

**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 «Вступ до технології Data Science»**

Тема:

**РЕАЛІЗАЦІЯ ПРОЦЕСІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ: МІНІ
ПРОЕКТИ В ГАЛУЗЯХ OLAP, Data Mining, Text Mining, Voice Recognition**

Виконав:

Студент 3 курсу кафедри ІПІ ФІОТ,
Навчальної групи ІП-11
Сідак К.І.

Перевірив:

Професор кафедри ОТ ФІОТ
Писарчук О.О.

Київ 2023

I. Мета роботи:

Виявити дослідити та узагальнити особливості реалізації процесів статистичного навчання із застосуванням методів обробки Big Data масивів та калмановської рекурентної фільтрації з використанням можливостей мови програмування Python.

II. Завдання:

Лабораторія провідної IT-компанії реалізує масштабний проект розробки універсальної платформи з обробки Big Data масиву статистичних даних поточного спостереження для виявлення закономірностей і прогнозування розвитку контрольованого процесу. Платформа передбачає розташування back-end компоненти на власному хмарному сервері з наданням повноважень користувачам заздалегідь адаптованого front-end функціоналу універсальної платформи.

II рівень складності 8 балів, викладених у табл.2.

2. Розробити програмний скрипт, що реалізує виявлення факту появи обраного товару в Інтернет магазині.

III. Виконання лабораторної роботи.

3.1. Створення файлу з обраними товарами

Спочатку я сформував файл .xlsx з моделями десяти різних товарів, для кожного з яких буде виявлено статус наявності.

1	model
2	Apple iPhone 15 Pro 128GB Black Titanium Dual SIM (MTQ43)
3	Dell S2722QC
4	Apple AirPods Pro 2 with Magsafe White (MQD83)
5	Xiaomi Dreame Bot W10 Pro
6	Apple MacBook Air 15" M2 256GB Starlight (MQKU3) 2023
7	Razer DeathAdder CHROMA (RZ01-01210100) Grade B2 Refurbished
8	Bluetti PowerOak EB70 716Wh 1000W 198889mAh Grey (EB70)
9	Apple iPad Pro 4 11" 2022 Wi-Fi 128GB M2 Space Gray (MNXD3)
10	Zotac GeForce GTX Titan X ArcticStorm (ZT-90402-10P)
11	MSI GeForce RTX 4090 SUPRIM LIQUID X 24G

Рис 3.1 – Excel таблиця з моделями десяти різних товарів

3.2. Парсинг даних про наявність обраних товарів в Інтернет магазині Stylus

Посилання на сайт магазину Stylus є наступним: <https://stylus.ua>. URL-адреса для пошуку конкретного товару є наступною: <https://stylus.ua/search?q=>, де в якості параметра потрібно передати назву товару. Якщо статус код не 200 в результаті GET запиту, то буде повернуто статус помилки та відповідне повідомлення. Крім того, якщо задана назва товару є неоднозначною, тобто результатом пошуку буде 2 або більше товари, то буде видано відповідно повідомлення та повернуто статус помилки. Аналогічний статус та відповідне повідомлення буде повернуто, якщо даного товару наразі взагалі не присутній в асортименті магазину.

```
def check_product_on_stylus(product_name):
    try:
        search_url = f"https://stylus.ua/search?q={product_name.replace(' ', '+')}}"
        response = requests.get(search_url)

        if response.status_code != 200:
            return 'error', "Failed to fetch the search results page from Stylus."

        soup = BeautifulSoup(response.text, features='html.parser')
        product_listing = soup.find(name='div', class_='product-listing-right')

        if product_listing:
            return 'error', f"The product name '{product_name}' is ambiguous."

        product = soup.find(name='div', class_='product-page-block')
        if not product:
            return 'error', f"The product '{product_name}' is not present on Stylus."
```

Рис. 3.2 – Пошук товару на сайті магазину Stylus

Якщо даний товар знайдено, то в HTML-кодї сторінки, який отриманий бібліотекою requests та розпарсений бібліотекою BeautifulSoup, буде наявний div класу “product-page-block”.

```
<div class="product-page-block">
  <h1 class="page-name">
    Монитор Dell S2722QC (210-BBRQ)
  </h1>
  <div class="" data-ec="{\"name\":\"Dell S2722QC (210-BBRQ)\",\"category\":\"Монітори\",\"brand\":\"Dell\",\"id\":\"985497\",\"price\":\"15278\"}" data-stock="out-of-stock" data-voted="false" id="product-layout">
    <div class="top-info">
      <div class="sku-block">
        Код товару:NBSP985497
      </div>
      <div class="reviews-block mobile">
        <div class="rating-block" data-rating="0">
          <div class="rating-block__empty">
            </div>
          <div class="rating-block__fill" style="width: 0%;>
            </div>
        </div>
      </div>
      <div class="availability-block mobile">
        <div class="availability not-available">
          Нет в наличии
        </div>
      </div>
    </div>
  </div>
```

