**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный университет**

**Кафедра прикладной математики и процессов управления.**

**отчет**

**по лабораторной работе №1**

**по дисциплине «Функциональное программирование»**

**на тему «Парсинг и анализ постов соцсетей»**

|  |  |  |
| --- | --- | --- |
| Студент гр. 22.Б16-пу |  | Поляков Л.С. |
| Преподаватель |  | Киямов Ж. У. |

**Санкт-Петербург**

**2023 г.**

**Оглавление**

[**1.** **Введение** 3](#_Toc144847601)

[**2.** **Цель работы** 3](#_Toc144847602)

[**3.** **Задача** 3](#_Toc144847603)

[**4.** **Описание программы** 3](#_Toc144847604)

[**5.** **Рекомендации пользователю** 4](#_Toc144847605)

[**6.** **Рекомендации программисту** 4](#_Toc144847607)

[**7.** **Контрольный пример** 5](#_Toc144847608)

[**8.** **Вывод** 7](#_Toc144847611)

**1. Введение**

Программа разработана для сбора и анализа информации о постах из социальных сетей VKontakte (VK) и Telegram. Она позволяет пользователю ввести необходимые параметры для сбора данных и выводит результаты анализа в графическом интерфейсе, созданном с использованием библиотеки tkinter.

**2. Цель работы**

Целью программы является сбор и анализ данных из VK и Telegram, а также предоставление пользователю удобного интерфейса для выполнения этих задач.

**3. Задачи:**

* Сбор данных о постах из VK и Telegram.
* Предварительная обработка текстов постов.
* Анализ популярных хэштегов и ключевых слов.
* Вывод результатов анализа в графическом интерфейсе.

**4. Описание программы**

Программа написана на языке Python и включает в себя следующий ряд компонентов:  
• Модули:

|  |  |
| --- | --- |
| *tkinter* | Необходим для создания графического интерфейса |
| *vk\_api* | Необходим для сбора информации о постах из VK |
| *os* | Необходим для отслеживания файла сессии |
| *telephon* | Необходим для сбора информации о постах из Telegram |
| *collections* | Необходим для вычисления наиболее часто встречаемых хэштегов и ключевых слов |
| *re* | Необходим для составления регулярных выражений |
| *threading* | Необходим для создания отдельных потоков (парсинг VK) |
| *asyncio* | Необходим для создания отдельных потоков (парсинг Telegram) |
| *time* | Необходим для выставления задержки (для предотвращения ошибок) |

• Функции:

|  |  |
| --- | --- |
| *collect\_data\_from\_vk* | Необходима для сбора информации о заданном количестве постов из VK |
| *collect\_data\_from\_telegram* | Необходима для сбора информации о заданном количестве постов из Telegram |
| *preprocess\_text* | Необходима для удаления лишних символов из текста |
| *analyze\_data* | Необходима для выделения ключевых слов и хэштегов |
| *process\_vk\_data\_thread* | Необходимы для сбора информации о постах и добавления ключевых слов и хэштегов в массив |
| *process\_telegram\_data\_thread* |
| *init\_telegram\_session* | Необходима для создания новой сессии для парсинга постов из Telegram |
| *start\_parsing* | Основная функция, содержащая в себе сбор начальных данных, создание потоков и вывод информации в окне |
| *start\_async\_parsing* | Функция для асинхронного запуска start\_parsing |

