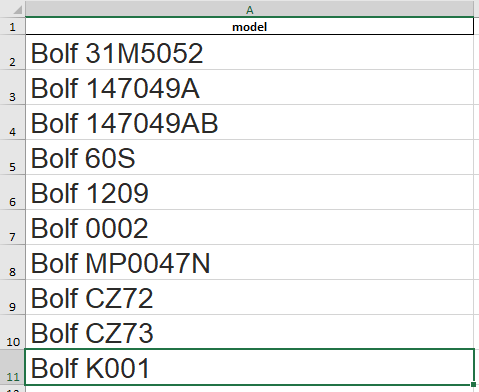
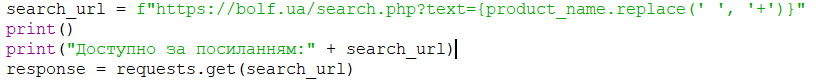


Рис 3.1 – Excel таблиця з моделями десяти різних товарів

Далі йде Парсинг даних про наявність обраних товарів в Інтернет магазині:



Посилання на сайт магазину Blof є наступним: <https://bolf.ua/>

URL-адреса для пошуку конкретного товару буде наступною: https://bolf.ua/search.php?text={product\_name.replace(' ', '+')},

де в якості product\_name.replace(' ', '+') наш параметр в url, який потрібно передати.

Далі за допомогою soup.find шукаємо наш товар, якщо товару не існує, то повернемо повідомлення, що товару цієї моделі не існує на сайті:

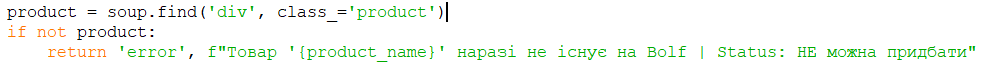


Рис. 3.2 – Пошук товару на сайті магазину Bolf

Якщо даний товар знайдено, то в HTML-коді сторінки, розпарсений бібліотекою BeautifulSoup, буде наявний div класу “product ”

