

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

**ТЕКСТОВЫЙ АНАЛИЗ ПУБЛИКАЦИЙ  
TELEGRAM-КАНАЛОВ**

**АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ**

студента 2 курса 248 группы  
направления 09.04.03 — Прикладная информатика

механико-математического факультета  
Сохинова Михаила Денисовича

Научный руководитель

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Заведующий кафедрой

д. ф.-м. н., доцент

\_\_\_\_\_

С. П. Сидоров

Саратов 2023

## **Введение**

**Актуальность темы.** Исследования современных инструментов цифровой экономики и оценка их влияния на хозяйственную деятельность имеет важное значение, поскольку связаны с научной оценкой широкомасштабного использования данных, информации, цифровых ресурсов практически во всех сферах социальной активности общества. В современном мире социальные сети стали неотъемлемой частью жизни многих людей, и среди них особенно популярным мессенджером является Telegram. С помощью этой платформы пользователи могут создавать и подписываться на различные каналы, которые предлагают контент на различные темы - от новостей до юмора. Таким образом, Telegram стал площадкой для распространения информации и взаимодействия между людьми.

**Практическая значимость.** В связи с этим возникает важная задача анализа контента публикаций в Telegram каналах. Текстовый анализ является одним из инструментов, позволяющих автоматически обрабатывать и анализировать большие объемы текстовой информации. Он может быть использован для выявления тематик каналов, определения настроения публикаций, анализа популярности и многих других задач. Применение методов текстового анализа в социальных медиа, включая Telegram, может быть полезным для многих областей, таких как маркетинг, исследование общественного мнения, мониторинг новостей и т.д.

Одной из конкретных задач, которые можно решить с помощью текстового анализа в Telegram, является анализ публикаций в каналах. Каналы Telegram - это платформы для распространения информации на определенную тематику, которые могут иметь большую аудиторию. Анализ публикаций в таких каналах может помочь понять, что интересует и вовлекает аудиторию, и на основе этого оптимизировать контент для достижения максимальной эффективности.

**Объект исследования.** Для данной работы был выбран Telegram канал «Банк России», на который подписано более 130 тысяч пользователей, он имеет подтвержденный статус и является одним из официальных источников информации Центрального Банка РФ.

**Целью магистерской работы** является исследование методов тек-

стового анализа, применение одного из них для анализа публикаций в Telegram канале с целью определения тематик и популярности контента, а также выявления вовлеченности аудитории и выявления зависимостей новостей от экономической ситуации в стране. В работе необходимо решить следующие задачи:

- изучение Telegram, как социального инструмента
- проведение обзора исследований по теме применения текстового анализа к социальным сетям
- описание статистических методов анализа данных
- анализ бизнес-процессов
- изучение текстового анализа, как метода анализа данных
- разработка программы для сбора данных с Telegram канала
- изучение средств проведения текстового анализа
- написание программного кода для проведения текстового анализа
- анализ полученных результатов

Таким образом, данная работа будет содержать следующие разделы:

- обзор исследований, связанных с текстовым анализом социальных сетей
- описание методологии исследований
- описание текстового анализа
- разработка программы для сбора данных
- реализация текстового анализа

Кроме того, магистерская работа содержит введение, заключение, список использованных источников и 2 приложения. Список использованных источников содержит 20 наименований. В приложении А приведён код для реализации программы для сбора и обработки данных из Telegram-канала. В приложении Б приведён код для проведения текстового анализа Telegram канала.

**Работа прошла апробацию** на различных конференциях, в частности, на ежегодной студенческой конференции «Актуальные проблемы математики и механики», которую проводил механико-математический факультет СГУ в апреле 2022 года, в секции «Анализ данных», в XI Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2022 года.

## Основное содержание работы

Для реализации текстового анализа был выбран один из самых популярных мессенджеров России - Telegram. По данным компании App Annie, в первом квартале 2021 года Telegram был скачан более 5 млн раз в России на устройствах на базе Android и iOS, что составляет 21% от общего числа скачиваний мессенджеров в России. Также он является популярным мессенджером среди российских политиков, бизнесменов и журналистов, которые используют его для коммуникации и обмена информацией. На данный момент количеством активных пользователей за месяц в Telegram 48 млн., а в ВКонтакте 100 млн. Стоит отметить, что сеть ВКонтакте была создана на 7 лет раньше, чем мессенджер Telegram.

