

**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 захворюваності на коронавірус у світі

**Студент**

групи КП-73

Булаєвський І.О.

\_\_\_\_\_

(підпис)

**Викладач**

к.т.н, доцент кафедри

СПіСКС

Петрашенко А.В.

\_\_\_\_\_

(підпис)

Київ – 2020

## **1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ**

Найменування: моніторингова система захворюваності на коронавірус у світі: можливе дослідження даних у відкритих онлайн джерелах, збір інформації про кількість щоденних нових захворювань та летальних випадків.

Галузь застосування: прогнозування розвитку епідемії в країнах, яких епідемія ще не пішла на спад, на прикладі країн, які вже зупинили поширення коронавірусу. Збір, аналіз та фільтрація даних з відкритих онлайн джерел. Визначення орієнтовних дат початку різних етапів виходу з карантину в Україні.

## **2. ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ**

Дата початку проекту - 26 лютого 2020 року (дата видачі завдання курсового проекту).

Дата закінчення проекту - ? 2020 року ( захист курсового проекту).

## **3. МЕТА РОЗРОБКИ**

Метою розробки даного курсового проекту є набуття виконавцем (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 летальні випадки та нові захворювання, що розміщені на статистичному веб-сайті: <https://www.worldometers.info/coronavirus/>.
- Засоби фільтрації та валідації даних:
  - Доповнення ПЗ з попереднього пункту функціоналом фільтрації та перевірки даних на коректність, відсіювання дублікацій.
- База даних:

- MongoDB
- Засоби реплікації даних:
  - оскільки для використання у даній розробці була обрана нереляційна база даних MongoDB, то як засіб реплікації використовуватимемо реплісети (Replica Sets).
- Засоби масштабування:
  - шардинг (sharding), який використовує MongoDB для обробки великої кількості даних.
- Засоби аналізу даних:
  - Pandas – бібліотека, яка використовуватиметься як надбудова до NumPy для структуризації роботи із масивами даних.
  - NumPy – бібліотека для роботи із великими масивами даних.
  - Matplotlib – бібліотека для візуалізації даних у вигляді 2D і 3D графіків.
  - Scikit-learn – бібліотека для використання алгоритмів машинного навчання.
- Задачі аналізу даних:
  - Структурувати усі отримані дані з різних веб-ресурсів про нові випадки: країна, кількість нових захворювань, кількість нових летальних випадків.
  - Здійснити валідацію даних та прибрати зайву інформацію із усіх масивів даних.
  - Об'єднати інформацію та заповнити пусті місця у даних.
  - Провести кореляцію по всіх даних.
  - У місцях, де найбільша кореляція, здійснити більш детальний аналіз.
- Засоби резервування та відновлення даних:
  - Передбачені при використанні Replica Sets у MongoDB.

## **5. ОБҐРУНТУВАННЯ ВИБОРУ СУБД**

В даному курсовому проєкті використовується MongoDB. Ця база даних є об'єктно орієнтованою та 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, яку буде використовувати ПЗ написане з використанням мови програмування Python. Це пришвидшить збір, збереження та отримання даних програмним забезпеченням. Оскільки MongoDB є представником NoSQL баз даних, вона не потребує жорсткої схеми даних, що дозволяє пришвидшити процес розробки та зробити його більш гнучким. Окрім цього дана СУБД 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 буде консольним (за завданням такий варіант реалізації інтерфейсу дозволено). Інтерфейсу користувача буде використовуватися для запуску на виконання ПЗ, його налаштування та передачі параметрів для збору та аналізу даних, генерації звітної інформації (графіків та діаграм) у вигляді збережених на диск файлів - зображень. Звітна інформація стосується візуалізації роботи засобів аналізу даних.

## **7. ВИБІР ЗАСОБІВ РОЗРОБКИ**

Мова програмування – Python 3.8. Дана мова програмування відносно проста та за думкою фахівців, які займаються розробкою ПЗ у сфері науки про дані, є найкращим рішенням для даного роду задач. Окрім цього для цієї мови існує велика кількість бібліотек для аналізу даних. Зокрема:

- sklearn — бібліотека алгоритмів машинного навчання, використовується для класифікації досліджених даних;
- pandas — бібліотека для обробки та аналізу даних, використовується для первинної обробки даних;
- matplotlib — бібліотека для графічного представлення даних, а саме для побудови графіків;
- numpy — математична бібліотека з підтриманням багатомірних масивів;

## 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	20.03.2020
3	Розробка засобів генерації даних.	15.04.2020
4	Додавання засобів фільтрації та валідації даних.	20.04.2020
5	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи.	25.04.2020
6	Додавання засобів аналізу даних (реалізацію алгоритмів буде запозичено у великих бібліотеках аналізу даних).	30.04.2020
7	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).	05.05.2020
8	Тестування програми	10.05.2020
9	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	15.05.2020
10	Захист курсової роботи	?

