

Тестирование методов обработки комментариев из Telegram-каналов и пабликов ВКонтакте для анализа социальных медиа

Б.А. Низомутдинов, О.Г. Филатова

Аннотация — В работе представлен сравнительный анализ двух официальных источников информации, на примере группы в социальной сети ВКонтакте и Telegram-канала главы Иркутской области. Объект исследования - комментарии, которые публикуются под официальными записями и постами. Используются автоматизированные методы сбора и обработки комментариев. Использовались API ВКонтакте и библиотека Tescrape для парсинга данных из Telegram. Функционал настольной версии мессенджера Telegram позволил выгрузить комментарии из любого открытого канала в формате JSON. Все полученные результаты были сохранены в BI систему Yandex Data Leans. Для собранной информации проведено тематическое моделирование, выделено 10 основных тем, которые обсуждаются в комментариях в ВКонтакте и Telegram. Для создания тематических моделей применялась модель Gensim библиотеки Python.

Сравнительный анализ выделенных тем показал, что в комментариях ВКонтакте и в Telegram заметны разные социально-экономические проблемы, поднимаемые жителями региона и они по-разному ранжируются. Используемые методы и подходы в целом показали эффективность как аналитическое средство для сбора и оценки комментариев из разных типов социальных медиа, позволяющее, в частности, выявлять их делиберативный потенциал.

Ключевые слова — Telegram, ВКонтакте, парсинг, тематическое моделирование, социальные медиа, делиберация

I. ВВЕДЕНИЕ

В настоящее время социальные сети и мессенджеры являются важными источниками информации для проведения исследований на различном общественно-политическом тематикам. С развитием технологий появились новые возможности по сбору и обработке той информации, которую публикуют пользователи в различных каналах, позволяющие без привлечения дополнительных средств проводить такие исследования.

За последнее десятилетие хорошо изучены возможности автоматизированных методов,

применяемых для изучения коммуникации в социальной сети ВКонтакте. Однако большая часть новейшей цифровой среды - это мобильные службы обмена сообщениями на базе смартфонов - платформы и приложения. Это меняет сетевой ландшафт, поскольку «мобильные приложения для обмена сообщениями и агрегаторы новостей становятся все более важными для людей» [1]. Происходит переход к распределенной медиасреде, которая меняет конфигурацию каналов связи и открывает новые возможности для исследователей из-за быстрого роста популярности групп в приложениях для обмена сообщениями [2]. В частности, в последние годы все большее внимание исследователей привлекает мессенджер Telegram как площадка для проведения исследований. В данной работе предпринимается попытка объединить возможности выгрузки комментариев из ВКонтакте и Telegram для проведения в дальнейшем сравнительного анализа таких комментариев.

Представленное в данной статье исследование проведено в продолжение исследования, описанного в предыдущей статье [3], где с помощью инструмента TGStat был осуществлен пилотный анализ каналов глав всех регионов РФ в Telegram: отобраны каналы, собрана статистическая информация, выгружены сообщения и проведен предварительный анализ текстовой информации. В результате кластерного анализа сообщений были выделены тематические кластеры, описанные ключевыми словами, позволяющими оценить основные темы, которые освещаются в аккаунтах официальных лиц регионов. Опробованный инструмент показал свою перспективность и в качестве продолжения данной исследовательской линии начался анализ коммуникационных каналов глав отдельных регионов, причем не только официальных каналов в Telegram, но и публичных страниц ВКонтакте. В данном случае для исследования были выбраны каналы публичной коммуникации главы Иркутской области. В дальнейшем этот анализ позволит, на наш взгляд, выявлять и сравнивать делиберативный потенциал разных социальных медиа для взаимодействия с общественностью главы любого другого региона.

II. ОБЗОР ИССЛЕДОВАНИЙ

Исследования делиберативного потенциала

Статья получена 20 апреля 2023. Исследование выполнено за счет гранта Российского научного фонда № 22-18-00364 «Институциональная трансформация управления электронным участием в России: исследование региональной специфики» (<https://rscf.ru/project/22-18-00364/>)

Б.А. Низомутдинов, Университет ИТМО (boris@itmo.ru)
О.Г. Филатова, Санкт-Петербургский государственный университет, Университет ИТМО (o.filatova@spbu.ru)

социальных медиа ведутся уже более 10 лет. Одними из первых таких работ, например, являются работы Jennifer Stromer-Galley, которая предложила систематический способ измерения того, что происходит в ходе обсуждений [4;5]. В 2015 году J. Stromer-Galley вместе с коллегами сделали шаг вперед и предложили классификацию типов выражения разногласия, используя методику дискурс-анализа. Были проанализированы качественные различия в формах выражения разногласий в офлайн и онлайн-делиберации, а также насколько сильно поддерживается несогласие в онлайн-среде [6]. Современные исследования свидетельствуют, что социальные медиа значительно демократизируют и упрощают процессы коммуникации, предоставляя возможность получения моментальной обратной связи [7]. Социальные медиа позволяют построить личные и более прочные взаимоотношения с сотнями и тысячами людей, узнать, что люди думают о власти на самом деле, разговаривая в неформальной обстановке [8]. Характер передаваемой информации – пожалуй, главное, определяющее свойство социальных медиа. Исследователи описывают его как «подлинность», «искренность», «эмоциональность». Полученные многочисленными зарубежными исследованиями данные показывают, что интернет расширяет репертуар коллективных действий, а цифровые инструменты социальных медиа во многих отношениях способствуют основанной на участии делиберативной демократии. Демократическая делиберация означает многообразие и доступность взглядов и интересов, что особенно важно в сложных и плюралистических обществах, отмеченных глубокими различиями и разногласиями [9].

