

**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

I. Мета:

Виявити, дослідити та узагальнити особливості інтелектуального аналізу даних та технологій OLAP, Data Mining, Text Mining, Voice Recognition.

II. 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:

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

З табл.2:

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

III. Результати виконання лабораторної роботи.

3.4.Блок схема алгоритму.

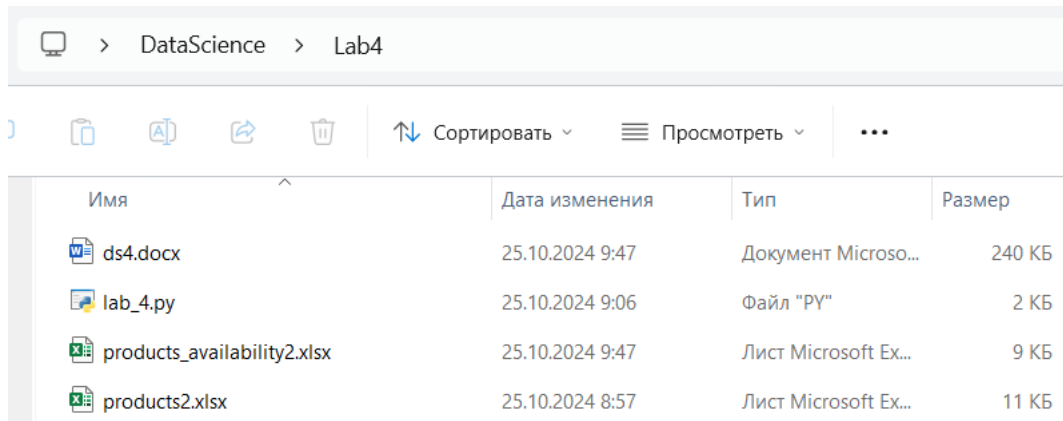


Рис.1 – Блок-схема алгоритму програми

3.1. 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.



Имя	Дата изменения	Тип	Размер
ds4.docx	25.10.2024 9:47	Документ Microso...	240 КБ
lab_4.py	25.10.2024 9:06	Файл ".py"	2 КБ
products_availability2.xlsx	25.10.2024 9:47	Лист Microsoft Ex...	9 КБ
products2.xlsx	25.10.2024 8:57	Лист Microsoft Ex...	11 КБ

Рис.2 – Структура проекту

ds4.py – файл програмного коду лабораторної роботи;

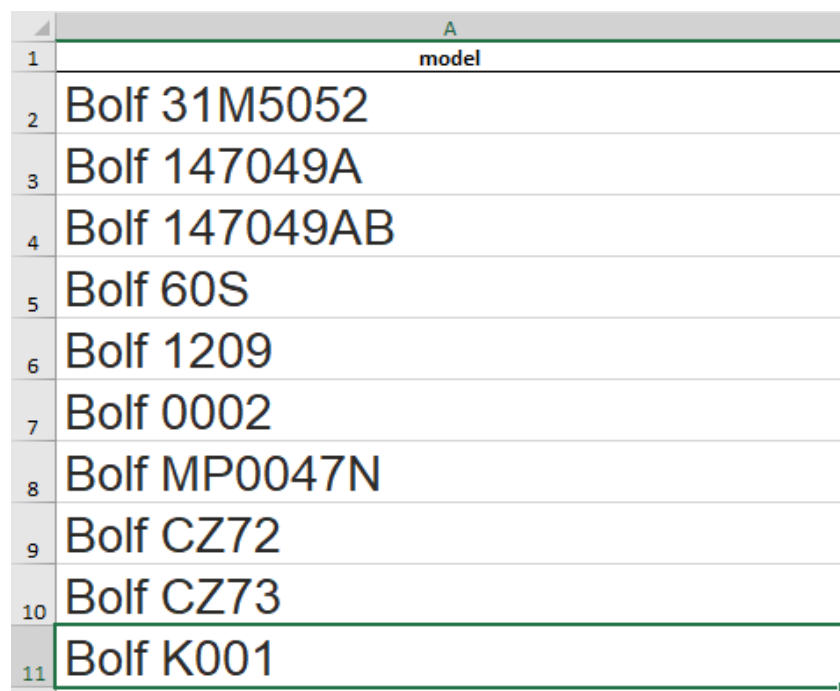
ds4.docx – файл звіту лабораторної роботи

products2.xlsx – початковий файл з моделями товару

products_availability2.xlsx – кінцевий файл з наявністю товару

3.2. 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.