Уже несколько лет тема влияния социальных сетей и мессенджеров на жизнь людей является одной из самых обсуждаемых, как в СМИ, так и в научных организациях. Интернет постепенно забирается во все сферы жизнедеятельности человека и это не может не волновать. В исследовании роли социальных сетей в жизни людей и их вовлеченности в них можно выделить несколько направлений:

- Исследования мотивации к использованию социальных сетей. В рамках этого направления ученые занимаются поиском и апробацией концепций и теорий, способных объяснить причины участия молодых людей в социальных сетях (например, теория обоснованного или разумного поведения - theory of reasoned action).
- Социально-психологические исследования социальных сетей. Представители данного подхода в своих работах чаще всего касаются вопроса взаимосвязи между такими социально-психологическими характеристиками индивида, как коммуникабельность, конфликтность и использованием социальных сетей.
- Исследования особенностей виртуальной коммуникации. В рамках данного направления исследователи чаще всего касаются различных аспектов виртуальной репрезентации пользователей, особенностей восприятия других участников социальной сети при общении онлайн. Здесь наиболее часто применяются теория знаков или теория приписывания.
- Исследования социальных сетей «как таковых». В это направление по-

падает большая часть эмпирических исследований использования людьми социальных сетей: основное внимание, как правило, сосредотачивается на изучении основных практик использования социальных сетей, выявлении аудитории общения, анализу связи поведения в социальных сетях с социально-демографическими характеристиками пользователей.

Можно сказать, что исследователи пытаются ответить на три главных вопроса: зачем люди идут в виртуальные социальные сети, что они там делают и как это отражается на их реальной социальной жизни.

Деятельность людей во множестве случаев предполагает работу с данными, а она в свою очередь может подразумевать не только оперирование ими, но и их изучение, обработку и анализ. Например, когда нужно уплотнить информацию, найти какие-то взаимосвязи или определить структуры. И как раз для аналитики в этом случае очень удобно пользоваться не только разными техниками мышления, но и применять статистические методы.

Особенностью методов статистического анализа является их комплексность, обусловленная многообразием форм статистических закономерностей, а также сложностью процесса статистических исследований.

В современном мире нельзя не сказать, про новый показатель меры - это показатель вовлеченности. Именно он служит оценкой для пользователей социальных сетей. Какое-то время эффективность использования социальных сетей для бизнеса определялась по числу подписчиков. В этих и других абсолютных показателях выражалась «мощность» аккаунта. Чем больше пользователей подписаны на обновления сообщества или канала, тем лучше. Но в определённый момент был задан вопрос: «А в чём польза от огромной базы подписчиков?»

Перед владельцами групп и бизнес-аккаунтов встаёт вопрос о приведении подписчиков в активное состояние — они должны приносить пользу сообществу, а не лежать мёртвым грузом. Польза — это интерес, внимание или, говоря точнее, вовлечённость.

Engagement Rate, или ER, в SMM — это показатель вовлеченности: метрика продвижения, которая отражает количество пользовательских реакций на контент. Измеряют ее в процентах: при 0% на пост никто не отреагировал,

при 100% каждый подписчик поставил лайк или написал комментарий.

Коэффициент вовлеченности аудитории напрямую влияет на охваты. В 2022 году в большинстве соцсетей работают умные новостные ленты. Они подстраиваются под интересы пользователей и показывают то, на что люди отреагируют с большей долей вероятности. Алгоритмы новостных лент учитывают поведение подписчиков в сети и показатели вовлеченности вашего контента.

Показатель вовлеченности в социальных сетях для разных временных промежутков считают, чтобы тестировать гипотезы. Например, публиковать одну неделю посты с утра, другую – вечером, и сравнивать, какие набирают больше реакций. Или выкладывать в один день только подборки товаров, в другой – обзоры подписчиков на них и смотреть, в какой день средний ER больше.

