

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

| | | |
|----------|---------------------------|----------|
| 1 | Мета роботи | 3 |
| 2 | Завдання | 4 |
| 3 | Код реалізації | 5 |
| 4 | Висновки | 8 |
| 5 | Контрольні питання | 9 |

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.

| № варіанта | Завдання |
|------------|--|
| 6 | <p><u>Використати існуючу БД (лаб. роб. 2,3)</u></p> <p>Запити:</p> <ol style="list-style-type: none"> 1. Вивести повні адреси JS скриптів (розширення js) і їх розмір, які запитали з комп'ютера з IP адресою 83.227.29.211, для яких запит завершився вдало. 2. Придумайте будь який запит з використанням UNION, в дослідити різні форми цієї команди. 3. Знайдіть розмір всього скачаного комп'ютером з IP адресою 83.227.29.211 |

3 Код реалізації







4 Висновки

Знайдено записи, з нестандартними діями, ідентифіковано IP адреси комп'ютерів з яких відбулась атака і передано ці записи до кіберполіції.

5 Контрольні питання

1. Що таке запити розширеного SQL (advanced SQL)?
2. Що таке регулярні вирази?
3. Для чого використовуються запити з командою UNION?
4. Що таке агрегатні функції?
5. Для чого використовуються запити з командою ORDER?
6. Для чого використовуються запити з командою HAVING?