Инженерный вестник Дона, №7 (2022)
ivdon.ru/ru/magazine/archive/n7y2022/7797

Сбор и обработка текстовых данных в контексте оценки социальных настроений: методологические аспекты

А.Н. Марьенков, А.И. Кривенко

Астраханский государственный университет, Астрахань

Аннотация: В данной статье рассмотрены аспекты сбора и обработки текстовых данных социальных медиа и СМИ в контексте оценки социальных настроений населения отдельных макрорегионов. Текстовые данные являются важным источником отражения настроений общества и с применением современных технических инструментов могут быть собраны, предобработаны и проанализированы с учетом специфики проверяемых гипотез и поставленных задач. Данный контекст является важным звеном системы комплексной безопасности макрорегионов в их культурном, социальном, экономическом, политическом аспектах. Обсуждаемые подходы включают в себя сбор релевантных данных, их обработку и анализ с применением инструментов интеллектуального анализа данных.

Ключевые слова: анализ текста, данные СМИ, обработка естественного языка, оценка социальных настроений, анализ данных

Введение

Большие массивы текстовых данных, источником которых являются популярные социальные сети, мессенджеры, блоги, веб-сайты СМИ, являются важным источником информации о настроении общества относительно процессов, происходящих в конкретном регионе или мире в целом. Анализ таких данных в динамике способен показать степень изменения отношения людей к той или иной проблеме, оценить влияние социума на конкретные группы людей, особенно в контексте локализации конкретной территории, а также приблизить попытки исследователей к выявлению и интерпретации характера идентификации населения с той или иной группой. В частности, если рассматривать территорию Прикаспийского региона, куда входят регионы России, а также Казахстан, Туркмения, Иран, Азербайджан [1], то в рамках поставленных гипотез можно говорить о попытке формирования «каспийской идентичности» [2] и оценке текущих настроений общества относительно рассматриваемых региональных и глобальных проблем. Данная парадигма встраивается в логику мониторинга

и обеспечения комплексной безопасности регионов в их культурном, социальном, экономическом, политическом аспектах. Ценным источником являются текстовые данные, которыми уже пользуется ряд исследовательских групп.

Данные социальных сетей активно используются для оценки качественных параметров отдельных процессов, проходящих в социокультурном пространстве. В частности, в монографии Han J., Kamber M., Pei J., [3] была рассмотрена методика измерения субъективного качества жизни в регионах Российской Федерации при расчёте индекса качества жизни машинными методами с целью установления корреляций для «этнических» зон. Отдельно можно выделить подходы по созданию модели машинного обучения для автоматического прогнозирования политических взглядов российских пользователей социальной сети «ВКонтакте» на основе микроподхода к анализу данных [4].

Использование данных социальных сетей нередко направлено на применение контент-анализа публикаций. Так, в исследовании по сетевой организация скулшутеров в социальной сети "ВКонтакте" на примере фанатского сообщества "керченского стрелка" [5] авторы рассматривают подход сетевого анализа деструктивных девиантных сообществ во «ВКонтакте» и проводят контент-анализ и анализ подписок участников сообществ для определения характера сетевых связей и степени организационной и тематической интегрированности членов сообществ.

Среди других подходов можно выделить работы, связанные с сентимент-анализом данных социальных сетей [6], исследования в области анализа политических публикаций в интернет-пространстве [7], а также методы сбора и анализа текста в интернете и цифрового следа пользователей [8], которые предлагают полезный инструментарий для работы с текстовыми данными из различных источников.

В данной статье обсуждаются методологические аспекты к построению рабочего процесса сбора и анализа релевантных текстовых данных для оценки социальных настроений населения.

Подходы к сбору данных

Как отмечалось ранее, данные являются критически важным звеном такого рода исследований и необходимо принимать во внимание сами источники (платформы) нужных данных (например, социальная сеть ВКонтакте), конкретизацию по точкам распространения данных (группы или лидеры мнений), а также механизмы сбора данных с таких платформ.

Среди релевантных источников открытых данных в контексте рассматриваемой проблематики являются следующие:

1. Данные из тематических групп социальной сети Вконтакте;
2. Данные из СМИ;
3. Данные из открытых пабликов Телеграм.

