

ОБЗОР МЕТОДОВ ПАРСИНГА КОММЕНТАРИЕВ TELEGRAM КАНАЛОВ ДЛЯ СБОРА ТЕКСТОВЫХ СООБЩЕНИЙ ГОРОЖАН

Низомутдинов Б.А. (Университет ИТМО)

В данной работе проведен обзор методов по сбору комментариев из каналов мессенджера Telegram, с целью дальнейшей выгрузки дискуссий и контент-анализа. Проведен обзор литературы, документации API Telegram, информации о библиотеках, размещенных в GitHub, а также описаний доступных онлайн-сервисов. Данное исследование проведено в рамках НИР Университета ИТМО № 622264 "Разработка сервиса выявления объектов городской среды общественной активности и ситуаций повышенного риска на основе текстовых сообщений горожан".

Введение.

Telegram за последние несколько лет стал привлекательной площадкой для проведения исследований, это подтверждает ряд исследований [1-3]. Этот инструмент коммуникации привлекает новых пользователей. Каналы Telegram идеально подходят для обмена информацией с друзьями и семьей или совместной работы в небольших командах, а также, для обсуждения городских проблем. Каналы - это инструмент для трансляции сообщений большой аудитории. Сегодня, анализ самих постов в сообществах уже хорошо изучен [4-5], в наших предыдущих работах был описан метод выгрузки сообщений из каналов, разработан конвертер для сохранения в формат CSV основных постов каналов. с дальнейшей обработкой с методами машинного обучения. Однако, не решенным остается вопрос выгрузки комментариев под самими постами, но именно комментарии представляют наибольший интерес в проведении дискурс-анализа. Большие массивы данных, которые генерируют сами жители в информационном пространстве социальных сетей, активно используются в мировой практике для мониторинга городской среды. Наша задача, найти метод для сбора таких комментариев.

Основная часть.

Важно заранее отметить, выгрузка самих сообщений - вопрос решенный, встроенные функции мессенджера позволяют выгружать текст постов из Telegram, в результате можно получить доступ ко всем данным в автономном режиме в формате JSON или в красиво отформатированном HTML, однако, для комментариев такой функции нет.

На первом шаге, для решения задачи сохранения комментариев было решено обратиться к документации API Telegram, однако, анализ документации показал, что доступных методов решения задачи нет.

На следующий шаг были проанализированы научные публикации в системе sciencedirect.com, использовались различные запросы для поиска публикаций по теме, например "Telegram" AND "comments", но поиск также не дал результатов. Была выдвинута гипотеза, что сервисы для работы с Telegram могут иметь подобный функционал, так, были изучены функции сервиса TGStat. TGStat.ru - сервис с открытой статистикой более 800 000 Telegram-каналов и чатов, однако, возможности сохранения комментариев обнаружено не было. Далее, была предпринята попытка поиска ботов, которые умеют выгружать комментарии, однако, поиск показал, что таких ботов не существует.

Положительный результат дал поиск решений на Github, была выявлена библиотека Telescape [6], в документации к которой представлена информация, что с помощью данной библиотеки можно выгрузить сообщения и комментарии, что решает поставленную задачу.

На текущем этапе работа по выгрузке не проводилась, цель работы - поиск инструмента для дальнейшего извлечения информации.

Дополнительно была найдена библиотека Telethon [7]. Эта библиотека предназначена для упрощения написания программ на Python, которые могут взаимодействовать с Telegram.

Выводы. Проведенное исследование демонстрирует, что готовых сервисов по выгрузке комментариев на текущий момент не представлено, однако, существуют различные наработки по решению данного вопроса. Полученные данные после выгрузки позволят проводить дискурс-анализ в городских каналах Telegram и изучать настроение горожан, а также, определять проблемные пространства. С помощью данного метода запланировано продолжение исследования коммуникации жителей с органами власти. На предыдущем этапе проводился анализ комментариев в социальной сети Вконтакте [8], теперь появится возможность сравнения коммуникативных практик в разных источниках.

Список использованных источников:

1. Денисенко С.А., Камус С.Ф., Пименов Ю.Д., Тергоев В.И., Папушев П.Г. Светосильный широкоугольный телескоп АЗТ-33ВМ // Оптический журнал. – 2009. – № 76(10). – С. 48–51.
2. Elebaed H., Saeid E., Elfatih Y. (2021) Assessment of the Applicability of using Telegram as a Learning Management System. International Journal of Information Technology and Language Studies
3. Hakim D. K., Nugroho S. A. (2019) Implementasi Telegram Bot untuk Monitoring Mikrotik Router. SAINTEKS. Vol 16, No 2 DOI:10.30595/st.v16i2.7132
4. Hrastinski, & Stefan. (2019). What do we mean by blended learning? TechTrends 63(5), 564-569.
5. Júnior M., Melo P., Kansaon D., Mafra V. (2022) Telegram Monitor: Monitoring Brazilian Political Groups and Channels on Telegram. Social and Information Networks. P.
6. Telescape. Электронный ресурс. URL: <https://github.com/PeterWalchhofer/Telescape>. Дата обращения 12.02.2023
7. Telethon.Электронный ресурс. URL: <https://docs.telethon.dev/en/stable/>. Дата обращения 12.02.2023
8. Nizomutdinov, B.A., Uglova, A.B. Communication Practicies in the Information Space of City Residents Interested in the Improvement and Greening of Urban Areas (on the Example of St. Petersburg) Proceedings of the 2022 Communication Strategies in Digital Society Seminar 2022, ComSDS 2022, 2022, pp. 23–26

Низомутдинов Б.А. (автор)

Подпись