**Санкт-Петербургский государственный университет**

**Факультет прикладной математики и процессов управления**

**ОТЧЁТ**

**по дисциплине “Функциональное программирование”**

**об многопоточном анализе данных из VK и Telegram**

Автор работы: Горбунов К. Д. Группа: 22.Б16-пу

Преподаватель: Киямов Ж. У.

Санкт-Петербург

2023 г.

**Содержание**

[**1. Введение** 4](#_Toc147273281)

[**2. Цель работы** 4](#_Toc147273282)

[**3. Задачи** 4](#_Toc147273283)

[**4. Техническое описание** 5](#_Toc147273284)

[**4.1 Используемые инструменты и библиотеки** 5](#_Toc147273285)

[**5 Описание программы** 5](#_Toc147273286)

[**5.1 Описание ключевых компонентов программы** 5](#_Toc147273287)

[**5.2 Общий ход программы** 7](#_Toc147273287)

[**6. Рекомендации пользователю** 9](#_Toc147273284)

[**7. Рекомендации программисту** 10](#_Toc147273286)

[**8. Вывод** 10](#_Toc147273284)

[**9. Библиография** 11](#_Toc147273284)

**1. Введение**

Данный отчет описывает программу для анализа текстовых данных из социальных сетей VK (ВКонтакте) и Telegram с использованием библиотек и инструментов Python. Программа предназначена для сбора, предобработки и анализа текстовых данных из указанных социальных платформ.

**2. Цель работы**

Целью данной работы является разработка программы, которая позволяет собирать текстовые данные из VK и Telegram, проводить предварительную обработку этих данных и анализировать их для выявления ключевых слов и фраз.

**3. Задачи**

Основными задачами программы являются:

* Сбор данных из VK и Telegram с использованием API соответствующих платформ.
* Предобработка собранных данных, включая токенизацию, приведение к нижнему регистру и удаление стоп-слов и пунктуации.
* Анализ текстовых данных для выявления ключевых слов и фраз.
* Отображение результатов анализа в графическом интерфейсе на основе библиотеки tkinter.

**4. Техническое описание**

**4.1 Используемые инструменты и библиотеки**

* Python: Язык программирования, на котором написана программа.
* VK API: Библиотека для работы с API VKontakte.
* Telethon: Библиотека для работы с Telegram API.
* NLTK и spaCy: Библиотеки для обработки естественного языка и токенизации текста.
* tkinter: Библиотека для создания графического интерфейса пользователя.
* threading: Модуль для многопоточности.
* queue: Модуль для работы с очередями данных.
* string: Модуль для работы со строками и пунктуацией.
* collections: Модуль для работы с контейнерами.

**4.2 Структура программы**

* perform: Основная функция, запускающая сбор данных из VK и Telegram, их предобработку и анализ, а также отображение результатов в графическом интерфейсе.
* scrape\_vk: Функция для сбора данных из VK.
* scrape\_telegram: Функция для сбора данных из Telegram.
* preprocess: Функция для предобработки собранных данных.
* analyze: Функция для анализа текста и выявления ключевых слов.
* Графический интерфейс с использованием tkinter для отображения результатов анализа.

**5** **Описание программы**

**5.1 Описание ключевых компонентов программы**

В таблице 1 представлено описание ключевых компонентов программы.

Таблица 1. Ключевые компоненты

|  |  |
| --- | --- |
| Компонент | Описание |
| perform | Главная функция, координирующая работу программы. Она запускает сбор данных из VK и Telegram, предобработку данных, а затем анализ и вывод результатов. |
| scrape\_vk и scrape\_telegram | Функции для сбора данных из соответствующих социальных сетей с использованием их API. |
| preprocess | Функция для предобработки текстовых данных, включая токенизацию, приведение к нижнему регистру и удаление стоп-слов и пунктуации. |
| analyze | Функция для анализа текста с использованием библиотеки spaCy и выявления ключевых слов и фраз. |
| Графический интерфейс на tkinter | Создает окно с кнопкой для запуска анализа и текстовыми полями для вывода результатов. |

**5.2 Общий ход программы**

Общий ход программы можно описать следующим образом:

1. Пользователь запускает программу и нажимает кнопку "Проанализировать данные".
2. Программа начинает сбор данных из VK и Telegram с помощью соответствующих функций.
3. Собранные данные подвергаются предобработке, где они токенизируются, приводятся к нижнему регистру и удаляются стоп-слова и пунктуация.
4. После предобработки данные анализируются для выявления ключевых слов и фраз.
5. Результаты анализа выводятся в графическом интерфейсе на tkinter в соответствующие текстовые поля.

На рис. 1 приведено окно пользователя.