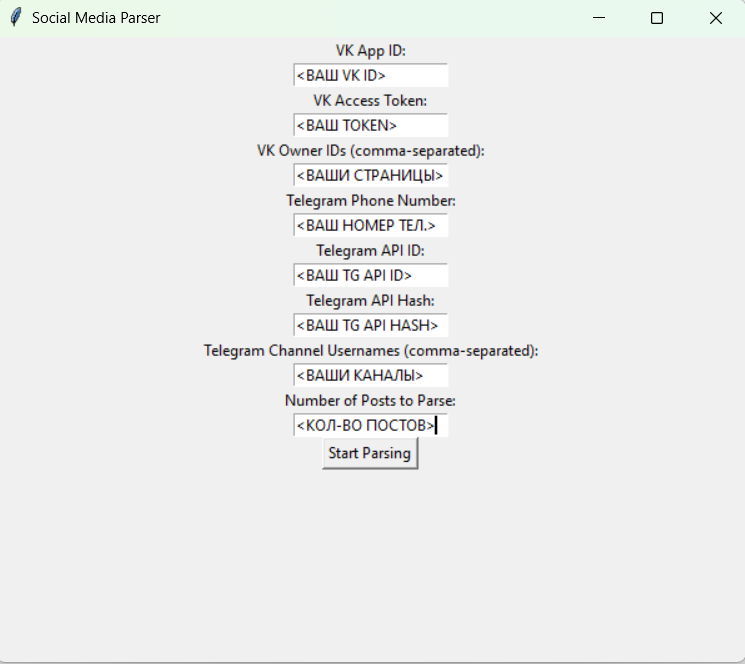
**5. Рекомендации пользователю**

* Введите корректные данные для каждой из социальных сетей (VK и Telegram).
* Убедитесь, что указанные вами группы и страницы доступны для просмотра.
* Выберите необходимое количество потоков для сбора данных с учетом производительности вашего компьютера и доступа к интернету.
* Убедитесь, что у вас установлен Python версии 3.8 и выше.