Показатель вовлеченности в Telegram можно определить по нескольким параметрам:

- Количество просмотров сообщений. Если сообщение получает много просмотров, то это говорит о высокой вовлеченности аудитории.
- Количество лайков и комментариев. Если сообщение получает много лайков и комментариев, то это тоже свидетельствует о высокой вовлеченности.
- Скорость реакции на сообщения. Если пользователи быстро отвечают на сообщения и активно общаются в чатах, то это также свидетельствует о высокой вовлеченности.
- Количество подписчиков. Если у канала или чата много подписчиков, то это может свидетельствовать о высокой вовлеченности аудитории.

Общий показатель вовлеченности в Telegram можно вычислить, например, по формуле: среднее количество просмотров/лайков/комментариев на сообщение, поделенное на количество подписчиков канала или чата.

Однако, следует учитывать, что показатели вовлеченности могут различаться в зависимости от тематики канала или чата, а также от активности администраторов в поддержании коммуникации и взаимодействии с аудиторией.

Социальные сети и мессенджеры - это ещё и базы данных, в которых

огромное количество информации. Иногда для анализа канала нужно выгрузить базу, а специальных встроенных функций для этого не предоставляется. Для этой задачи можно использовать такой инструмент, как парсер.

Парсинг — автоматизированный сбор и систематизация информации из открытых источников с помощью скриптов. Другое название этого процесса — веб-скрейпинг.

Скрипты, которые собирают и систематизируют информацию, называются парсерами. Они работают так:

- ищут источники по заданным параметрам — например, парсеру можно дать список сайтов, а он найдёт на них страницы с ценами;
- извлекают из источников нужную информацию — несколько строк текста, ссылку или сумму;
- преобразуют информацию — например, парсер может взять фрагмент из HTML-документа и преобразовать его в текст без кода;
- сохраняют информацию в нужном формате — например, в виде списка или таблицы в Excel.

Парсеры работают на разных языках программирования — Python, JavaScript, PHP 5 и других.

Смысл парсинга в том, чтобы ускорить рутинную работу. Чтобы собрать и сохранить в таблице список из тысячи статей на сайте, человек потратит часы. Парсер сделает эту работу за несколько минут. Парсер ускоряет работу в сотни раз и допускает меньше ошибок, чем человек.

Парсинг страниц, ведущий сбор информации из веб-контента сайтов, используется в различных сферах — для продаж, маркетинга, финансов, электронной коммерции, сбора информации по конкурентам и так далее.

В данной работе будет использована только часть парсинга, а именно обработка данных и конвертация их в другой формат. Сама программа для сбора и обработки данных будет мною разработана на языке Python версии 3.8. Python — это язык программирования, который широко используется в интернет-приложениях, разработке программного обеспечения, науке о данных и машинном обучении.

Также для реализации программы нам потребуется социальный мессенджер Telegram, из которого мы будем вытаскивать историю чата или канала,

а затем уже эти данные конвертировать в нужный для нас формат и с нужной информацией.

В качестве среды разработки выбрана оболочка PyCharm, так как она удобна в использовании и имеет бесплатную версию.

Стоит отметить, что для реализации программного кода, сначала необходимо выгрузить историю чата или канала в формате json непосредственно из Telegram. Это делается с помощью десктопной версии приложения. Файл с историей чатов необходимо сохранить в ту же папку, где находится сама программа, то есть файл с расширением .py. В самой программе указывается название данного файла, чтобы можно было к нему обратиться и взять оттуда информацию. После запуска программы, в папке создастся файл формата csv, в котором и будет результат сбора данных по конкретному каналу или чату.

