**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 Telegram каналів

**Студент Садрицький С.В. \_\_\_\_\_\_\_\_\_**

**групи ​ КП-61(62)** (ПІБ) (підпис)

**Викладач**

**к.т.н, доцент кафедри**

**СПіСКС**  **Петрашенко А.В.** \_\_\_\_\_\_\_\_\_

(підпис)

Київ – 2019

**1. НАЙМЕНУВАННЯ ТА ГАЛУЗЬ ЗАСТОСУВАННЯ РОЗРОБКИ**

Найменування: моніторингова система каналів Telegram: можливе замірювання cтатистики повідомлень та активності підписок на вибраний канал в месенджері Телеграм.

Галузь застосування: інтернет-маркетинг.

2.**ДАТА ПОЧАТКУ ТА ЗАКІНЧЕННЯ ПРОЕКТУ**

Дата початку проекту – 23 березня 2019 року (дата видачі завдання для курсового проекту).

Дата закінчення проекту – 20 травня 2019 року (захист курсового проекту).

3. **МЕТА РОЗРОБКИ**

Метою розробки даного курсового проекту є набуття практичних навичок розробки сучасного програмного забезпечення, що взаємодіє з постреляційними базами даних, інструментальними засобами аналізу великих обсягів даних, а також здобуття навичок оформлення відповідного текстового, програмного та графічного матеріалу у формі проектної документації.

4. **ВИМОГИ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ**

Проект передбачено реалізувати в середовищі розробки PyCharm на ОС Linux(Ubuntu). Для реалізації даного курсового проекту буде спроектовано базу даних у СКБД MongoDB.

* **генерація даних**

Дані в БД буде отримано шляхом запитів до API месенджера Telegram. Ці дані включають текстові та цифрові значення, такі як число підписників на каналі, кількість постів та підписок за деякий проміжок часу тощо.

* **фільтрація і валідація даних**

Для вибірки і фільтрації даних передбачається використання можливостей Aggregation Framework.

* **засоби реплікації**

Реплікація БД проекту буде реалізована через вбудовані засоби MongoDB, як replicaSet, Mongos тощо.

* **засоби масштабування**

Для забезпечення реплікації буде використано вбудовані засоби шардінгу MongoDB.

* **засоби аналізу даних**

Для аналізу даних з БД передбачається використання Aggregation Framework та за потреби інших бібліотек мови python, як pandas, sklearn тощо.

* **засоби резервного копіювання і відновлення даних**

Для забезпечення резервного копіювання і відновлення буде також використано вбудовані засоби СКБД.

5. **ОБҐРУНТУВАННЯ ВИБОРУ СУБД**

**MongoDB** — [документо-орієнтована система керування базами даних](https://uk.wikipedia.org/wiki/%D0%94%D0%BE%D0%BA%D1%83%D0%BC%D0%B5%D0%BD%D1%82%D0%BE-%D0%BE%D1%80%D1%96%D1%94%D0%BD%D1%82%D0%BE%D0%B2%D0%B0%D0%BD%D0%B0_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D0%BA%D0%B5%D1%80%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F_%D0%B1%D0%B0%D0%B7%D0%B0%D0%BC%D0%B8_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85) (СКБД), яка не потребує опису схеми таблиць. Це є основним критерієм вибору, адже дані, отримані через API Telegram можуть мати різнорідний характер. Більше того, оскільки вимоги проекту передбачають розгортання БД на декількох комп’ютерах. MongoDB - компроміс між швидкими і масштабованими системами, що оперують даними у форматі ключ/значення, і [реляційними СКБД](https://uk.wikipedia.org/wiki/%D0%A0%D0%B5%D0%BB%D1%8F%D1%86%D1%96%D0%B9%D0%BD%D0%B0_%D0%A1%D0%9A%D0%91%D0%94), функціональними і зручними у формуванні запитів.

**6. ВИМОГИ ДО ІНТЕРФЕЙСУ КОРИСТУВАЧА**

Інтерфейс користувача буде консольним (за завданням такий варіант реалізації інтерфейсу дозволено). Задачею інтерфейсу користувача є взаємодія через консольні команди: діставання потрібних даних через API Telegram, вибірка даних, аналіз засобами сторонніх бібліотек, запуск/завершення роботи програми, генерація та вивід результатів у файли та в консоль.

**7. ВИБІР ЗАСОБІВ РОЗРОБКИ**

Мова програмування – Python 3.6. Була вибрана саме ця мова, оскільки вона відносно проста в освоєнні та має широкий спектр різних сторонніх бібліотек для аналізу та обробки даних.

Плередбачається використання наступних бібліотек:

* **telethon** - python-обгортка над стандартним Telegram API.
* **вбудовані можливості СКБД MongoDB +Aggregation Framework** для вибірки та фільтрації даних в БД.

За необхідністю у додатковому аналізі можуть бути використані наступні бібліотеки:

* **pandas** — бібліотека для первинної обробки та аналізу даних,
* **sklearn**  — бібліотека алгоритмів машинного навчання для класификації досліджених даних;
* **matplotlib** — бібліотека для графічного представлення даних(для побудови графіків тощо)
* **numpy** — математична бібліотека з підтриманням багатомірних масивів і широкого спектру математичних операцій.

8. **ЕТАПИ РОЗРОБКИ**

|  |  |  |
| --- | --- | --- |
| **№** | **Назва етапів розроблення** | **Термін виконання** |
| **1** | Затвердження теми курсової роботи. Розроблення та узгодження технічного завдання. | 9.04.19 |
| **2** | Аналіз постановки задачі | 14.04.19 |
| **3** | Розробка засобів отримання даних. | 19.04.19 |
| **4** | Додавання засобів фільтрації та валідації даних. | 24.04.19 |
| **5** | Реалізація зберігання, реплікації та масштабування інформації розробленої моніторингової системи. | 29.04.19 |
| **6** | Додавання засобів аналізу даних (реалізацію алгоритмів буде запозичено у великих бібліотеках аналізу даних). | 4.05.19 |
| **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). | 9.05.19 |
| **8** | Тестування програми | 14.05.19 |
| **9** | Аналіз результатів. Підготовка матеріалів курсового проекту та оформлення пояснювальної записки | 17.05.19 |
| **10** | Захист курсової роботи | 20.05.19 |