Для каждого из этих источников необходимо указать точки сбора: это могут быть конкретные группы в социальной сети, конкретные сайты региональных СМИ и блогов (если мы рассматриваем региональную тематику) и релевантные паблики в Телеграм. Стоит подчеркнуть, что при рассмотрении задачи, локализованной под региональный контекст, следует учитывать особенности выбора именно региональных сообществ. В частности, определение принадлежности сообщества к конкретному региону можно проводить на основе анализа подписчиков (критерий: более 50% подписчиков проживает в рассматриваемом регионе; это указано в их профиле). Кроме того, необходимо учитывать и количество пользователей таких групп, а также частоту размещения контента.

Выявление лидеров (общественных) мнений необходимо для выявления центров распространения информации и формирования контента. Контент от лидеров мнений необходимо учитывать при сборе и аналитике данных. В

частности, для социальной сети Вконтакте подход к выявлению таких аккаунтов следующий:

- Выявляем интересующие группы;
- Скачиваем данные этой (этих) группы через специализированные платформы или через API (Application Programming Interface, программный интерфейс приложения) ВКонтакте;
- С использованием свободно распространяемого программного обеспечения Gephi (gephi.org/) выявляем пользователей с большими социальными связями;
- Находим id (идентификаторы) этих пользователей, смотрим их контент на предмет релевантности, при необходимости скачиваем данные.

Пример построения подобного социального графа для выявления лидеров мнений показан на рис.1.

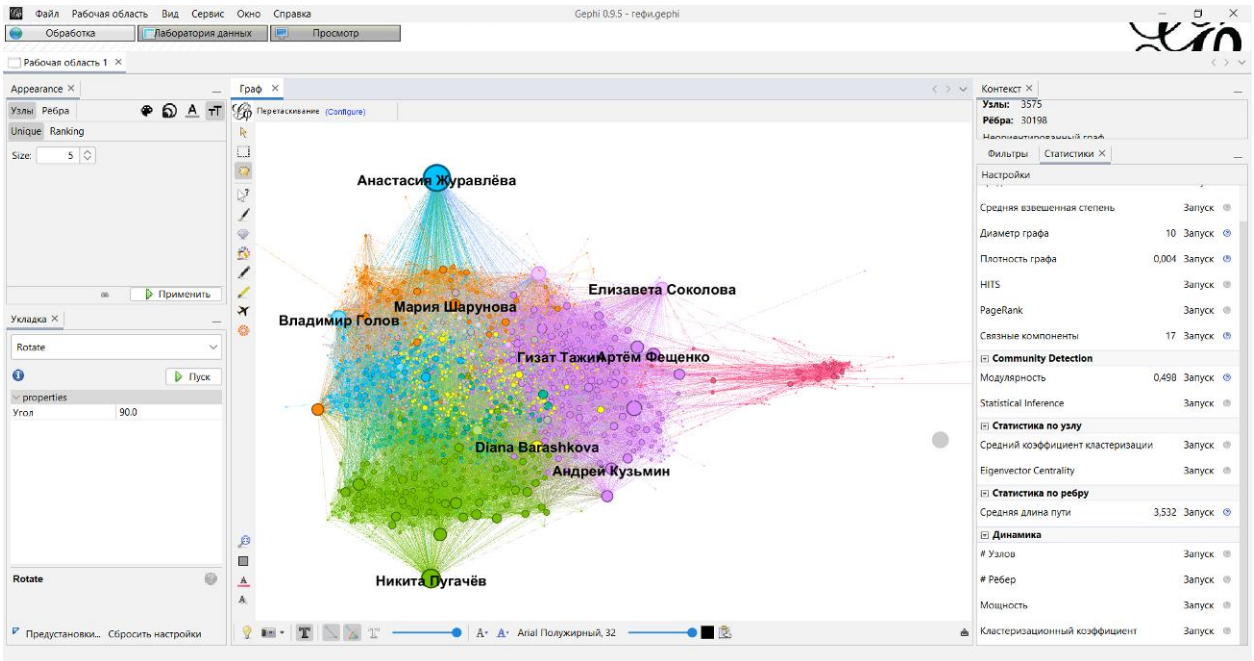


Рис. 1. – Построение графа на данных тематической группы Вконтакте с помощью Gephi.

Рисунок представлен не только в виде визуализации самого графа, а также с учетом демонстрации работы в данном инструменте.