В настоящее время общепризнанным является тот факт, что правительства должны стремиться к сотрудничеству с гражданами, чтобы принимать сбалансированные коллективные решения и совместно разделять ответственность за них. В частности, была установлена взаимосвязь между предоставлением качественной информации в социальных сетях и доверием граждан к правительственной структуре [10]. Исследования также показывают, что в равной степени людей интересует реакция на их мнения других граждан, а не только власти. Очень часто именно дискуссии среди самих граждан позволяют выявить все многообразие и глубину коллективного мнения [11]. Поэтому мы не можем обойти стороной исследования комментариев и дискуссий, которые ведут в интернете простые граждане, обсуждая разные актуальные социально-политические темы.

К настоящему времени ученые проанализировали, с интеграцией в исследования автоматизированных и машинных методов обучения, является ли комментирование постов в социальных сетях открытым и видимым для всех, как организованы интернет-дискуссии с точки зрения содержания и стиля, как такие дискуссии влияют на участников, а также на лиц, принимающих решения [12; 13]. Однако технической стороне вопроса, касающейся возможности выгрузки комментариев, посвящено не так уж много работ.

Так, в настоящее время разрабатываются методы для диагностики восприятия городского пространства пользователями социальных сетей и возможность использования этих данных для формирования программ реорганизации инфраструктуры городской среды [14]. Оцениваются значимые корреляции между ценностно-смысловыми характеристиками личности пользователей и данными их профилей из социальных сетей для прогнозного и аналитического управления городским пространством. Проводится анализ комментариев пользователей социальных сетей для оценки уровня социальной напряженности. Разработаны подходы к измерению социальной напряженности в отдельных регионах России на основе анализа сообщений пользователей в социальной сети ВКонтакте (VK) [15].

Изучаются коммуникативные тактики и лингвистические приемы авторов популярных политических Telegram-каналов [16], анализируется их воздействия на читателя и определяется их роль в формировании общественного мнения.

Рассматривается влияние Telegram на информационную повестку дня в России [17-18] и делается вывод о том, что возрастает влияние информационного наполнения мессенджера Telegram на информационную повестку дня в России. На конкретных примерах исследуются причины популярности Telegram и его специфика, прогнозируются перспективы дальнейшего развития данного мессенджера в качестве нового ресурса и инструмента информационного обеспечения внутренней и внешней политики, цифровой дипломатии.

Предпринимаются попытки изучения политических слухов в мессенджерах [19], ведутся исследования в области возможностей мессенджера Telegram как инструмента формирования имиджа политического лидера, в ходе чего выявляются приемы и методы размещения информации в социальной сети, которые показывают высокую эффективность в работе по связям с общественностью в политике [20].

Поэтому мы полагаем, что изучение дискурсивных практик комментирования и обсуждения в мессенджере Telegram и социальной сети ВКонтакте является важным компонентом изучения как делиберативного потенциала цифрового пространства в целом, так и отдельных социальных медиа, в частности,

III. ДИЗАЙН ИССЛЕДОВАНИЯ

Важно понимать, что ВКонтакте и Telegram отличаются друг от друга по своей сущности. Это два инструмента, которые могут решать как разные, так и одинаковые задачи.

Telegram – это мессенджер, используемый для обмена сообщениями и файлами между пользователями. Он позволяет создавать чаты, группы и каналы для общения, передавать файлы до 2 Гб в одной передаче, всегда работает в режиме реального времени. Telegram имеет интеллектуальную систему безопасности, которая предоставляет конфиденциальность и защиту данных пользователей. Telegram за 2022 год в два раза нарастил

количество подписчиков русскоязычных каналов – с 1 до 2 млрд., по данным TGStat, в два раза выросло и число суточной аудитории в России. На конец марта 2023 года каналы в Telegram ведут главы 80 регионов России.

ВКонтакте (VK) – это социальная сеть, которая также позволяет общаться с другими пользователями, но основным функционалом является обмен контентом, таким как фотографии, видео, аудио и текстовые сообщения. ВКонтакте также имеет функцию создания групп и форумов. Сообщения в ВКонтакте имеют расширенные функционалы. По данным We are social в России ВКонтакте пользуются 75,3 % пользователей (Telegram – 64,4%). Главы только двух регионов в России не имеют личных страниц «ВКонтакте».