Полный код программы изложен в приложении А. Ниже показаны результаты работы программы по сбору данных из Telegram канала «Банк России».



```
15     if "chats" not in jdata:
16         parse_telegram_to_csv(jdata)
17     else:
18         for chat in jdata["chats"]["list"]:
19             parse_telegram_to_csv(chat)
```

in: parser

C:\Users\sohin\Anaconda3\envs\pythonProject1\python.exe "C:/Users/sohin/"  
Чат БанкРоссии был успешно спарсен

Process finished with exit code 0

Рисунок 1 - Успешное выполнение программы

Имя	Дата изменения	Тип	Размер
БанкРоссии.csv	20.12.2022 9:19	Файл Microsoft E...	809 КБ
result.json	20.12.2022 9:06	JSON File	2 125 КБ
parser.py	19.12.2022 20:17	Python File	5 КБ
video_files	20.12.2022 9:18	Папка с файлами	
stickers	20.12.2022 9:18	Папка с файлами	

Рисунок 2 - Папка с результатами



A	B	C	D	E	F	G
id сообщения	отправитель	id отправителя	ответы на сообщения	дата	тип сообщения	контент сообщения
3	Банк России	channel1680634724	-1	25.02.2022 11:19	link	Банк России принимает все необходимые м
4	Банк России	channel1680634724	-1	25.02.2022 11:20	text	! Что важно знать о банках, в отношении ко
5	Банк России	channel1680634724	-1	25.02.2022 11:32	text	Текущий уровень ликвидности российской ко
7	Банк России	channel1680634724	-1	25.02.2022 14:37	text	📄 Карты всех платежных систем в России г
8	Банк России	channel1680634724	-1	25.02.2022 17:06	link	👉 Приняты новые меры по поддержке фина
9	Банк России	channel1680634724	-1	25.02.2022 18:54	text	Банк России проведет 28 февраля 2022 года
10	Банк России	channel1680634724	-1	25.02.2022 19:11	link	Условия заключения сделок «валютный св
11	Банк России	channel1680634724	-1	25.02.2022 20:45	text	Мы включили на своём сайте дополнитель
12	Банк России	channel1680634724	-1	25.02.2022 21:06	text	Увеличили выдачу наличных денег банкам
13	Банк России	channel1680634724	-1	25.02.2022 23:08	text	Четыре человека, поставившие 🗳, вы этим
14	Банк России	channel1680634724	-1	25.02.2022 23:12	poll	82676
15	Банк России	channel1680634724	-1	25.02.2022 23:31	text	Жаль, очень много ответов «нет», и очень
16	Банк России	channel1680634724	-1	27.02.2022 12:37	text	👉 Банк России располагает необходимыми
17	Банк России	channel1680634724	-1	27.02.2022 15:08	text	📄 Небольшое уточнение по работе pay-se
18	Банк России	channel1680634724	-1	27.02.2022 17:06	link	📄 В дополнение к ранее принятым решен
19	Банк России	channel1680634724	-1	27.02.2022 20:21	text	👉 Банк России с 28 февраля 2022 года возо
20	Банк России	channel1680634724	-1	28.02.2022 0:12	text	Нам сообщают, что на нашем канале появи
21	Банк России	channel1680634724	-1	28.02.2022 9:03	text	Торги на валютном рынке, денежном рынк
22	Банк России	channel1680634724	-1	28.02.2022 9:12	link	Дополнительные меры для обеспечения у
23	Банк России	channel1680634724	-1	28.02.2022 10:47	text	👉 Совет директоров Банка России принял р
24	Банк России	channel1680634724	-1	28.02.2022 10:50	link	👉 Заявление Председателя Банка России п
25	Банк России	channel1680634724	-1	28.02.2022 10:55	text	👉 На основании предложений Министерств
26	Банк России	channel1680634724	-1	28.02.2022 11:43	link	Расширяем объем обеспечения кредитны
27	Банк России	channel1680634724	-1	28.02.2022 13:25	text	! Важно: ранее принятые решения о време
28	Банк России	channel1680634724	-1	28.02.2022 14:04	text	👉 В связи со складывающейся ситуацией мы
29	Банк России	channel1680634724	-1	28.02.2022 16:58	link	Трансляция заявления на сайте – www.cbr.r
30	Банк России	channel1680634724	-1	28.02.2022 19:04	text	! Сегодня торги на Московской бирже зако
31	Банк России	channel1680634724	-1	28.02.2022 20:09	text	Как регулятор мы не вправе давать совет
32	Банк России	channel1680634724	-1	28.02.2022 20:09	text	Обращение инвестиционного сообщества !
33	Банк России	channel1680634724	-1	28.02.2022 20:10	text	Но долгосрочные запреты могут создать ха
34	Банк России	channel1680634724	-1	28.02.2022 22:32	text	📄 В связи со складывающейся ситуацией Б
35	Банк России	channel1680634724	-1	01.03.2022 15:52	text	! Важно! Люди по всей стране получают о
36	Банк России	channel1680634724	-1	01.03.2022 16:09	text	!! До 1 трлн руб. из ФНБ правительство выд
37	Банк России	channel1680634724	-1	01.03.2022 18:27	link	📄 Напоминаем, что на нашем официально
38	Банк России	channel1680634724	-1	02.03.2022 9:01	text	👉 Банк России принял решение не возобно

