

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Департамент программной инженерии

СОГЛАСОВАНО

УТВЕРЖДАЮ

НИУ ВШЭ, Приглашенный
Преподаватель

Академический руководитель
образовательной программы
«Программная инженерия» профессор
департамента программной инженерии,
канд. техн. наук

_____ А. А. Топтунов
« ____ » _____ 2024 г.

_____ Н. А. Павлочев
« ____ » _____ 2024 г.

ВИЗУАЛИЗАЦИЯ СОЦИАЛЬНОГО ГРАФА
ПОЛЬЗОВАТЕЛЯ TELEGRAM

Пояснительная записка

Лист УТВЕРЖДЕНИЯ

RU.17701729.05.04-01 ТЗ 01-1-ЛУ

Исполнитель: Студент группы БПИ 213
_____ Н. А. Бирюлин
« ____ » _____ 2024 г.

УТВЕРЖДЁН
RU.17701729.05.04-01 ТЗ 01-1-ЛУ

ВИЗУАЛИЗАЦИЯ СОЦИАЛЬНОГО ГРАФА
ПОЛЬЗОВАТЕЛЯ TELEGRAM

Пояснительная записка

RU.17701729.05.04-01 ТЗ 01-1

Листов 12

Инв. № подл	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Содержание

1	Введение	3
1.1	Наименование программы	3
2	Основание для разработки	4
2.1	Документы, на основании которых ведется разработка	4
3	Технические характеристики	5
3.1	Назначение программы	5
3.2	Алгоритм работы программы	5
3.3	Входные и выходные данные	6
4	Технические и программные средства	8
4.1	Необходимые технические средства	8
4.2	Обоснование необходимых технических средств	8
4.3	Необходимые программные средства	8
4.4	Обоснование необходимых программных средств	8
5	Ожидаемые технико-экономические показатели	9
5.1	Экономическая эффективность	9
5.2	Предполагаемая потребность	9
5.3	Преимущества разработки по сравнению с отечественными и зарубежными аналогами	9
6	Приложение 1. Терминология	10

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. Введение

1.1. Наименование программы

1.1.1. Наименование программы на русском языке

Визуализатор социального графа «Telegram».

1.1.2. Наименование программы на английском языке

«Telegram» social graph visualizer.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. Основание для разработки

2.1. Документы, на основании которых ведется разработка

Учебный план подготовки бакалавров по направлению 09.03.04 «Программная инженерия» и утвержденная академическим руководителем программы тема курсового проекта.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. Технические характеристики

3.1. Назначение программы

3.1.1. Функциональное назначение программы

Функциональным назначением программы является сводный и реляционный анализ данных, экспортируемых из аккаунтов Telegram.

3.1.2. Эксплуатационное назначение программы

Приложение позволяет пользователю находить связи между пользователями в выгрузке сообщений из Telegram. После импорта .json файла с данными, пользователь может получить для заданного списка пользователей граф их связей. Под связями подразумеваются как прямые (А писал В), так и косвенные (А и В писали в один чат, А упоминал В в сообщении).

3.2. Алгоритм работы программы

Перед проведением анализа пользователь выбирает .json файл на своем устройстве. Этот файл передается в веб-приложение. При этом файл не загружается на удаленный сервер - вся обработка производится на устройстве. Это, кроме прочего, позволяет экономить время на передаче файлов (экспорт данных может достигать нескольких гигабайт).

После загрузки вычисляется граф зависимостей, имеющий следующую структуру:

```
class Graph {  
    chatIdToName = {};  
    chatNameToId = {};  
  
    actorIdToName = {};  
    actorNameToId = {};  
  
    chatIdToMessages = {}; // Messages sent  
    actorIdToMessages = {};  
  
    actorToChatMapping = {}; // User writing something in chat. [user_id][chat_id] = [  
        messages]  
    actorToActorMapping = {}; // User mentioning another user. [user_a_id][user_b_id] = [  
        messages]  
  
    chatCategories = {};  
}
```

Собираются отношения Telegram ID к видимым именам для всех объектов (пользователи, чаты, каналы). Они используются для упрощения взаимодействия с пользователем (автодополнение, имена в графах вместо ID).