В нашем исследовании мы рассмотрим функции, которые есть в обоих инструментах – возможность публиковать новости и оставлять комментарии под ними, дают такую возможность. С разной реализацией, но пользователи ВКонтакте и Telegram могут оставлять записи в своих каналах, а другие пользователи могут их комментировать, если такая опция не отключена.

Для дальнейшего описания важно отметить, что в качестве источников информации для исследования используются два коммуникативных канала – сообщества ВКонтакте и публичные каналы в Telegram.

Каналы в Telegram – это вид публичных чатов, которые могут быть созданы любым пользователем или группой пользователей. Владелец канала может публиковать любой контент: текст, фотографии, видео, аудиозаписи, файлы и т.д. Пользователи могут подписаться на канал и получать уведомления о новых публикациях. Как правило, каналы используются для распространения новостей, информации о товарах или услугах, рекламы, а также для связи с аудиторией и создания сообществ. Кроме того, при наличии включенной опции, в каналах можно комментировать записи.

Сообщество ВКонтакте – это группа людей, объединенных вокруг общей тематики (например, музыкальное сообщество, группа фитнеса, туристическое сообщество и т.д.), которые могут обмениваться информацией, обсуждать темы и создавать события. В сообществе пользователя могут опубликовывать статьи, фото и видео, задавать вопросы и комментировать, если такая возможность не отключена администратором.

В данной работе важно, что оба эти инструмента дают возможность оставлять комментарии под сообщениями. Мы предприняли попытку проанализировать, как строится диалог в двух разных источниках: под публикациями ВКонтакте и под публикациями в Telegram. В дальнейшем это позволит проанализировать коммуникативные практики пользователей, выявить отличия в практике ведения диалога на двух разных площадках и есть ли отличия в уровне доверия информации.

Для тестирования методов в нашем исследовании мы выбрали канал в Telegram и группу ВКонтакте одного главы региона – губернатора Иркутской области Игоря Ивановича Кобзева. Мы исходили из универсальности

нашей методики и возможности ее применения в дальнейшем к любому объекту исследования.

Подчеркнем, что в ходе исследования анализировался не контент самих записей, а комментарии, оставленные под постами губернатора. Необходимо отметить, что среди глав российских регионов, которые имеют свои каналы в Telegram, только 41% предоставляют возможность комментирования своих постов, в 59% посты закрыты для комментирования. Среди открытых для комментирования каналов Telegram-канал главы Иркутской области занимает четвертое место по количеству подписчиков (рис. 1). Следует также отметить, что из подборки постов, собранных для количественного анализа, был исключен канал главы Чеченской республики https://t.me/RKadyrov_95, так как он имеет максимальное число подписчиков (2,6 млн), а столь высокий показатель не позволяет построить графики для сопоставления позиций других каналов, чьи значения не достигают даже 100 тыс. подписчиков. Лидирующие позиции Telegram-канала главы Чеченской республики отмечались во многих публикациях и рейтингах.



Рисунок 1. Количество подписчиков в публичных каналах глав регионов с возможностью комментирования

В целом Сибирский федеральный округ (СФО), куда входит Иркутская область, является одним из наименее развитых в цифровой сфере округов России [3]. При этом Иркутская область (наряду с Красноярским краем и Новосибирской областью) имеет наибольший потенциал для развития электронного взаимодействия власти и общества в СФО. В Иркутской области самый высокий индекс цифровой среды [21] и очень заметна личность губернатора, который имеет стабильный

медиаиндекс [22]. Процент жителей, подписанных на госпаблики, невысок, однако аудитория является одной из самых активных по федеральному округу – можно предположить, что данный показатель связан с рейтингом губернатора и говорит о степени доверия к власти.

IV. СБОР ДАННЫХ

Для решения задачи использованы автоматизированные методы выгрузки информации, а также использованы ранее разработанные инструменты для парсинга данных.

Для сбора информации из социальной сети задействовано API ВКонтakte. API (Application Programming Interface) ВКонтakte предоставляет возможность разработчикам использовать функционал социальной сети для создания своих приложений.

API ВКонтakte представляет собой набор методов, которые разработчик может вызывать из своей программы для взаимодействия социальной сети, таких как получение информации о профилях пользователей, доступ к новостной ленте и др. Для использования API ВКонтakte разработчик сначала должен получить ключ доступа (access token). API VK позволяет получать и передавать множество данных, включая информацию о пользователях и сообществах, подписках и т.д. API также позволяет выполнять различные операции с данными, например, создавать, редактировать и удалять записи пользователей и сообществ, добавлять и удалять пользователей из списков друзей и многое другое.

Для разработчиков, помимо других функций, возможна выгрузка комментариев из открытых групп. Что и было проделано в данной работе. Единственная доработка - все данные были деперсонализированы. В настоящей работе сохранялся только текст комментариев, без персональных данных.