Рисунок 3 - Результат работы в формате csv

После того, как с помощью разработанной программы был получен файл с необходимыми нам данными, мы можем приступить к анализу. Для начала нужно разобраться, какие методы анализа существуют и какой более эффективен для решения поставленной задачи.

Сейчас активно развивается такой метод исследования, как контент-анализ, который используется для анализа содержания различных видов коммуникаций, таких как тексты, изображения, аудио- и видеозаписи.

Этот метод используется для выявления определенных тенденций, тем, образцов поведения, отношений и других ключевых характеристик в текстовых и других формах коммуникации.

Контент-анализ является методом исследования, который используется уже более полувека. В начале своего развития он применялся в области журналистики и исследований рекламы, однако с течением времени стал широко использоваться в различных областях. Современные методы контент-анализа развиваются в соответствии с требованиями рынка и развития технологий. В частности, расширяется область применения контент-анализа за счет возможностей анализа больших объемов данных и автоматизации процесса ис-

следования. Сейчас в развитие контент-анализа активно вкладываются также исследователи из различных областей, включая информационные науки, компьютерные технологии, социологию и другие. Они работают над улучшением методов анализа текста, развитием алгоритмов и технологий, которые позволяют более эффективно проводить контент-анализ и получать более точные результаты.

В процессе контент-анализа, исследователь применяет систематический подход для категоризации и кодирования содержания, используя заранее определенные критерии и категории. Затем проводится статистический анализ полученных данных, который позволяет получить количественные результаты и вывести выводы о содержании и характеристиках исследуемых материалов. Контент-анализ широко используется в различных областях, включая научные исследования, маркетинг, политику, социологию, психологию и другие области, где требуется анализировать большие объемы текстовых или других видов информации.

Так как, контент-анализ это метод, который в основном применяется при различных типах данных, при обработке изображений, аудио и видео материалов, в данной работе уместней будет использовать текстовый анализ.

Текстовый анализ - это процесс извлечения и анализа информации из текстовых данных с использованием методов компьютерной лингвистики и машинного обучения. Он может быть применен к различным типам текстов, таким как социальные медиа сообщения, новостные статьи, электронные письма, научные публикации и т.д.

Цель текстового анализа - понимание смысла и содержания текста, выявление тем, настроений, мнений и других параметров. Для этого используются различные методы анализа, такие как статистический анализ, морфологический анализ, синтаксический анализ, семантический анализ и др.

Текстовый анализ может использоваться для решения различных задач, таких как:

- автоматическое распознавание языка и определение его структуры;
- категоризация и классификация текстов;
- выявление тематик, настроений и мнений в текстах;
- автоматический анализ текстовых ответов в опросах и анкетах;

- автоматический поиск ключевых слов и фраз в текстах;
- автоматический перевод текстов на другой язык.

В зависимости от конкретной задачи и используемых методов, текстовый анализ может требовать различных инструментов и технологий, таких как языковые модели, алгоритмы классификации и кластеризации, методы искусственного интеллекта и машинного обучения и другие.