Кроме того, сообщения группируются по чатам и по пользователям. На этом этапе строится базовая связь чатов с пользователями (пользователь писал в чат) и пользователей с пользователями (пользователь упоминал пользователя в одном из своих сообщений). Стоит заметить, что построение связей упоминаний пользователей в может найти интересные связи, но при этом приводит к созданию в графе несуществующих связей. Например, при тестировании обнаружилось, что очень много пользователей связаны с одним конкретным пользователем. Причина была в том, что имя пользователя - "Да что, конечно, много где упоминалось в переписке.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Данные вычисляются один раз, после чего построение графа использует заранее просчитанный набор связей. Это позволяет значительно уменьшить задержку при взаимодействии с пользователем и сделать скорость построения графа незначительной. Так, даже если время импорта файла будет порядка 20-30 секунд, непосредственно построение графа всегда будет измеряться в миллисекундах.

3.3. Входные и выходные данные

3.3.1. Организация входных данных

Входные данные - .json файл, экспортированный из приложения Telegram. //TODO Ссылка на доку по экспорту

3.3.2. Обоснование метода организации входных данных

Для подобного анализа требуется, с одной стороны, проанализировать большой набор данных (речь может идти о миллионах сообщений), с другой стороны, сделать это без потери доверия пользователя.

Анализ экспорта позволяет нам получить все сообщения не взаимодействуя с Telegram напрямую (а такое большое количество запросов к API могло бы привести к блокировке), что делает сервис надежнее и быстрее. Более того, разница по скорости - не просто вопрос нескольких секунд: при прямом взаимодействии с Telegram речь идет о замедлении как минимум на несколько порядков, т.к. Telegram ограничивает количество запросов, которые могут быть выполнены сторонними клиентами.

Кроме того, так пользователю не нужно давать доступ к своему Telegram-аккаунту (что, для большинства пользователей, было бы неприемлемо). В совокупности с обработкой данных исключительно на устройстве пользователя это обеспечивает максимальный уровень анонимности и приватности.

3.3.3. Организация выходных данных

Базовая статистика (количество пользователей в экспорте, контактов, базовая информация о аккаунте) выводится в текстовом виде. Анализ отношений пользователей выводится на экран в виде интерактивного графа, отображающего пользователей и ассоциированные с ними чаты. В качестве дополнительной информации на графе по разному (тип и цвет объекта) отображаются разные элементы графа: пользователи, личные чаты, публичные и приватные группы, каналы. С каждым объектом указывается его уникальный Telegram ID.

3.3.4. Обоснование метода организации выходных данных

Базовая статистика - набор из нескольких чисел и имен, простое текстовое представление для нее ожидаемо, нет повода от него отказываться.

Представление связей между пользователями в виде интерактивного графа мотивированно доступностью пользователям: на маленьких примерах и статическая картинка, и интерактивный вариант, который можно подвигать или приблизить могут быть достаточно удобны, но на больших, запутанных схемах интерактивность значительно улучшает пользовательский опыт. Цветовая маркировка различных элементов улучшает читаемость графа, особенно в случаях конфликта имен (например, так проще отличить пользователя "Котенок" и канал с картинками с аналогичным названием). Указание Telegram ID решает случаи конфликтов, когда объек-

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ты одного типа имеют одинаковое имя (например, два разных пользователя, но оба с именем "Олег").

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. Технические и программные средства

4.1. Необходимые технические средства

Пупупу

4.2. Обоснование необходимых технических средств

Пупупу

4.3. Необходимые программные средства

Пупупу

4.4. Обоснование необходимых программных средств

Пупупу

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

5. Ожидаемые технико-экономические показатели

5.1. Экономическая эффективность

Расчёт экономической эффективности в рамках работы не предусмотрен.

5.2. Предполагаемая потребность

Пупупу

5.3. Преимущества разработки по сравнению с отечественными и зарубежными аналогами

Пупупу

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

6. Приложение 1. Терминология

Пупупу

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Список литературы

- [1] ГОСТ 19.101-77. ЕСПД. Виды программ и программных документов. — М.: ИПК Издательство стандартов, 2001.
- [2] ГОСТ 19.201-78. ЕСПД. Техническое задание. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [3] ГОСТ 19.301-79. ЕСПД. Программа и методика испытаний. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [4] ГОСТ 19.401-78. ЕСПД. Текст программы. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [5] ГОСТ 19.404-79. ЕСПД. Пояснительная записка. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [6] ГОСТ 19.505-79. ЕСПД. Руководство оператора. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [7] ГОСТ 19.504-79. ЕСПД. Руководство программиста. Требования к содержанию и оформлению. — М.: ИПК Издательство стандартов, 2001.
- [8] ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. — М.: ИПК Издательство стандартов, 2001.
- [9] ГОСТ 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды. — М.: Изд-во стандартов, 1997.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.05.04-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Лист регистрации изменений

[illegible]