Более сложной задачей стал парсинг комментариев из мессенджера Telegram, так как обзор литературы не выявил доступных методов для этого. Поэтому мы обратились к репозиторию GitHub для поиска решений. В итоге была отобрана библиотека Telescape. Telescape - библиотека на языке Python, которая используется для парсинга данных из Telegram. Она позволяет автоматизировать процесс сбора данных из чатов, каналов и групп в Telegram с помощью скриптов на Python.

Telescape предоставляет API для работы с Telegram и позволяет получать информацию о сообщениях, пользователях, чатах и других объектах Telegram. Кроме того, библиотека Telescape позволяет извлекать данные из структурированных сообщений (например, таблиц, списков и пр.) для еще более глубокого анализа данных. Telescape может использоваться для многих задач, таких как мониторинг публикаций в Telegram-каналах, анализ сообщений в Telegram-группах или для отслеживания активности конкурентов в Telegram. Он является эффективным инструментом для сбора данных, особенно в контексте анализа социальных медиа.

Однако библиотека Telescape не обновлялась последние 2 года и часть функций в настоящее время не

поддерживается. В итоге, решение было найдено в самом мессенджере – функционал настольной версии приложения дает выгрузить комментарии из любого открытого канала в формате JSON.

Мы провели выгрузку комментариев ко всем постам в группе ВКонтakte и выгрузили все комментарии из канала в Telegram. Итоговая база сохранена в формате JSON.

В таблице 1 приведены итоговые данные по сбору информации, все полученные результаты были сохранены в BI систему Yandex Data Leans. Yandex Data Lens - это инструмент, представленный компанией Яндекс, является бесплатным сервисом для визуализации и анализа больших данных. Он позволяет быстро и удобно обрабатывать данные различных форматов, а затем визуализировать их в виде графиков, диаграмм, карт и других элементов.

Было определено общее количество записей, комментариев и подписчиков. Данные приведены в таблице 1.

Таблица 1. Количественные показатели собранной информации в группе ВКонтakte и канале Telegram

	ВКонтakte	Telegram
Количество подписчиков	20310	25263
Количество записей	3535	2744
Дата регистрации	30.12.2019	25.02.2022
Собрано комментариев	22813	48454

V. ОБРАБОТКА ДАННЫХ

Далее, для более качественного анализа были использованы методы на основе машинного обучения. Было поставлено 2 задачи (см. рис. 2) - провести тематическое моделирование для выявления основных тем, на которые оставляют комментарии пользователи (метод кластеризации) и классифицировать записи по заранее размеченным темам (метод классификации).

Тематическое моделирование - это метод анализа данных, который позволяет выявлять темы или скрытые смыслы в большом объеме текстовых данных. Он используется для автоматической категоризации документов на основе сходства их содержания.



Рисунок 2. Общая схема сбора и обработки информации

Тематическое моделирование производит кластеризацию текстов по темам, вычленимым из

набора слов или фраз, которые часто встречаются в этих текстах. Каждый кластер соответствует конкретной теме и может быть описан набором ключевых слов. Эти ключевые слова отражают содержание текстов в кластере.

Он упрощает анализ текстовых данных, позволяет выявлять скрытые связи и закономерности и помогает принимать обоснованные решения на основе данных.

В нашей работе мы использовали модель Gensim. Gensim - это библиотека для Python, которая используется для высокоуровневого решения задач по работе с естественным языком (Natural Language Processing, NLP). Одной из наиболее популярных и интересных функциональных возможностей библиотеки является возможность создания тематических моделей.

Тематическая модель Gensim - это модель, созданная на основе алгоритма LDA (Latent Dirichlet Allocation), который используется для поиска скрытых тем в большом количестве текстовых данных. Алгоритм LDA помогает выделить наиболее вероятные темы в текстовых коллекциях на основе вероятностной модели. В обработке естественного языка скрытое распределение Дирихле (LDA) представляет собой генеративную статистическую модель, которая объясняет набор наблюдений с помощью ненаблюдаемых групп, и каждая группа объясняет, почему некоторые части данных похожи. LDA является примером тематической модели. При этом наблюдения (например, слова) собираются в документы, и присутствие каждого слова относится к одной из тем документа. Каждый документ содержит небольшое количество тем.

В итоге, для каждого из собранных массивов были выявлены 10 тем, которые описываются словами, приведенными в таблице 2 и 3. Каждая тема в обобщенном виде сведена и проранжирована в порядке значимости в таблице 4.