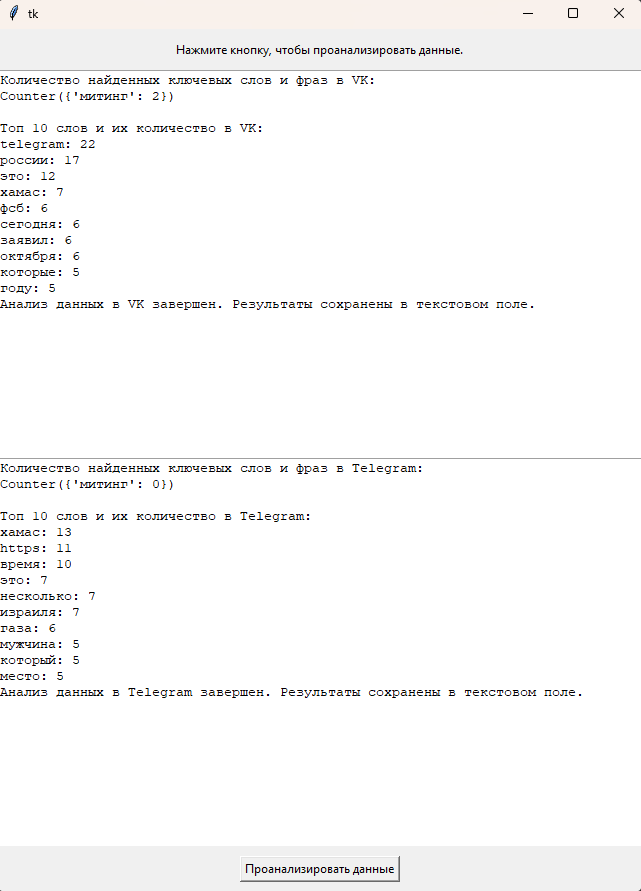


Рис. 1 Окно пользователя

**6. Рекомендации пользователю**

* Убедитесь, что ваши ключи доступа (API-ключ VK, ID и хэш Telegram) корректно указаны в коде программы.
* Перед началом работы удостоверьтесь, что у вас есть доступ к интернету, так как программа использует API социальных сетей для сбора данных.

**7. Рекомендации программисту**

* **Проверка наличия необходимых библиотек**: Убедитесь, что на устройстве, на котором будет запускаться программа, установлены все необходимые библиотеки, указанные в техническом описании программы (например, NLTK, spaCy, tkinter и др.). В случае отсутствия какой-либо библиотеки, установите ее с помощью менеджера пакетов Python (например, pip).
* **Обновление библиотек**: Проверьте актуальность установленных библиотек и обновите их до последних версий, чтобы избежать возможных проблем с совместимостью.
* **Создание виртуального окружения**: Рекомендуется создать виртуальное окружение Python для данной программы. Это позволит изолировать зависимости программы от других проектов и обеспечит чистую среду для работы программы.
* **Настройка безопасности**: Проверьте настройки брандмауэра и антивирусного ПО, чтобы убедиться, что программа имеет доступ к сети и не блокируется. В случае необходимости, добавьте исключения для портов и протоколов, которые используются программой (например, HTTP/HTTPS).
* **Проверка доступа к интернету**: Удостоверьтесь, что устройство, на котором будет работать программа, имеет доступ к интернету. Программа использует API социальных сетей для сбора данных, поэтому соединение с сетью обязательно.
* **Настройка окружения**: При необходимости настройте окружение для работы с API VK и Telegram, включая правильное размещение и обновление ключей доступа, API ID и хэшей. Убедитесь, что ключи доступа актуальны и имеют необходимые разрешения.

**8. Вывод**

Программа для анализа текстовых данных из социальных сетей VK и Telegram разработана и успешно выполняет сбор, предобработку и анализ данных. Результаты анализа выводятся в графическом интерфейсе, что делает использование программы более удобным. Для дальнейшего развития программы можно добавить дополнительные функции и улучшения.

**Библиография**

**Python (официальный сайт)**:

* Сайт: <https://www.python.org/>

**Библиотеки, используемые в программе**:

**vk\_api (работа с VK API)**:

* PyPI: <https://pypi.org/project/vk-api/>
* Репозиторий: <https://github.com/python273/vk_api>

**Telethon (работа с Telegram API)**:

* PyPI: <https://pypi.org/project/telethon/>
* Репозиторий: <https://github.com/LonamiWebs/Telethon>

**NLTK (библиотека обработки естественного языка)**:

* PyPI: <https://pypi.org/project/nltk/>
* Репозиторий: <https://github.com/nltk/nltk>

**spaCy (библиотека обработки естественного языка)**:

* PyPI: <https://pypi.org/project/spacy/>
* Репозиторий: <https://github.com/explosion/spaCy>

**tkinter (графический интерфейс Python)**:

* Сайт: <https://docs.python.org/3/library/tkinter.html>

**threading (модуль многопоточности Python)**:

* Сайт: <https://docs.python.org/3/library/threading.html>

**queue (модуль очередей Python)**:

* Сайт: <https://docs.python.org/3/library/queue.html>

**string (модуль строк Python)**:

* Сайт: <https://docs.python.org/3/library/string.html>

**collections (модуль коллекций Python)**:

* Сайт: <https://docs.python.org/3/library/collections.html>

GitHub-https://github.com/Gorbunov-Konstantin/vk\_and\_telegram\_statistic.git