

**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 аналізу популярності мобільних
додатків на сервісах цифрової дистрибуції Google Play Store та Apple App
Store**

Студент

групи КП-82

Мельничук О. Г.

(ПІБ)

(підпис)

Викладач

**к.т.н, доцент кафедри
СПіСКС**

Петрашенко А.В.

(підпис)

Київ – 2021

1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ

Найменування: моніторингова система аналізу популярності мобільних додатків на сервісах цифрової дистрибуції Google Play Store та Apple App Store.

Галузь застосувань: ринок мобільних застосунків.

2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ

Дата початку проекту - 5 березня 2021 року (дата видачі завдання курсового проекту).

Дата закінчення проекту - 20 травня 2021 року захист курсового проекту).

3. МЕТА РОЗРОБКИ

Метою розробки курсового проекту є набуття студентом практичних навичок розробки сучасного програмного забезпечення, що взаємодіє з постреляційними базами даних, а також здобуття навичок оформлення відповідного текстового, програмного та ілюстративного матеріалу у формі проектної документації.

У результаті виконання курсового проекту студент повинен вміти розробляти програмне забезпечення для постреляційних баз даних, володіти основами використання СУБД, а також інструментальними засобами аналізу великих обсягів даних.

4. ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

- Підсистема попередньої обробки даних містить у собі:
 - Засоби генерації даних: має бути реалізовано ПЗ для збору інформації про мобільні застосунки.
- Засоби фільтрації та валідації даних:
 - Доповнення ПЗ з попереднього пункту функціоналом фільтрації та перевірки даних на коректність, відсіювання дублікатів.
- База даних:

- MongoDB
- Засоби реплікації даних:
 - MongoDB підтримує реплікацію з допомогою replica sets (наборів реплік) та master-slave. В даному проекті будуть використовуватися replica sets.
- Засоби масштабування:
 - MongoDB імплементує шардинг (sharding – метод розподілу даних між кількома вузлами) як метод масштабування.
- Засоби аналізу даних:
 - NumPy — бібліотека з математичними функціями та підтримкою багатомірних масивів;
 - Scrapy — бібліотека для збору даних з веб-сайтів;
 - Pandas — аналітична бібліотека для обробки та аналізу даних;
 - Matplotlib — бібліотека для представлення даних в графічному вигляді;
 - Scikit-learn — бібліотека для машинного навчання що побудована на NumPy та Matplotlib;
 - StatsModels — бібліотека для статистичного моделювання даних.
- Задачі аналізу даних:
 - Структурувати усі отримані дані з різних веб-ресурсів про популярність мобільних застосунків на платформах.
 - Провести валідацію даних та прибрати непотрібну інформацію.
 - Об'єднати інформацію та заповнити пусті місця у масивах даних.
 - Провести кореляцію даних та провести детальний аналіз у відповідних місцях кореляції.
 - Знайти зв'язок між популярністю та категорією мобільних додатків; сформулювати можливий прогноз на базі отриманих даних.
- Засоби резервування та відновлення даних:
 - Передбачені при використанні replica sets у MongoDB.

5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД

В даному курсовому проєкті буде використана MongoDB.

MongoDB є нереляційною (not only SQL) документо-орієнтованою базою даних з відкритим кодом. MongoDB дозволяє зберігати великі масиви неструктурованих даних і не потребує опису схем таблиць. MongoDB забезпечує швидку обробку великих наборів даних та їх приведення до нормальних форм. MongoDB підтримує горизонтальне масштабування з допомогою шардингу (метод розподілу даних між кількома вузлами).

6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА

Інтерфейс користувача буде реалізований у консольному вигляді.

Інтерфейс користувача буде використовуватися для запуску на виконання ПЗ, його налаштування та передачі параметрів для збору та аналізу даних, генерації звітної інформації у вигляді графіків та діаграм що зберігаються на диск в якості зображень. Звітна інформація передбачає візуалізацію роботи засобів аналізу даних.

7. ВИБІР ЗАСОБІВ РОЗРОБКИ

Для проєкту була вибрана мова програмування Python версії 3.9. Дана мова програмування є простою та має достатню кількість аналітичних бібліотек для задач подібного роду. Бібліотеки:

- Scrapy — бібліотека для збору даних з веб-сайтів;
- NumPy — бібліотека з математичними функціями та підтримкою багатомірних масивів;
- Pandas — аналітична бібліотека для обробки та аналізу даних;
- Matplotlib — бібліотека для представлення даних в графічному вигляді;
- Scikit-learn — бібліотека для машинного навчання що побудована на NumPy та Matplotlib;
- StatsModels — бібліотека для статистичного моделювання даних.

8. ЕТАПИ РОЗРОБКИ

№	Назва етапів розроблення	Термін виконання
1	Розроблення та узгодження теми роботи та технічного завдання	25.03.2021
2	Розробка засобів збору та генерації даних	10.04.2021
3	Реалізація засобів фільтрації та валідації даних	15.04.2021
4	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи	17.04.2021
5	Додавання засобів аналізу даних (реалізацію алгоритмів буде запозичено у великих бібліотеках аналізу даних).	20.04.2021
6	Додавання засобів резервування та відновлення даних (призначені для оперативного та пакетного збереження фрагментів та усієї бази даних з можливістю її відновлення з урахуванням необхідності підключення додаткового комп'ютера як елемента горизонтального масштабування).	27.04.2021
7	Тестування програми	1.05.2021
8	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	10.05.2021
9	Захист курсової роботи	20.05.2021