

**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: Система аналізу поширення захворюваності на “Covid-19”

**Студент**

групи КП-83

Ландо Максим Юрійович

\_\_\_\_\_  
(підпис)

**Викладач**

к.т.н, доцент кафедри СПіСКС

Петрашенко А.В.

\_\_\_\_\_  
(підпис)

Київ – 2021

## **ЗМІСТ**

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

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

Тема роботи: Система аналізу поширення захворюваності на “Covid-19”.

Галузь застосування: складання висновків щодо тенденцій поширення захворюваності на “Covid-19”.

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

Дата початку проекту: 22 лютого 2021 року (дата видачі завдання курсового проекту).

Дата закінчення проекту: 25 червня 2021 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: дані беруться з відкритого датасету за посиланням [Download historical data \(to 14 December 2020\) on the daily number of new reported COVID-19 cases and deaths worldwide \(europa.eu\)](https://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&code=sdg-11-6-2&plugin=1)
- Засоби фільтрації та валідації даних:
  - Доповнення ПЗ з попереднього пункту функціоналом фільтрації отриманих з датасету даних, виправлення неправильних форматів та значень
- База даних:
  - MongoDB

- Засоби аналізу даних:
  - Pandas – бібліотека, яка використовуватиметься як надбудова до NumPy для структуризації роботи із масивами даних.
  - Matplotlib – бібліотека для візуалізації даних у вигляді 2D і 3D графіків.
- Задачі аналізу даних:
  - Структурувати усі отримані дані веб-ресурсу.
  - Здійснити валідацію даних та прибрати зайву інформацію із усіх масивів даних.
  - Об'єднати інформацію та заповнити пусті місця у даних.
  - Отримати кількість випадків захворюваності та смертність за певний період.
  - Знайти середні та медіанні значення по захворюваності та смертності у країні.
  - Виявити тенденції по захворюваності та смертності у країні.
  - Виявити країни з найбільшою та найменшою кількістю хворих та смертністю.
- Резервування та відновлення даних
  - Засоби й ресурси, які входять до складу сервера бази даних 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. Звітна інформація стосується візуалізації роботи засобів аналізу даних.

## **7. ВИБІР ЗАСОБІВ РОЗРОБКИ**

Мова програмування – Python 3.9. 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:

- pandas — бібліотека для обробки та аналізу даних, використовується для первинної обробки даних;
- matplotlib — бібліотека для графічного представлення даних, а саме для побудови графіків.
- pymongo – бібліотека, яка дозволяє взаємодіяти з базою даних MongoDB, створювати зв'язок з нею, отримувати та передавати дані.

## 8. ЕТАПИ РОЗРОБКИ

№	Назва етапів розроблення	Термін виконання
1	Затвердження теми курсової роботи. Опрацювання відповідної літератури. Розроблення та узгодження технічного завдання.	22.02.2021
2	Аналіз постановки задачі	28.02.2021
3	Розробка засобів генерації даних.	15.03.2021
4	Додавання засобів фільтрації та валідації даних.	20.03.2021
5	Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи.	25.03.2021
6	Додавання засобів аналізу даних (реалізацію алгоритмів буде запозичено у великих бібліотеках аналізу даних).	30.03.2021
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).	05.04.2021
8	Тестування програми	10.04.2021
9	Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки	15.05.2021
10	Захист курсової роботи	25.05.2021