**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

Лошак В.І.

**Перевірив:**

Професор кафедри ОТ ФІОТ

Писарчук О.О.

**Київ 2023**

**І. Мета роботи:**

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.**

Розробити програмний скрипт, що реалізує виявлення факту появи обраного товару в

Інтернет магазині.

**ІІІ. Результати виконання лабораторної роботи.**

Для виконання лабораторної були використані бібліотеки python такі як: selenium— збереження вмісту сторінки html. Також BeautifulSoup для обробки сторінки та пошуку даних на ній.

Для парсингу обрано сайт rozetka, а точніше його сторінку за посиланням: <https://rozetka.com.ua/ua/mobile-phones/c80003/producer=apple/#search_text=iphone> що містить дані про останні моделі айфона. Мета лабораторної — зреагувати на вихід нового iphone і зберегти дані про нього в файл.

Оскільки rozetka використовує механізм блокування парсингу, використано headless мод доступу до браузеру бібліотеки selenium щоб отримати дані.

**A screen shot of a computer code

Description automatically generated**

Рис. 1 — завантаження даних в програму для подальшої обробки.



Рис. 2 – Дані з назвою

A screen shot of a computer program

Description automatically generated

Рис. 3 – Дані з ціною

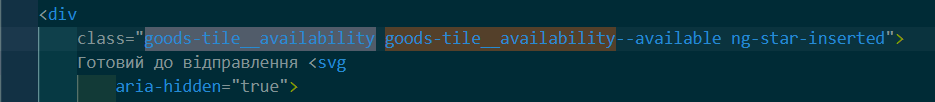


Рис. 4 – Дані про наявність

Для видобутку необхідних даних зі коду сторінки використано наступний скрипт.

A screen shot of a computer code

Description automatically generated

Рис. 5 – Вивантаження необхідних даних про товари внутрішньої структури даних

A screen shot of a computer code

Description automatically generated

Рис. 6 — код використаний для збереження необхідних даних в файл з часовою відміткою для подальшого аналізу.

Після збереження вивантажимо файл в програму щоб пересвідчитись в його цілісності і достовірності.

A screenshot of a computer

Description automatically generated

Рис.7 —Дані про результати пошуку iphone 14.12.2023

**A screenshot of a computer program

Description automatically generated**

Рис.7 —Дані про результати пошуку iphone 14.12.2023(вигляд всередині csv файлу)

**IV. Висновки.**

У ході цієї роботи, я практикував парсинг сайтів для отримання потрібних даних, форматування даних та їх підготовку до статистичного навчання та оцінки. Варто зазначити, що функціонал такої програми підлягає вдосконаленню в подальшому, існує багато опцій. Наприклад, використати email розсилку яка буде сповіщати користувача, що товар з’явився у продажу. Можлива розробка систем e-commerce на основі даного функціоналу.

Виконав: студент ФІОТ Лошак В.І. ІП-11