TF и IDF (TF-IDF) - это два показателя, используемых в текстовом анализе для оценки важности слов в документе или коллекции документов.

TF (term frequency) - это показатель, который показывает, как часто определенное слово встречается в документе. Вычисляется путем деления количества вхождений слова в документ на общее количество слов в документе. Чем выше TF, тем более важное слово в конкретном документе.

IDF (inverse document frequency) - это показатель, который показывает, насколько часто определенное слово встречается во всей коллекции документов. IDF вычисляется путем логарифмирования отношения общего числа документов в коллекции к числу документов, содержащих это слово. Чем выше IDF, тем более редкое и важное слово в коллекции документов.

TF-IDF - это произведение двух показателей, которое позволяет оценить важность слова в конкретном документе и всей коллекции документов. Чем выше значение TF-IDF для слова, тем более важное оно в конкретном документе и всей коллекции документов.

TF-IDF анализ используется для решения различных задач, таких как категоризация и классификация текстов, поиск ключевых слов и фраз, выявление тематик и другие.

Существует множество средств для проведения текстового анализа, некоторые из них:

1. Python библиотеки: Python имеет несколько библиотек для обработки естественного языка, таких как NLTK, SpaCy, TextBlob, Gensim и многие другие.
2. Коммерческие пакеты: Существует множество коммерческих пакетов, таких как RapidMiner, KNIME, IBM Watson, SAS, которые предоставляют инструменты для анализа текста.
3. Онлайн-инструменты: Существует множество онлайн-инструментов, та-

ких как Google Cloud Natural Language API, Amazon Comprehend, IBM Watson Natural Language Understanding, которые предоставляют возможности для анализа текста.

4. Открытые исходные коды: Существуют открытые исходные коды, такие как BERTopic, которые предоставляют возможности для тематического моделирования текста.

В качестве инструмента в данной работе было решено остановиться на открытом исходном коде - BERTopic.

BERTopic - это пакет Python для тематического моделирования текста на основе алгоритма BERT (Bidirectional Encoder Representations from Transformers), который является одним из наиболее продвинутых алгоритмов обработки естественного языка. BERTopic позволяет проводить анализ текста и выявлять темы, на которые он может быть разбит. BERTopic имеет следующие преимущества:

- Использование предварительно обученной модели BERT позволяет получать высокую точность в выделении тематических кластеров и понимании смысла текста.
- Автоматически определяет оптимальное количество тем в тексте, что облегчает процесс анализа текста.
- Позволяет создавать качественные визуализации тематической структуры, что облегчает понимание текста и его интерпретацию.
- Обладает высокой скоростью обработки текста и может обрабатывать большие объемы данных.
- Позволяет выявлять доминирующие слова и ключевые фразы в каждой теме, что упрощает процесс интерпретации результатов.
- Позволяет проводить анализ текста как с использованием всего корпуса текстов, так и с использованием отдельных документов, что делает его более гибким и универсальным инструментом для анализа текста.

Сам программный код писался и запускался на сервисе Kaggle, так как необходимо было подключать сразу несколько библиотек. Ниже представлено несколько графических визуализаций проведенного текстового анализа. Полные результаты отражены в самой работе, а программный код вынесен в приложение Б.

[16]:

	Topic	Count	Name
0	-1	207	-1_россии_банк_года_также
1	0	55	0_цен_инфляция_году_инфляции
2	1	52	1_денежно_ставки_ставок_года
3	2	49	2_кредитов_кредитам_кредитных_банк
4	3	41	3_сша_валюте_долларов_средства
5	4	31	4_банка_россии_председателя_состоится
6	5	28	5_мск_московской_режиме_торги
7	6	23	6_расписок_бумаг_бумаги_акции
8	7	21	7_осаго_страхования_страховщики_страховой
9	8	20	8_мошенники_злоумышленники_данные_мошенников
10	9	18	9_финансовой_финансового_рынка_стабильности
11	10	17	10_фондов_рублей_сбп_доходность
12	11	17	11_банки_россии_вклада_банк
13	12	14	12_грамотности_финансовой_онлайн_сессия
14	13	13	13_аукцион_проведет_репо_мая
15	14	13	14_правила_инвесторов_стандарт_защиты
16	15	12	15_климата_предприятий_индикатор_бизнес
17	16	12	16_подробности_интервью_трансляция_
18	17	11	17_ожидания_инфляционные_ценовые_настроения
19	18	11	18_вебинар_предпринимателей_желающие_задать

