

Міністерство освіти і науки України  
НТУУ «Київський політехнічний інститут»  
Фізико-технічний інститут

**Системи та засоби інтерактивної аналітики**  
Лабораторна робота No5  
Автоматизація збирання і обробки інформації  
Варіант No6

**Виконав:**  
Студент 4-го курсу  
групи ФІ-21  
Климентьєв Максим  
**Перевірив:**

---

# Зміст

<b>1</b>	<b>Мета роботи</b>	<b>3</b>
<b>2</b>	<b>Завдання</b>	<b>4</b>
<b>3</b>	<b>Код реалізації</b>	<b>6</b>
<b>4</b>	<b>Висновки</b>	<b>9</b>
<b>5</b>	<b>Контрольні питання</b>	<b>10</b>

# **1 Мета роботи**

Навчитися автоматизовано отримувати інформацію з сторінки сайта, попередньо її обробляти і записувати у базу даних для подальшої роботи.

## 2 Завдання

Основна задача - навчитися обробляти інформацію автоматизовано. **Всі дії** зробити за допомогою програми на мові програмування, яку ви знаєте.

### 1. Імпорт інформації.

Файл завантажити у БД, якщо Ваш комп'ютер повільний, можна з файла взяти менший об'єм інформації.

**На 14 або 15 балів** файл імпортувати у базу як є, кожен рядок -> один запис у базі (ID, TEXT), потім SQL запитом з регекспами перетворит таблицю на таблицю з записами де кожне поле логфайла відповідає полю у таблиці (ID, IP, DATA, URL, RETCODE, SIZE, ...). Потім створити SQL запит, яким порахувати вказане у п2. І програмно виміряти час, за який виконуються пп 2-3. Потім зробити аналогічні дії без використання БД, програмно виміряти час, за який виконуються пп 1-2 і порівняти ресурсоемність двох підходів.

**На 0-13 балів** зробити пункт 2 будь-яким способом, без порівняння ресурсоемності.

### 2. Аналіз даних.

Знайти розміри скачаних даних за кожним кодом стану (всі, не тільки 2xx), які були скачані з певної IP адреси, адреси для кожного варіанту наведені у Табл. 5.1.

### 3. Візуалізація.

Вивести діаграму розподілу скачаного для перших 3х кодів (з попереднього пункта) і зберегти її у файл (програмно).

4. Вивести всю можливу інформацію про IP адресу (країна, місто, провайдер і т.д.).

5. Постаратися все зробити у одній програмі. IP - задати змінною для того щоб можна бу

6. Створити звіт. Приєднати до класу.

- У звіті навести все необхідне для повторення і перевірки ваших дій (діаграму БД, SQL запити для створення БД і таблиць, структуру БД, і т.д). Навести знімки екрана, які підтверджують виконані дії.
- У протоколі SQL запити наводити у текстовому вигляді щоб їх можна було редагувати і модифікувати під час захисту.
- Зробити висновки по роботі і занести їх у звіт.

7. Підготувати відповіді на контрольні питання (для офлайн захисту навести їх у протоколі, розкрити сутність, навести приклади).

## 8. Захистити роботу.

Можна використовувати будь яку мову програмування, “картинку” треба згенерувати “статично”, без використання online бібліотек на зразок Chart.js, Google Charts або Chartist, але можна їх використати додатково.

№ варіанта	Завдання
6	46.125.249.79

### 3 Код реалізації

```
(.env) PS C:\GitHub\InteractiveAnalytic> & C:/GitHub/InteractiveAnalytic/.env/Script
      status total_size
0      200    1560515.0
1      404       2324.0
status
200    1560515
404      2324
Name: size, dtype: int64
[Time Spent] SQL: 58.1743 sec | Python 306.1213 sec
```



## Аналіз логів сервера

Recreate data

Напишіть IP для аналізу

46.125.249.79

SQL: 1.0488 сек | Python: 0.3018 сек

### Using Database

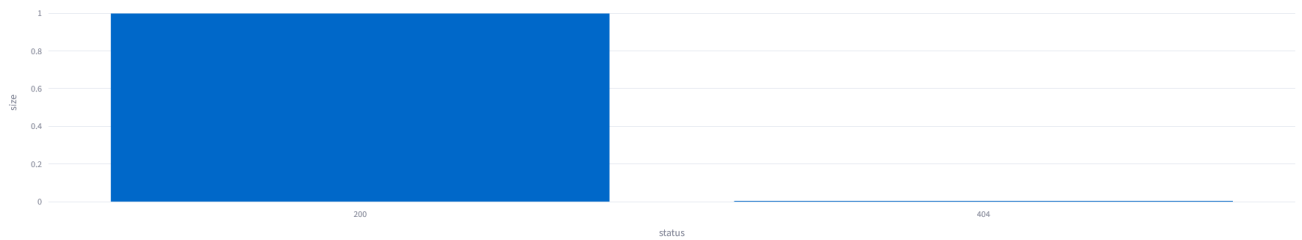
	status	total_size
0	200	1560515
1	404	2324

### Using Python

	status	size
0	200	1560515
1	404	2324

📊 📄 📁 📂 📅 📆 📇 📈 📉 📊

### Розподіл скачаного за кодами



Download chart as PNG





## Інформація про IP:

```
▼ {  
  "status" : "success"  
  "country" : "Austria"  
  "countryCode" : "AT"  
  "region" : "9"  
  "regionName" : "Vienna"  
  "city" : "Vienna"  
  "zip" : "1020"  
  "lat" : 48.217  
  "lon" : 16.3944  
  "timezone" : "Europe/Vienna"  
  "isp" : "T-Mobile Austria"  
  "org" : "T-Mobile Austria GmbH"  
  "as" : "AS8412 T-Mobile Austria GmbH"  
  "query" : "46.125.249.79"  
}
```



## **4 Висновки**

Створено програму для автоматизованого імпорту, обробки та аналізу log-файлів із подальшим збереженням у базі даних. Реалізовано побудову SQL-запитів для вибірки даних за IP-адресою, підрахунку обсягів завантажень за кодами стану та візуалізації результатів у вигляді діаграми.

## 5 Контрольні питання

1. Як можна отримувати дані з бази даних за допомогою SQL-запиту і відображати їх на сторінці за допомогою PHP та GD, або у тих мовах і засобах, які ви використали?

Для реалізації відображення даних з бази даних у Python можна використати бібліотеки `sqlite3`, `duckdb`, `matplotlib`, `plotly`, `pandas` та фреймворк `Streamlit`.