Таблица 2. Выделенные слова тем для массива комментариев из социальной сети ВКонтакте

№ п.п.	Topics — Выделенные кластеры «Комментарии ВКонтакте»
1	иркутской, уточните, региона, ответ, вера, регулирования, сообщение, санкции, вопрос, газификации
2	товаров, рублей, детей, маски, новость, снт, ребёнок, украинцы, дети
3	ремонт, работы, иркутской, администрации, дороги, города, работ, дорог, района, территории
4	молодцы, иркутской, здравоохранения, помогать, правительство, праздником, записи, право, благодарю
5	бог, дай, здоровья, героям, труд, галина, молодец, внимание, жизни, другим
6	детей, социальной, дом, социального, режим, вопрос, бред, выплаты, дома
7	дороги, питания, кут, усть, магазинах, ангарского, автомобильной, производство, округа, городского
8	цен, первой, необходимости, сахар, рф, товары, хлеб, необоснованного, фас

9	память, светлая, вечная, иркутской, рф, суда, нашим, бумагу
10	нато, рождения, магазина, страны, вообще, kobzev, нашей, требуется

Таблица 3. Выделенные слова тем для массива комментариев из мессенджера Telegram

№ п.п.	Topics — Выделенные кластеры «Комментарии в Telegram»
1	номеру, рекомендуем, действий, транспорта, соответствии, благодарю, план, иркутска, дорожной
2	пункт, ждём, подскажите, ждать, вообще, обращения, звонила, банк, удачи, усолье
3	большое, срок, новосибирске, поддержку, утро, здоровья, вода, вопросом, успехов
4	федерации, работ, российской, мобилизации, вопросу, военнослужащего, рф, рублей, денежное, довольствие
5	работы, тел, ремонт, дороги, дорог, строительство, телефона, части, усть, школы
6	мо, военную, сво, месту, личные, бог, семье, военкомат, отправили, дай
7	детей, рождения, муж, дети, прошу, мужа, школы, сказали, школу, внимание
8	обороны, днём, обращение, ребята, центр, завтра, больницы, поздравляю, детской, график
9	военнослужащих, здоровья, воды, горячей, линии, связь, юрге, сил, граждане, желаю
10	информацию, иркутской, обратиться, сообщение, можете, ответ, ноября, министерства, военной, граждан

Таблица 4. Ранжирование тем для массива комментариев из ВКонтакте и из Telegram

Темы ВКонтакте	Место	Темы Telegram
Региональные экономические проблемы	1	Транспорт, дороги
Защита детей	2	Горячая линия
Ремонт дорог	3	Поздравления, благодарности
Поздравления, благодарности	4	Денежное довольствие военнослужащих, мобилизация
Социально-экономические проблемы, поддержка героев	5	Дороги (ремонт дорог)
Социальные проблемы	6	Вопросы СВО
Региональные экономические проблемы	7	Поддержка семьи и детства
Цены на товары первой необходимости	8	Обращения военнослужащих
Память о земляках	9	Горячая линия
Патриотизм	10	Обращения военнослужащих

Из таблиц следует, что для каждого коммуникационного канала можно выделить свой набор тем и к тому же эти темы по-разному ранжируются.

Сравнительный анализ наглядно показывает, что среди тем, поднимаемых в комментариях во ВКонтакте, преобладают социально-экономические проблемы, среди комментариев в Telegram больше вопросов, посвященных поддержке военнослужащих и проведению специальной военной операции. Заметно, что Telegram чаще используется именно для вопросов и обращений, требующих оперативного реагирования, проведения горячих линий и т.д., что, видимо определяется его спецификой как мессенджера. Преобладают такие слова, как «обращение», «сообщение», «обратиться», «подскажите», которые практически не встречаются в комментариях ВКонтакте. На наш взгляд, в тематике комментариев заметна и разница в возрасте аудитории того и другого канала коммуникации, которую необходимо учитывать в процессе коммуникации главы региона с общественностью – на странице во ВКонтакте губернатора Иркутской области она более возрастная.

В целом исследование подтверждает мнение заместителя генерального директора АНО «Диалог Регионы» Андрея Цепелева о том, что «связки ВКонтакте и Telegram в абсолютном большинстве регионов для полноценной коммуникации с интернет-аудиторией» [23].

VI. ЗАКЛЮЧЕНИЕ

Проведенное исследование не является полностью завершенным и будет продолжено. Но уже сейчас можно с достаточной уверенностью сказать, что используемые методы и подходы показывают эффективность как аналитическое средство для сбора и оценки комментариев в разных социальных медиа и помогают охарактеризовать последние с точки зрения их делиберативного потенциала.

Исследование комментариев, размещаемых пользователями в социальных медиа, является важным методом исследования при изучении разных форм гражданского диалога и участия граждан в управлении. Размещение постов в социальных медиа является формой такого диалога, когда происходит не только обмен мнениями на нижнем, индивидуальном коммуникационном уровне, но формируются групповые, солидарности, выражающие общественное мнение различных слоев общества. Понимание таких процессов необходимо для поиска путей повышения эффективности вовлечения граждан в диалог с властью и участия в государственном управлении.

Важность используемого инструментария состоит в возможности выгрузки и структурирования текстов комментариев таким образом, чтобы не только осуществить смысловой анализ путем выявления базового смысла и намерения автора комментария, но и выявить состава и размер групп, считающих какие-то важные для них позиции более значимыми. К сожалению, во многих работах, относящихся к политическим наукам, исследования делиберативных процессов, зачастую очень поверхностно затрагивают лингвистические параметры, являющиеся первостепенными для исследований дискурсов разных социальных сетей, ограничиваясь лишь подсчетом количественных показателей. В то же время понимание

процесса возникновения и изменения мнений людей невозможно достичь при использовании только лингвистически ориентированных методов текстового анализа. Результаты, представленные в настоящей статье, не претендуя на всеобъемлющий статус, призваны в некоторой мере восполнить данные проблемы.