Рисунок 4 - Самые популярные темы

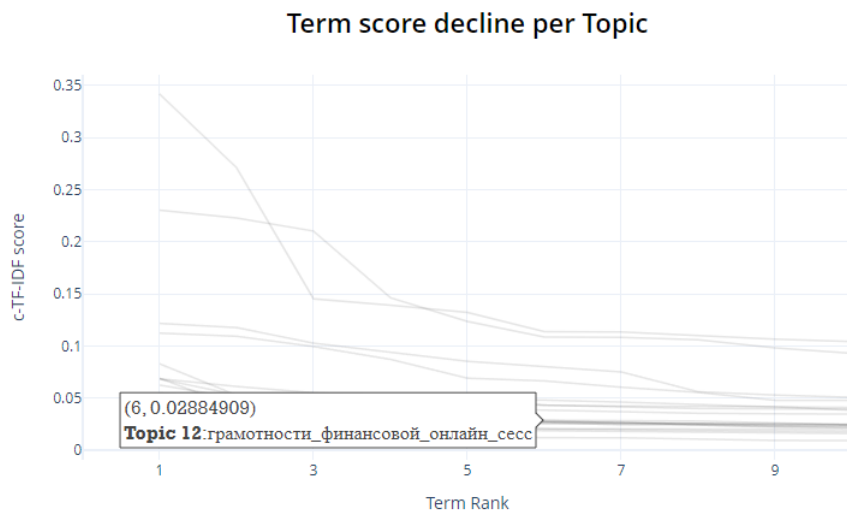


Рисунок 5 - График оценки c-TF-IDF



Рисунок 6 - Карта межтематических расстояний

## Заключение

В ходе работы было проведено исследование, которое позволило выявить наиболее популярные темы, обсуждаемые в Telegram канале «Банк России», а также определить взаимосвязь тематики публикаций с состоянием экономики в стране. Для проведения этого исследования были выполнены следующие задачи:

- рассмотрение Telegram, как инструмента социальной активности
- проведение обзора проведенных исследований по теме текстового анализа
- рассмотрение нескольких видов статистических методов анализа данных
- изучение текстового анализа и его методов
- разработка программы для сбора и обработки сообщений из Telegram канала
- написание программного кода на языке Python для проведения текстового анализа
- обработка полученной информации и вывод зависимостей постов в Telegram канале от экономики страны

Тематический анализ был использован для выявления наиболее часто обсуждаемых тем в каналах. Было выявлено несколько основных тем, которые отражают вектор направления страны в сфере экономики и отношение аудитории к определенным направлениям развития Центрального Банка РФ. Эта информация может быть полезна для рекламодателей и маркетологов, которые могут использовать ее для создания более эффективных рекламных кампаний и увеличения вовлеченности аудитории к той или иной тематике.

В целом, результаты исследования подтверждают эффективность применения методов текстового анализа для анализа публикаций в Telegram каналах и могут быть полезны для маркетологов, рекламодателей и исследователей социальных медиа.

Однако, для более точных и полных результатов анализа необходимо учитывать различные факторы, такие как специфика аудитории, региональные особенности и конкуренцию в отрасли.

В дальнейшем, можно проводить более глубокий анализ содержания

публикаций, использовать более сложные методы машинного обучения и анализировать другие социальные медиа.

Результаты работы имеют практическую ценность, как для государственных учреждений, так и для обычных пользователей. Данную модель проведения текстового анализа при работе с социальной сферой можно использовать в любых направлениях. Это может быть очень полезным инструментом для улучшения взаимодействия между организацией и пользователями, а также для получения более качественной информации из текстовых сообщений.