	A
1	model
2	Bolf 31M5052
3	Bolf 147049A
4	Bolf 147049AB
5	Bolf 60S
6	Bolf 1209
7	Bolf 0002
8	Bolf MP0047N
9	Bolf CZ72
10	Bolf CZ73
11	Bolf K001

Рис 3.1 – Excel таблиця з моделями десяти різних товарів

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

```
search_url = f"https://bolf.ua/search.php?text={product_name.replace(' ', '+')}"
print()
print("Доступно за посиланням:" + search_url)|
response = requests.get(search_url)
```

Посилання на сайт магазину Bolf 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:

```
product = soup.find('div', class_='product')|
if not product:
    return 'error', f"Товар '{product_name}' наразі не існує на Bolf | Status: НЕ можна придбати"
```

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”

```
<section id="search" class="search products d-flex flex-wrap mb-2 mb-sm-3 row mx-sm-0" data-listing_name="">
<div class="product col-6 col-sm-4 pt-3 pb-md-3 tag-new tag-distinguished loaded" data-product_id="95932" data-prod_id="95932" data-prod_cat="Kurtki zimowe" data-prod_brand="
J.STYLE" data-prod_position="1" data-prod_price="1699.00">
<a class="product_icon d-flex justify-content-center align-items-center loaded sec_img" data-product-id="95932" href="https://bolf.ua/product-ukr-95932-%D0%A7%D0%BE%D1%80%D0%
%BD%D0%B0-%D1%81%D1%82%D1%8C%D0%BE%D0%B1%D0%B0%D0%B0-%D1%87%D0%BE%D0%B8%D0%BE%D0%B2%D1%96%D1%87%D0%B0-%D0%B7%D0%B8%D0%BC%D0%BE%D0%B2%D0%B0-%D0%BA%D1%83%D1%80%D1%82%D0%
BA%D0%B0-Bolf-31M5052.html" title="Чорна стьобана чоловіча зимова куртка Bolf 31M5052">
<picture>
<source media="(max-width: 420px)" srcset="https://static1.bolf.ua/ukr_is_Чорна-стьобана-чоловіча-зимова-куртка-Bolf-31M5052-95932.jpg">

</picture>
<strong class="label_icons">
<span class="label newproducts">Новинка</span>
<span class="label distinguished">
<span class="sale-percent">-44%</span>
</span>
</strong>
</a>
<a class="product_buttons_favorite" rel="nofollow" title="Додати до обраного" href="https://bolf.ua/basketchange.php?product=95932&change=add&mode=2&size=311">
<i class="icon-wishes"></i>
</a>
<div class="product_name_price_wrapper">
<h3>
<a class="product_name" href="https://bolf.ua/product-ukr-95932-%D0%A7%D0%BE%D1%80%D0%B0%D0%B0-%D1%81%D1%82%D1%8C%D0%BE%D0%B1%D0%B0%D0%BD%D0%B0-%D1%87%D0%BE%D0%B8%D0%
%BD%D0%B2%D1%87%D0%B0-%D0%B7%D0%B8%D0%BC%D0%BE%D0%B2%D0%B0-%D0%BA%D1%83%D1%80%D1%82%D0%BA%D0%B0-Bolf-31M5052.html" title="Чорна стьобана чоловіча зимова куртка
Bolf 31M5052" style="min-height: 31px;">
Чорна стьобана чоловіча зимова куртка Bolf 31M5052
</a>
</h3>
<div class="product_prices --omnibus">
<del class="price --max">3 049,00 ГРН</del>
<strong class="price">1 699,00 ГРН</strong>
<span class="omnibus_price">
<span class="omnibus_price_text">Найнижча ціна за останні 30 днів до знижки:</span>
<span class="omnibus_price_value">1 699,00 ГРН</span>
</span>
</div>
</div>
<div class="product_quick_cart">
<form class="product_quick_cart_form" action="https://bolf.ua/basketchange.php" method="post">
<input type="hidden" class="mode" name="mode" value="1">
<input type="hidden" name="product" value="95932">
<div class="product_sizes_list">
<input type="hidden" name="size" value="">
<label class="--disabled" data-amount="0" data-type="309" data-prod-id="95932">M/54</label>
<label class="--disabled" data-amount="0" data-type="310" data-prod-id="95932">L/56</label>
<label data-amount="1" data-type="311" data-prod-id="95932">XL/58</label>
```