Конечная цель наших исследований заключается в том, чтобы выявить разницу между комментариями к постам, которые могут рассматриваться как форма электронного участия в политике (если принять, что они дискурсивно делиберативны, а значит – демократичны) на площадках разных типов социальных медиа. При этом важно, чтобы лица, принимающие решения, опирались на результаты исследований, учитывающих всю множественность и неоднозначность высказываемых в цифровом публичном пространстве мнений.

БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда № 22-18-00364 «Институциональная трансформация управления электронным участием в России: исследование региональной специфики» (<https://rscf.ru/project/22-18-00364/>)

БИБЛИОГРАФИЯ

- [1] Rainie L., Smith A., Schlozman K., Brady H., Verba S. Social media and political engagement. Pew Research: Internet Project, 2012.
- [2] Marquart F., Ohme J., Möller J. Following Politicians on Social Media: Effects for Political Information, Peer Communication, and Youth Engagement. // *Media and Communication*. 2020. Vol. 8, № 2. P. 197-207.
- [3] Чугунов А.В., Низомутдинов Б.А., А.А. Будяк А.А. Telegram каналы глав субъектов Российской Федерации: тестирование исследовательского инструментария // *International Journal of Open Information Technologies*. 2022. Т. 10, № 11, С. 141-146.
- [4] Stromer-Galley J. Measuring deliberation's content: A coding scheme // *Journal of Public Deliberation* 2007. Vol. 3 (1). P. 31–35.
- [5] Stromer-Galley J., Muhlberger P. Agreement and disagreement in group deliberation: Effects on deliberation satisfaction, future engagement, and decision legitimacy // *Political Communication*. 2009. Vol. 26 (2). P.173–192.
- [6] Stromer-Galley J., Bryan, L., Bimber, B. Context and medium matter: Expressing disagreements online and face-to-face in political deliberations // *Journal of Public Deliberation*, 2015. Vol. 11 (1). Article 1.
- [7] Scudder M.F. *Beyond Empathy and Inclusion: The Challenge of Listening in Democratic Deliberation*. Oxford: Oxford University Press, 2020.
- [8] Pflughoeft B.R., Schneider I.E. Social media as E-participation: Can a multiple hierarchy stratification perspective predict public interest? // *Government Information Quarterly*. 2020. Vol. 37 (1). Article 101422. DOI: 10.1016/j.giq.2019.101422.
- [9] Oliveira C. Proposed solutions to citizen engagement in virtual environments of social participation: a systematic review // *International Journal of Electronic Governance*. 2020. Vol. 12 (1). P. 76 – 91. DOI: 10.1504/IJEG.2020.106994.
- [10] Arshad S., Khurram S. Can government's presence on social media stimulate citizens' online political participation? Investigating the influence of transparency, trust, and responsiveness // *Government Information Quarterly*. 2020. Vol. 37 (3). Article 101486. DOI: 10.1016/j.giq.2020.101486.
- [11] Misnikov Y. How to read and treat online public discussions among ordinary citizens beyond political mobilisation: Empirical evidence from the Russian-language online forums // *Digital Icons: Studies in Russian, Eurasian and Central European New Media*. Leeds, 2012. Vol. 7. P. 1–37.
- [12] Fournier-Tombs E., Di Marzo Serugendo G. DelibAnalysis: Understanding the Quality of Online Political Discourse with Machine Learning // *Journal of Information Science*. 2020. Vol. 46 (6). P. 810–922.