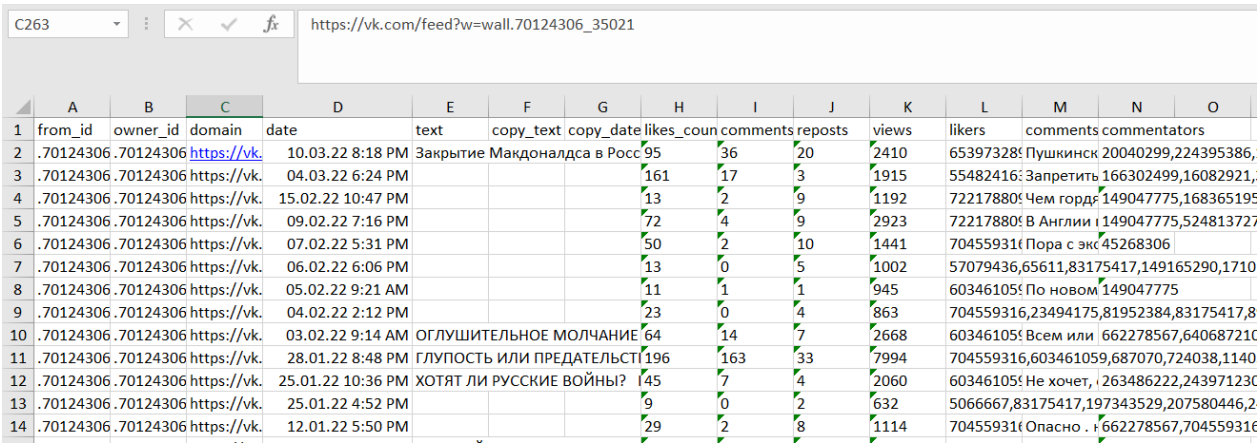
Сбор данных из рассматриваемой социальной сети посредством API имеет ограничение на скачивание количества постов в секунду и ограничение на сутки по выгрузке данных в 5000 постов на один аккаунт пользователя [9]. Как отмечалось ранее, существует ряд коммерческих решений, которые предоставляют возможность скачивать обезличенные данные из разных источников в большем объеме и с меньшими ограничениями. На рис.2 представлен пример кода парсинга данных из социальной сети Вконтакте.

```python
import vk_api, json, os
from datetime import datetime
from time import time
def auth_handler():
    """ При двухфакторной аутентификации вызывается эта функция.
    """
    # Код двухфакторной аутентификации
    key = input("Enter authentication code: ")
    # Если: True - сохранить, False - не сохранять.
    remember_device = True

    return key, remember_device

def parse_posts(tools, post_from_wall, groups_id, groups_name):
    time_to_parse = time()
    posts_data = {
            "Group name": groups_name,
            "Group id": groups_id,
            "Posts" : []
        }
    # print(type(post_from_wall))
    for count, current_post in enumerate(post_from_wall):
        all_comments_in_post = []
        print('*'*48, f'\nNow we are parsing post: https://vk.com/{groups_name}?w=wall-{groups_id}_{current_post["id"]}')
        comments_from_post = tools.get_all_slow('wall.getComments', 100, {
                                                'owner_id' : -groups_id,
                                                'post_id' : current_post['id'],
                                                'need_likes': 1
                                                })
        for current_commentary in comments_from_post['items']:
            try:
                commentary_text = current_commentary['text']
                commentary_likes_count = current_commentary['likes']['count']
            except:
                commentary_text, commentary_likes_count = "", ""
            data_from_commentary = {
```

Рис. 2. – Пример парсинга данных из социальной сети В контакте

Получаемые в результате такой операции данные хранятся в формате csv. На рис.3 представлен пример хранения данных, их структура и метрики.

Данные включают в себя ID группы, дату публикации, текст публикации, количество комментариев, текст комментариев, количество лайков, репостов, а также ID пользователей групп, для дальнейшего анализа аудитории сообществ из выборки.



