

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО”**

**Факультет прикладної математики
Кафедра програмного забезпечення комп’ютерних систем**

**КУРСОВИЙ ПРОЕКТ
ТЕХНІЧНЕ ЗАВДАННЯ
з дисципліни “Бази даних”**

спеціальність 121 – Програмна інженерія

на тему: Моніторингова система терористичних актів у світі

Студент

групи КП-71

Крючек Євгеній Іванович

(підпис)

Викладач

к.т.н, доцент кафедри

СПіСКС

Петрашенко А.В.

(підпис)

Київ – 2020

ЗМІСТ

	стор.
1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ	3
2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ	3
3. МЕТА РОЗРОБКИ	3
4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	3
5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД	4
6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА.....	4
7. ВИБІР ЗАСОБІВ РОЗРОБКИ.....	5
8. ЕТАПИ РОЗРОБКИ	5

1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ

Найменування: моніторингова система терористичних актів у світі.

Галузь застосування: сфера національної безпеки, запобігання терористичним актам.

2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ

Дата початку проекту: 26.02.2020

Дата закінчення проекту: 21.06.2020

3. МЕТА РОЗРОБКИ

Метою розробки є створення програмного забезпечення, яке буде взаємодіяти з постріляційною базою даних, проводити заміри досліджуваного об'єкта і подальший аналіз, оцінку, порівняння отриманих результатів для виявлення певних закономірностей, тенденцій, змінних і їх динаміки. А також здобуття навичок використання СУБД та інструментальними засобами підтримки розробки додатків для подібних баз даних, оформлення відповідного текстового, програмного та ілюстративного матеріалу у формі проектної документації.

4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

- Вимоги до генерації даних:

Для генерації даних слід використовувати відкриті набори даних з сайту [Kaggle - розділ datasets](https://www.kaggle.com/START-UMD/gtd). Для цього слід розробити відповідний компонент ПЗ, який буде використовувати дані про терористичні акти, які доступні за посиланням у форматі csv <https://www.kaggle.com/START-UMD/gtd>

- Вимоги до фільтрації та валідації даних:

Щоразу при отриманні даних про терористичний акт їх слід фільтрувати наступним чином: в масиві повинні залишитися записи лише з унікальними ідентифікаторами інциденту (штучний числовий ідентифікатор). При виявленні декількох таких записів слід

проігнорувати той запис, який надійшов пізніше. Також слід відфільтровувати ті записи про інциденти, в яких відсутні обов'язкові поля: рік, країна, регіон, статус акту, чи був акт самогубством, тип цілі акту та тип озброєння.

- Вимоги до засобів реплікації та масштабування:

Розроблений програмний комплекс має використовувати СУБД MongoDB, масштабування виконується горизонтально, за допомогою сегментування (sharding) об'єктів баз даних (розподілення їх частин по різних вузлах кластера); реплікацію досягти шляхом використання replica set.

- Вимоги до аналізу даних:

Отримані дані повинні бути проаналізовані створеним програмним забезпеченням в наступний спосіб: оцінювати кількість терористичних актів по рокам, кількість актів по країнам та регіонам, оцінювати співвідношення вдалих/невдалих актів та актів, які були/не були спробою самогубства та створювати статистику типів жертв та використаної зброї. Для реалізації алгоритмів використовувати математичні бібліотеки або методи Big Data тощо.

- Вимоги до резервування та відновлення даних:

Резервування даних здійснювати шляхом використання реплікації. У разі втрачання зв'язку з певним вузлом, при підключенні вже нового вузла дані мають бути відновлені в ньому. За потреби здійснювати backups чи snapshots та відновлення з них.

5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД

Для поставленої задачі було обрано СУБД MongoDB — об'єктно орієнтована СУБД, що дозволяє зберігати великі масиви неструктурованих даних. На відміну від SQL баз даних ми можемо зберігати дані у “сирому” об'єктному вигляді, який використовується програмою та є більш близьким за структурою до моделі даних, що пришвидшить збір, збереження та отримання

та аналіз даних програмним забезпеченням. Дана СУБД також підтримує горизонтальне масштабування за допомогою шардингу з метою зменшення навантаження на кожен окремий вузол шляхом розподілення навантаження між ними всіма.

6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА

Інтерфейс користувача виконати в консольному варіанті. Повинні бути доступні наступні розділи (пункти в меню) в інтерфейсі: налаштування засобів та підсистем (генерації даних, масштабування, резервування та відновлення даних), запуск/завершення їх роботи, генерація звітної інформації (графіків, діаграм) у вигляді збережених файлів-зображень. Звітна інформація стосується візуалізації роботи засобів аналізу даних (визначення трендів, регресії, кореляції).

7. ВИБІР ЗАСОБІВ РОЗРОБКИ

У якості засобів розробки було обрано мову Python 3.6. Для взаємодії з СУБД – бібліотеку PyMongo. Для аналізу даних повинні використовуватися відповідні бібліотеки напрямку Data Science чи Big Data, як-от: numpy, pandas, matplotlib, scipy.

8. ЕТАПИ РОЗРОБКИ

№	Назва етапів розроблення	Термін виконання
1	Затвердження теми курсової роботи. Опрацювання відповідної літератури. Розроблення та узгодження технічного завдання.	12.03.2020
2	Аналіз постановки задачі	22.03.2020
3	Розробка засобів генерації даних.	05.04.2020
4	Додавання засобів фільтрації та валідації даних.	12.04.2020

5	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи.	19.04.2020
6	Додавання засобів аналізу даних.	03.05.2020
7	Додавання засобів резервування та відновлення даних (з урахуванням необхідності підключення додаткового комп'ютера як елемента горизонтального масштабування).	10.05.2020
8	Тестування програми	15.05.2020
9	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	17.05.2020
10	Захист курсової роботи	21.05.2020