Рис. 3.3 – Фрагмент HTML-коду з потрібним div

Можна побачити, що у div класу “availability-block-mobile” міститься один внутрішній div, який містить текстову інформацію про наявність товару. Таким чином, відповідно до цього текстового рядка я визначаю наявність даного товару та повертаю відповідний статус та повідомлення.

```
model = product.find('h1', class_='page-name').text
availability = product.find('div', class_='availability-block mobile').find('div').text
result = f"The product '{model}' is currently "

if availability == 'Нет в наличии':
    status = 'out of stock'
    result += 'out of stock on Stylus.'
elif availability == 'Доступен предзаказ':
    status = 'available on preorder'
    result += 'available on preorder on Stylus.'
else:
    status = 'available'
    result += 'available on Stylus.'

return status, result
```

Рис. 3.4 – Парсинг наявності товару

3.3. Перевірка наявності для усіх товарів з файлу

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.

```
if __name__ == '__main__':
    products_df = pd.read_excel('products.xlsx')
    products_df['status'], products_df['time'] = check_products(products_df['model'])
    products_df.to_excel(excel_writer='products_availability.xlsx', index=False)
```

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

```
def check_products(products):
    statuses, times = [], []

    for product in products:
        status, message = check_product_on_stylus(product)
        current_time = datetime.datetime.now().strftime("%Y-%m-%d %H:%M:%S")
        statuses.append(status if status != 'error' else '')
        times.append(current_time)
        print(f'{current_time} {message}')

    return statuses, times
```

Рис. 3.6 – Перевірка наявності кожного товару

```
2023-11-05 19:34:45 The product 'Apple iPhone 15 Pro 128GB Black Titanium Dual SIM (MTQ43)' is currently available on preorder on Stylus.
2023-11-05 19:34:46 The product 'Монитор Dell S2722QC (210-BBRQ)' is currently out of stock on Stylus.
2023-11-05 19:34:47 The product 'Наушники Apple AirPods Pro 2 with Magsafe White (MQD83)' is currently available on Stylus.
2023-11-05 19:34:49 The product 'Робот уборщик Xiaomi Dreame Bot W10 Pro' is currently available on Stylus.
2023-11-05 19:34:50 The product 'Apple MacBook Air 15" M2 256GB Starlight (MQKU3) 2023' is currently available on Stylus.
2023-11-05 19:34:51 The product 'Razer DeathAdder CHROMA (RZ01-01210100) Grade B2 Refurbished' is currently out of stock on Stylus.
2023-11-05 19:34:52 The product 'Зарядная станция Bluetti PowerOak EB70 716Wh 1000W 198889mAh Grey (EB70) Международная версия' is currently available on Stylus.
2023-11-05 19:34:53 The product 'Планшет Apple iPad Pro 4 11" 2022 Wi-Fi 128GB M2 Space Gray (MNXD3)' is currently available on Stylus.
2023-11-05 19:34:54 The product 'Видеокарта Zotac GeForce GTX Titan X ArcticStorm (ZT-90402-10P)' is currently out of stock on Stylus.
2023-11-05 19:34:55 The product 'Видеокарта MSI GeForce RTX 4090 SUPRIM LIQUID X 24G' is currently available on Stylus.

Process finished with exit code 0
```

Рис. 3.7 – Повідомлення про наявність кожного товару

1	model	status	time
2	Apple iPhone 15 Pro 128GB Black Titanium Dual SIM (MTQ43)	available on preorder	2023-11-05 19:34:45
3	Dell S2722QC	out of stock	2023-11-05 19:34:46
4	Apple AirPods Pro 2 with Magsafe White (MQD83)	available	2023-11-05 19:34:47
5	Xiaomi Dreame Bot W10 Pro	available	2023-11-05 19:34:49
6	Apple MacBook Air 15" M2 256GB Starlight (MQKU3) 2023	available	2023-11-05 19:34:50
7	Razer DeathAdder CHROMA (RZ01-01210100) Grade B2 Refurbished	out of stock	2023-11-05 19:34:51
8	Bluetti PowerOak EB70 716Wh 1000W 198889mAh Grey (EB70)	available	2023-11-05 19:34:52
9	Apple iPad Pro 4 11" 2022 Wi-Fi 128GB M2 Space Gray (MNXD3)	available	2023-11-05 19:34:53
10	Zotac GeForce GTX Titan X ArcticStorm (ZT-90402-10P)	out of stock	2023-11-05 19:34:54
11	MSI GeForce RTX 4090 SUPRIM LIQUID X 24G	available	2023-11-05 19:34:55

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

IV. Висновок

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 Stylus 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 наявність товар певну кількість разів за заданий часовий проміжок, щоб моніторити його наявність та надсилати повідомлення в разі його появи або ж у вигляді десктопного повідомлення, або бота у конкретному месенджері. Варто зазначити, що для реалізації подібного функціоналу бажано, щоб програма була задеплована на сервері, тому в рамках лабораторної роботи я реалізував варіант функціоналу, описаний вище.