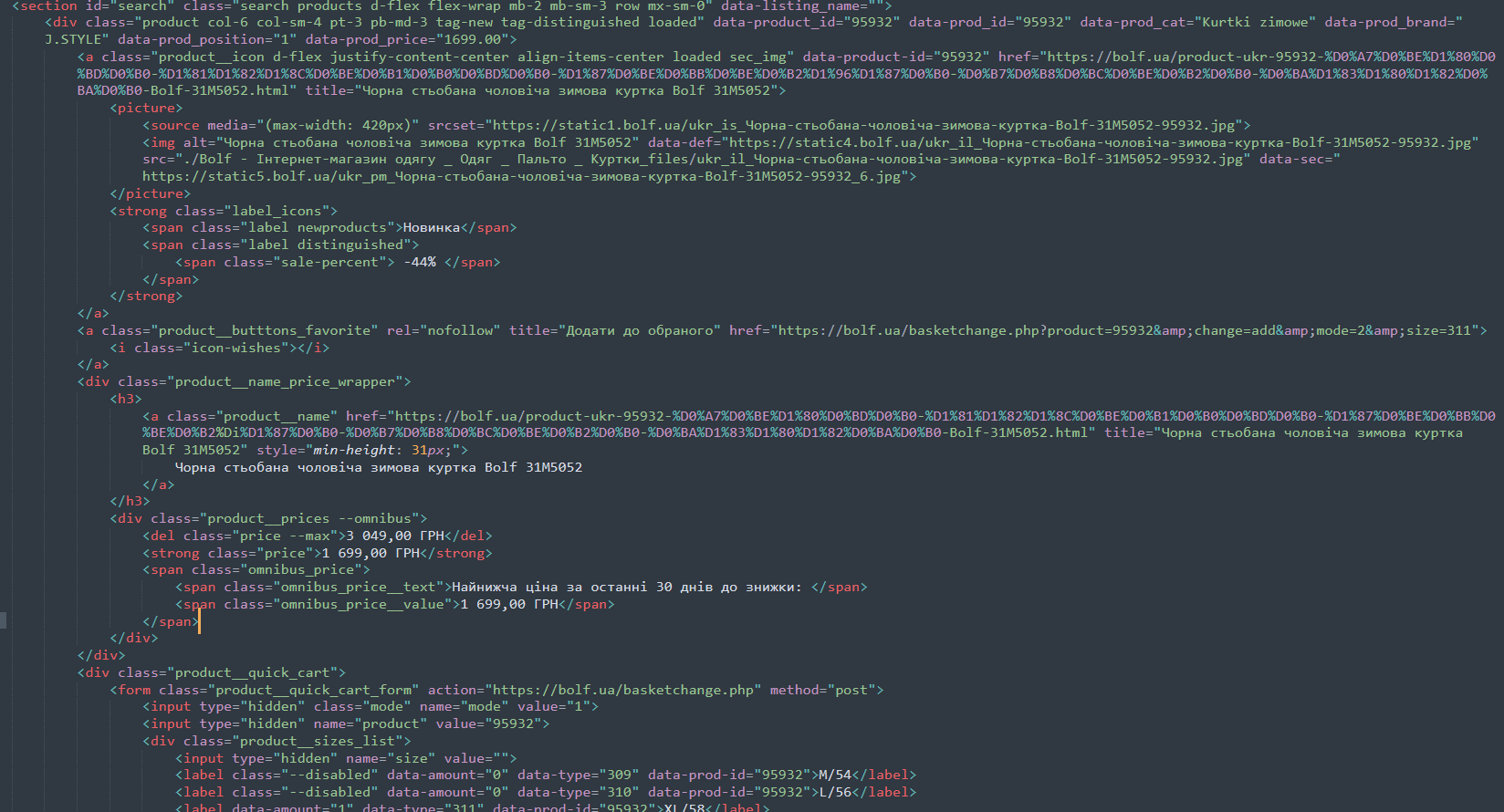


Рис. 3.3 – Фрагмент HTML-коду з потрібним div

Можна побачити, що у цьому div клас є “a”, класу class\_='product\_\_name', який містить текстову інформацію про детальну назву товару. Таким чином, якщо існує div класу product\_\_quick\_cart\_form, який потрібен для додавання у кошик, відповідно до цього текстового рядка я визначаю наявність даного товару та повертаю відповідний статус та результат

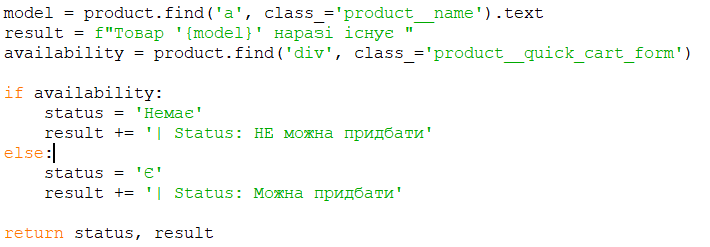


Рис. 3.4 – Парсинг наявності товару

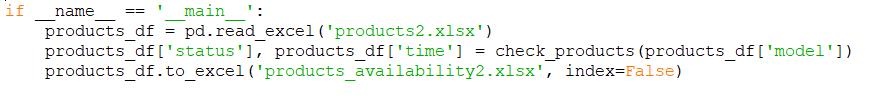


Рис. 3.5 – Перевірка наявності товарів та завантаження отриманих даних у xlsx файл

Для перевірки наявності всіх товарів з файлу я за допомогою бібліотеки Pandas завантажив дані з xlsx файлу в датафрейм та створив два стовпці зі статусом наявності та часом, коли була здійснена перевірка наявності. Отриманий датафрейм я завантажив у новий xlsx файл.

Для отримання даних про наявність усіх товарів я викликаю відповідну функцію, яка повертає списки зі статусами та часом перевірки й видає повідомлення для кожного товару про його наявність.

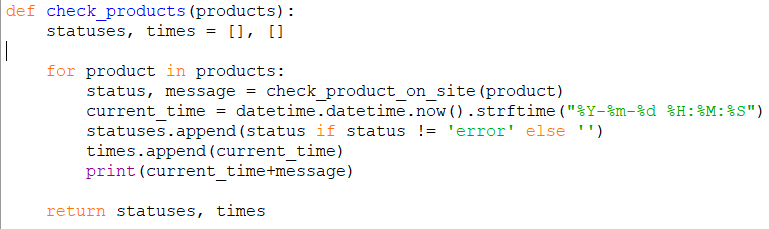


Рис. 3.6 – Перевірка наявності кожного товару



Рис. 3.7 – Повідомлення про наявність кожного товару

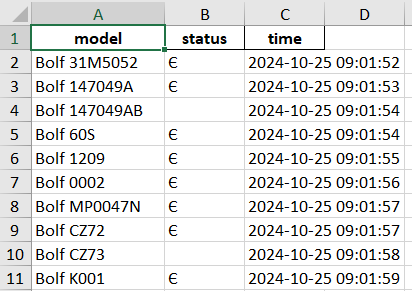


Рис. 3.8 – Вихідний файл з інформацією про наявність кожного товару

* 1. **Програмний код.**

Програмний код послідовно реалізує алгоритм рис.1 та спрямовано на отримання результатів, поданих вище.

При цьому використано можливості Python бібліотек: pip; pandas; numpy; sklearn; matplotlib.

Контексні коментарі пояснюють сутність окремих скриптів наведеного коду програми.

* 1. **Аналіз результатів відлагодження та верифікації результатів роботи програми.**

Результати відлагодження та тестування довели працездатність розробленого коду. Це підтверджується результатами розрахунків, які не суперечать теоретичним положенням.

Верифікація функціоналу програмного коду, порівняння отриманих результатів з технічними умовами завдання на лабораторну роботу доводять, що усі завдання виконані у повному обсязі.

**IV. Висновки.**

Отже, в ході даної лабораторної я застосував на практиці навички парсингу даних з сайту та файлу для виявлення наявності на поточний час в Інтернет магазині Bolf певних обраних товарів, що зберігаються у відповідному xlsx файлі, та завантаження отриманих даних у новий xlsx файл.

Дану програму можна модифікувати шляхом отримання інформації про різноманітність одного товару(Колір, Розмір і тд), наявність товару певну кількість разів за заданий часовий проміжок, щоб моніторити його нявність та надсилати повідомлення в разі його появи. Варто зазначити, що реалізація лабораторної роботи має доволі спрощений функціонал.