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

```

model = product.find('a', class_='product__name').text
result = f"Товар '{model}' наразі існує "
availability = product.find('div', class_='product__quick_cart_form')

if availability:
    status = 'Немає'
    result += '| Status: НЕ можна придбати'
else:
    status = 'Є'
    result += '| Status: Можна придбати'

return status, result

```

Рис. 3.4 – Парсинг наявності товару

```

if __name__ == '__main__':
    products_df = pd.read_excel('products2.xlsx')
    products_df['status'], products_df['time'] = check_products(products_df['model'])
    products_df.to_excel('products_availability2.xlsx', index=False)

```

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.

```

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

    for product in products:
        status, message = check_product_on_site(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(current_time+message)

    return statuses, times

```

Рис. 3.6 – Перевірка наявності кожного товару

```
= RESTART: C:\Users\Саша Головня\Desktop\DataScience\Lab4\lab_4.py

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+31M5052
2024-10-25 09:01:52Товар 'Чорна стьобана чоловіча зимова куртка Bolf 31M5052' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+147049A
2024-10-25 09:01:53Товар 'Чорний чоловіча толстовка з капюшоном і принтом Bolf 147049A' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+147049AB
2024-10-25 09:01:54Товар 'Bolf 147049AB' наразі не існує на Bolf | Status: НЕ можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+60S
2024-10-25 09:01:54Товар 'Толстовка чоловіча з капюшоном і принтом cipa Bolf 60S' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+1209
2024-10-25 09:01:55Товар 'Чоловічий лонгслів без принта антрацитовий Bolf 1209' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+0002
2024-10-25 09:01:56Товар 'Чорний велюровий чоловічий спортивний костюм Bolf 0002' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+MP0047N
2024-10-25 09:01:57Товар 'Чорні чоловічі джински regular fit Bolf MP0047N' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+CZ72
2024-10-25 09:01:57Товар 'Бордова кепка, бейсболка Bolf CZ72' наразі існує | Status: Можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+CZ73
2024-10-25 09:01:58Товар 'Bolf CZ73' наразі не існує на Bolf | Status: НЕ можна придбати

Доступно за посиланням:https://bolf.ua/search.php?text=Bolf+K001
2024-10-25 09:01:59Товар 'Елегантна чорна чоловіча краватка Bolf K001' наразі існує | Status: Можна придбати
>>>|
```

Рис. 3.7 – Повідомлення про наявність кожного товару

	A	B	C	D
1	model	status	time	
2	Bolf 31M5052	Є	2024-10-25 09:01:52	
3	Bolf 147049A	Є	2024-10-25 09:01:53	
4	Bolf 147049AB		2024-10-25 09:01:54	
5	Bolf 60S	Є	2024-10-25 09:01:54	
6	Bolf 1209	Є	2024-10-25 09:01:55	
7	Bolf 0002	Є	2024-10-25 09:01:56	
8	Bolf MP0047N	Є	2024-10-25 09:01:57	
9	Bolf CZ72	Є	2024-10-25 09:01:57	
10	Bolf CZ73		2024-10-25 09:01:58	
11	Bolf K001	Є	2024-10-25 09:01:59	

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

3.3. 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.

3.4. 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 Volf 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 наявність та надсилати повідомлення в разі його появи. Варто зазначити, що реалізація лабораторної роботи має доволі спрощений функціонал.