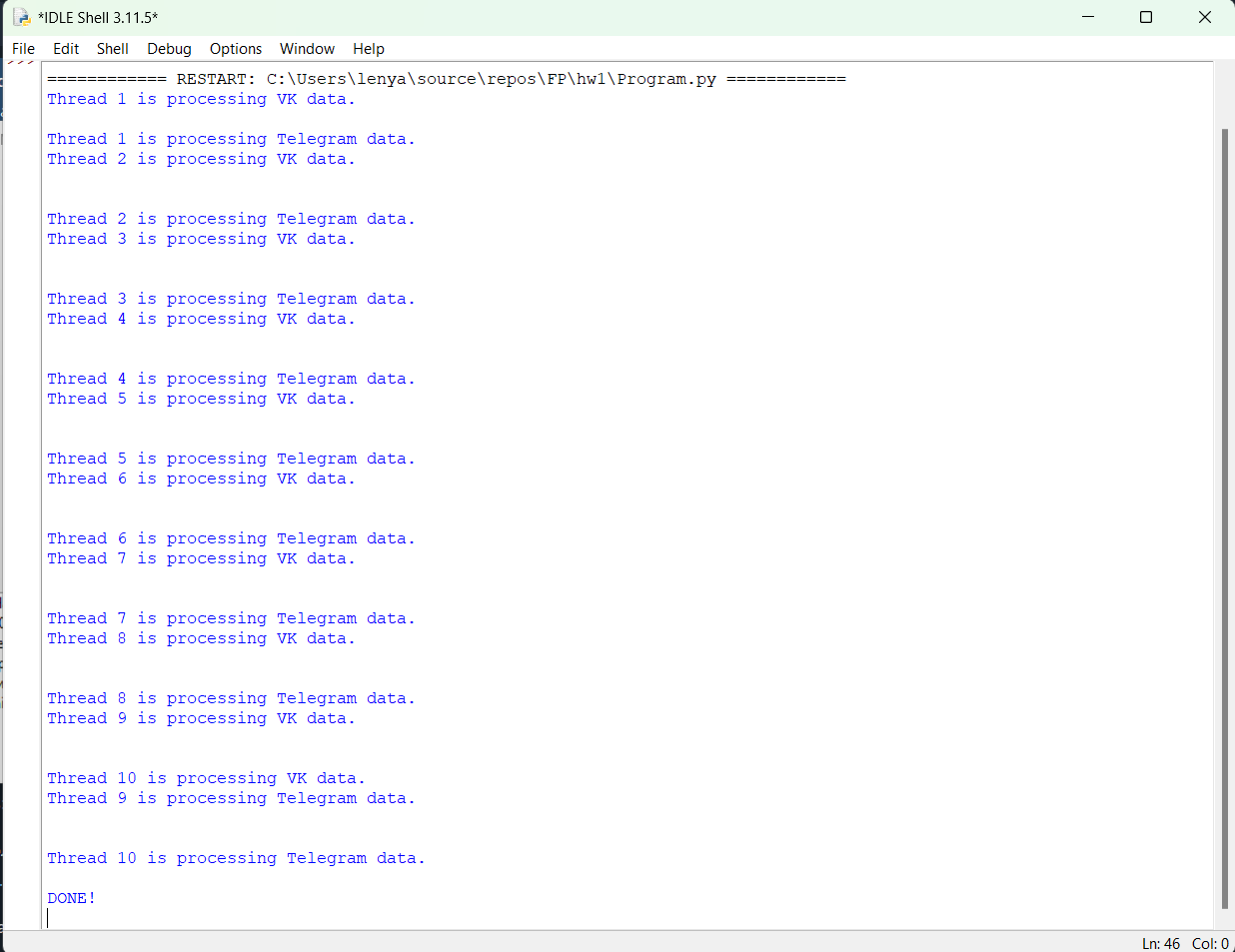
**6. Рекомендации программисту**

* Прокомментируйте код более подробно, чтобы облегчить понимание его работы.
* Добавьте обработку ошибок и исключений для улучшения надежности программы.
* Рассмотрите возможность улучшения пользовательского интерфейса, добавив больше информации и функций для пользователя.
* Убедитесь, что у вас установлен Python версии 3.8 и выше.

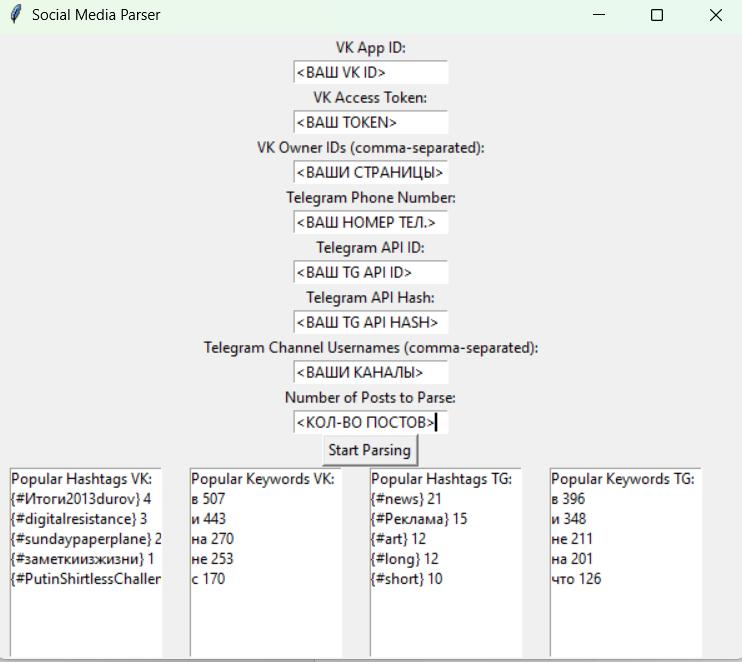
**7. Контрольный пример**



*Рис. 1 (При запуске программы открывается такое окно. В его полях необходимо ввести свои данные, а после нажать на кнопку “Start Parsing”)*

**

*Рис. 2 (После нажатия на кнопку “Start Parsing” в консоли начнется вывод сообщений о процессе парсинга. Отедльно выводится информация о VK и Telegram. Об оконачании сбора постов сигнализирует надпись “DONE!”)*

**

*Рис. 3 (В изначальном окне появится информация о самых популярных ключевых словах и хэштегов из каждой соцсети. Формат вывода: <Ключевое слово/хэштег> <колиество во всех постах>)*

**8. Вывод**

Программа успешно выполняет сбор данных из социальных сетей VK и Telegram, их анализ и вывод результатов в графическом интерфейсе. Дальнейшие улучшения и оптимизации могут сделать программу еще более полезной и удобной для пользователей.

Ссылка на GitHub: -