- [13] Begen P., Misnikov Y., Filatova O. Application of automated tools in researching internet discourses: Experience of using the recurrent neural networks for studying discussions on pension reform // 21st Conference on Scientific Services and Internet, SSI-2019. CEUR Workshop Proceedings. 2020. Vol. 2543. P. 336–344.
- [14] Nizomutdinov B., Uglova A. Application of data from social networks for value-based management of city development programs // ICEGOV 2021: 14th International Conference on Theory and Practice of Electronic Governance October 2021. P. 520–522.
- [15] Donchenko D., Ovchar N., Sadovnikova N., Parygin D., Shabalina O., Ather D. Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension // Procedia Computer Science. 2017. Vol. 119. P. 359-367.
- [16] Струкова И. Ю. Характерологические особенности современного политического интернет-дискурса на примере мессенджера Telegram // Язык и культура: взгляд молодых: Материалы III Международной научной конференции студентов и школьников в рамках Международного Кирилло-Мефодиевского фестиваля славянских языков и культур, Москва, 26–28 мая 2020 г. М.: Государственный институт русского языка им. А.С. Пушкина, 2020. С. 441-445.
- [17] Агабаева С. М. Ш. Особенности влияния новостных Telegram-каналов на формирование информационной повестки дня в России // Успехи гуманитарных наук. 2022. № 11. С. 43-48.
- [18] Смирнов К. Д. Анонимные политические Telegram-каналы как средство влияния на информационную повестку дня // Возможности и угрозы цифрового общества: Материалы конференции Всероссийской научно-практической конференции, Ярославль, 22 апреля 2020 года / Под редакцией А.В. Соколова, А.А. Фролова. Ярославль: ООО "Цифровая типография", 2020. С. 215-218.
- [19] Иванов Д. С. Агрегация политических слухов в мессенджере Телеграм: опыт Волгоградской области 2018 года // Международный научно-исследовательский журнал. 2019. № 12-2 (90). С. 156-158. DOI 10.23670/IRJ.2019.90.12.079.
- [20] Жердева Ю. Р. Мобильное приложение "Telegram" как инструмент формирования имиджа политика // Литературоведение, лингвистика и коммуникативистика: направления и тенденции современных исследований: Материалы II Всероссийской заочной научной конференции, Уфа, 12 февраля 2018 года / Отв. ред. А. В. Курочкина. Уфа: Башкирский государственный университет, 2018. С. 127-128.
- [21] Волкова Н. Н., Романюк Э. И. Развитие цифровой среды российских регионов // Проблемы развития территории. 2019. №5 (103). С. 38-52
- [22] Рейтинги губернаторов в СМИ по МедиаИндексу // Медиалогия. URL: <https://www.mlg.ru/ratings/vlast/governors/11710/>
- [23] Исследование компании Polilog // Полилог. Экспертиза. URL: https://t.me/polylog_expertise/554
- Низомутдинов Борис Абдуллохонвич**, ведущий аналитик Центра технологий электронного правительства Института дизайна и урбанистики Университета ИТМО, Санкт-Петербург (<https://itmo.ru/>), email: boris@itmo.ru, elibrary.ru: authorid=794641, ORCID: [orcidID= 0000-0002-4090-9564](https://orcid.org/0000-0002-4090-9564).
- Филатова Ольга Георгиевна**, канд. философ. наук, доцент Санкт-Петербургского государственного университета, старший научный сотрудник Центра технологий электронного правительства Института дизайна и урбанистики Университета ИТМО, Санкт-Петербург, email: o.filatova@spbu.ru, elibrary.ru: authorid=625633, ORCID: [orcidID= 0000-0001-9568-1002](https://orcid.org/0000-0001-9568-1002).

Testing Methods for Processing Comments from Telegram Channels and Public VKontakte to Analyze the Social Media

Boris A. Nizomutdinov, Olga G. Filatova

Abstract— The paper presents a comparative analysis of two official sources of information, using the example of a group in the social network VKontakte and the Telegram channel of the head of the Irkutsk region. The object of research is the comments that are published under official records and posts. Automated methods for collecting and processing comments were used. The VKontakte API and the Telescope library were used to parse data from Telegram. The functionality of the desktop version of the Telegram messenger made it possible to upload comments from any open channel in JSON format. All the results obtained were saved in the Yandex Data Leans BI system. For the collected information, thematic modeling was carried out, 10 main topics were identified, which are discussed in the comments on VKontakte and on Telegram. The Gensim model of the Python library was used to create thematic models. A comparative analysis of the selected topics showed that in the comments on VKontakte and on Telegram, different socio-economic problems raised by the inhabitants of the region are noticeable and they are ranked differently. The methods and approaches used, in general, have shown their effectiveness as an analytical tool for collecting and evaluating comments from different types of social media, which, in particular, makes it possible to identify their deliberative potential.

Keywords: Telegram, VKontakte, parsing, topic modeling, social media, deliberation

REFERENCES

- [1] Rainie L., Smith A., Schlozman K., Brady H., Verba S. Social media and political engagement. Pew Research: Internet Project, 2012.
- [2] Marquart F., Ohme J., Möller J. Following Politicians on Social Media: Effects for Political Information, Peer Communication, and Youth Engagement. // *Media and Communication*. 2020. Vol. 8 (2). P. 197-207.
- [3] Chugunov A.V., Nizomutdinov B.A., Budyak A.A. Telegram channels of heads of constituent entities of the Russian Federation: testing research tools // *International Journal of Open Information Technologies*. 2022, Vol. 10 (11). P. 141-146.
- [4] Stromer-Galley J. Measuring deliberation's content: A coding scheme // *Journal of Public Deliberation* 2007. Vol. 3 (1). P. 31–35.
- [5] Stromer-Galley J., Muhlberger P. Agreement and disagreement in group deliberation: Effects on deliberation satisfaction, future engagement, and decision legitimacy // *Political Communication*. 2009. Vol. 26 (2). P.173–192.
- [6] Stromer-Galley J., Bryan, L., Bimber, B. Context and medium matter: Expressing disagreements online and face-to-face in political deliberations // *Journal of Public Deliberation*, 2015. Vol. 11 (1). Art. 1.
- [7] Scudder M.F. Beyond Empathy and Inclusion: The Challenge of Listening in Democratic Deliberation. Oxford: Oxford Univ. Press, 2020.
- [8] Pflughoeft B.R., Schneider I.E. Social media as E-participation: Can a multiple hierarchy stratification perspective predict public interest? // *Government Information Quarterly*. 2020. Vol. 37 (1). Article 101422. DOI: 10.1016/j.giq.2019.101422.
- [9] Oliveira C. Proposed solutions to citizen engagement in virtual environments of social participation: a systematic review // *International Journal of Electronic Governance*. 2020. Vol. 12 (1). P. 76 – 91. DOI: 10.1504/IJEG.2020.106994.
- [10] Arshad S., Khurram S. Can government's presence on social media stimulate citizens' online political participation? Investigating the influence of transparency, trust, and responsiveness // *Government Information Quarterly*. 2020. Vol. 37 (3). Article 101486. DOI: 10.1016/j.giq.2020.101486.
- [11] Misnikov Y. How to read and treat online public discussions among ordinary citizens beyond political mobilisation: Empirical evidence from the Russian-language online forums // *Digital Icons: Studies in Russian, Eurasian and Central European New Media*. Leeds, 2012. Vol. 7. P. 1–37.
- [12] Fournier-Tombs E., Di Marzo Serugendo G. DelibAnalysis: Understanding the Quality of Online Political Discourse with Machine Learning // *Journal of Information Science*. 2020. Vol. 46 (6). P. 810–922.
- [13] Begen P., Misnikov Y., Filatova O. Application of automated tools in researching internet discourses: Experience of using the recurrent neural networks for studying discussions on pension reform // 21st Conference on Scientific Services and Internet, SSI-2019. CEUR Workshop Proceedings. 2020. Vol. 2543. P. 336–344.
- [14] Nizomutdinov B., Uglova A. Application of data from social networks for value-based management of city development programs // ICEGOV 2021: 14th International Conference on Theory and Practice of Electronic Governance October 2021. P. 520–522.
- [15] Donchenko D., Ovchar N., Sadovnikova N., Parygin D., Shabalina O., Ather D. Analysis of Comments of Users of Social Networks to Assess the Level of Social Tension // *Procedia Computer Science*. 2017. Vol. 119. P. 359-367.
- [16] Strukova I. Yu. Characterological features of modern political Internet discourse on the example of the Telegram messenger // *Language and culture: the view of young people: Materials of the III International scientific conference of students and schoolchildren in the framework of the International St. Cyril and Methodius Festival of Slavic Languages and Cultures, Moscow, May 26–28, 2020*. M: State Institute of the Russian Language. A.S. Pushkin, 2020. P. 441-445.
- [17] Agabaeva S. M. Sh. Features of the influence of Telegram news channels on the formation of the information agenda in Russia // *Successes in the Humanities*. 2022. No. 11. P. 43-48.
- [18] Smirnov K. D. Anonymous political Telegram channels as a means of influencing the information agenda // *Opportunities and threats of the digital society: Proceedings of the conference of the All-Russian Scientific and Practical Conference, Yaroslavl, April 22, 2020 / Under edited by A.V. Sokolova, A.A. Frolova. Yaroslavl: Limited Liability Company "Digital Typography", 2020. P. 215-218.*
- [19] Ivanov D.S. Aggregation of political rumors in the Telegram messenger: the experience of the Volgograd region in 2018 / D.S. Ivanov // *International Scientific Research Journal*. 2019. No. 12-2(90). P. 156-158. – DOI 10.23670/IRJ.2019.90.12.079.
- [20] Zherdeva Yu. R. Mobile application "Telegram" as a tool for forming the image of a politician // *Literary criticism, linguistics and communication studies: directions and trends of modern research: Proceedings of the II All-Russian Correspondence Scientific Conference, Ufa, February 12, 2018 / Ed. ed. A. V. Kurochkina. Ufa: Bashkir State University, 2018. P. 127-128.*
- [21] Volkova N. N., Romanyuk E. I. Development of the digital environment of Russian regions // *Problems of territory development*. 2019. No. 5 (103). P.38-52.
- [22] Ratings of governors in the media according to MediaIndex // *Medialogy*. URL: <https://www.mlg.ru/ratings/vlast/governors/11710>.
- [23] Polilog research // *Polilog. Expertise*. URL:https://t.me/polylog_expertise/554

Boris A. Nizomutdinov, Leading Analyst of E-Governance Center, Institute of Design and Urban Studies, ITMO University (<http://itmo.ru/>),

Saint-Petersburg, email: boris@itmo.ru, elibrary.ru: authorid=794641, scopus.com: authorid=56938763900, ORCID: orcidID=0000-0002-4090-956

Olga G. Filatova, Ph.D., Associate Professor St. Petersburg State University; Senior Researcher of E-Governance Center, Institute of Design and Urban Studies, ITMO University, St. Petersburg, email: o.filatova@spbu.ru, elibrary.ru: authorid=625633, ORCID: orcidID= 0000-0001-9568 -1002.