C263 | https://vk.com/feed?w=wall.70124306_35021

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	from_id	owner_id	domain	date	text	copy_text	copy_date	likes_coun	comments	reposts	views	likers	comments	commentators	
2	.70124306	.70124306	https://vk.	10.03.22 8:18 PM	Закрытие Макдоналдса в Росс			95	36	20	2410	65397328	Пушкинск	20040299,224395386,	
3	.70124306	.70124306	https://vk.	04.03.22 6:24 PM				161	17	3	1915	55482416	Запретить	166302499,16082921,	
4	.70124306	.70124306	https://vk.	15.02.22 10:47 PM				13	2	9	1192	72217880	Чем горда	149047775,168365195	
5	.70124306	.70124306	https://vk.	09.02.22 7:16 PM				72	4	9	2923	72217880	В Англии	149047775,524813727	
6	.70124306	.70124306	https://vk.	07.02.22 5:31 PM				50	2	10	1441	70455931	Пора с эк	45268306	
7	.70124306	.70124306	https://vk.	06.02.22 6:06 PM				13	0	5	1002	57079436,65611,83175417,149165290,1710			
8	.70124306	.70124306	https://vk.	05.02.22 9:21 AM				11	1	1	945	60346105	По новом	149047775	
9	.70124306	.70124306	https://vk.	04.02.22 2:12 PM				23	0	4	863	704559316,23494175,81952384,83175417,8			
10	.70124306	.70124306	https://vk.	03.02.22 9:14 AM	ОГЛУШИТЕЛЬНОЕ МОЛЧАНИЕ			64	14	7	2668	60346105	Всем или	662278567,64068721	
11	.70124306	.70124306	https://vk.	28.01.22 8:48 PM	ГЛУПОСТЬ ИЛИ ПРЕДАТЕЛЬСТ			196	163	33	7994	704559316,603461059,687070,724038,1140			
12	.70124306	.70124306	https://vk.	25.01.22 10:36 PM	ХОТЯТ ЛИ РУССКИЕ ВОЙНЫ?			45	7	4	2060	60346105	Не хочет,	263486222,243971230	
13	.70124306	.70124306	https://vk.	25.01.22 4:52 PM				9	0	2	632	5066667,83175417,197343529,207580446,2			
14	.70124306	.70124306	https://vk.	12.01.22 5:50 PM				29	2	8	1114	70455931	Опасно .	662278567,704559316	

Рис. 3. – Пример собранных данных постов выбранной группы социальной сети Вконтакте

Другой важный источник данных – текстовые данные СМИ – аналогично может собираться через адаптированные парсеры сайтов (например, на основе популярных библиотек языка Python) или через специализированные коммерческие решения (например, Крибрум [10]). Важную роль здесь играет фокусировка на необходимом контенте, который может быть уточнен через так называемые лингвистические маркеры [11] для их использования в дальнейшем в запросах.

Способы формирования маркерных слов для отбора релевантного новостного контента включают в себя следующие:

- изучение предметной области;
- анализ литературы;
- экспертная оценка;
- специализированные словари (например, словарь эмотивной лексики ([12] или словарь оценочных слов и выражений русского языка РуСентиЛекс [13];
- использование специальных сервисов (например, ЯндексВордстат, GoogleTrends).

Сформулированный набор маркерных слов затем используется (и при необходимости уточняется) для сбора более релевантных текстовых данных.

Данные из открытых релевантных пабликов мессенджера Телеграм также могут быть собраны через API мессенджера. Подробно это описано в исследовании Карабака И. И., Зорина К. А., Ажмухамедова И. М. [14].

В зависимости от конкретизации задач проводимого исследования, необходимо фиксировать при сборе данных также регион (примерные координаты) собираемых данных для того, чтобы в дальнейшем получить возможность визуализации получаемых результатов на карте.

Подготовка данных

В большинстве случаев необходима предварительная обработка и подготовка собранных данных. Предобработку данных необходимо проводить и корректировать на основе потребностей тех инструментов (и алгоритмов), которые будут использоваться на этапе анализа и моделирования.

Процедура подготовки текстовых данных, как правило, включает в себя следующие пункты [15]:

- очистка данных с удалением ненужных атрибутов;
- очистка текста от ненужных символов (пустые символы, непечатные символы);
- изменение регистра;
- токенизация;
- удаление стоп-слов;
- фильтрация по длине/частоте;
- стемминг;
- лемматизация;
- векторизация;
- работа с пропущенными значениями;
- работа с дубликатами;
- исправление аббревиатур.

В зависимости от используемых инструментов, модули очистки данных часто уже встроены в специализированные библиотеки (например, библиотека nltk в Python [16]) или готовые решения (например, Orange Data Mining [17], PolyAnalyst [18], KNIME [19]).

Для целей классификации нужны размеченные данные по тем классам (меткам), которые отражают цели проекта. Для этого необходимо сформулировать эти метки/классы. В последующем скачанные тексты будут автоматически проклассифицированы для их последующего использования.

Классы в рамках задачи оценки социальных настроений могут быть следующими:

- безопасность;
- экология;
- отношения между людьми и общее эмоциональное состояние;
- протестный потенциал:
    - политический протест;
    - социальный протест;
    - культурный протест;
- отношение к власти;
- отношение к конкретной стране;
- религия;
- миграция.

Сама разметка данных является непростым процессом и требует привлечения дополнительных ресурсов. С одной стороны, разметку данных можно делегировать на сторонние ресурсы – например, использовать возможности краудсорсинговой платформы Яндекс Толока [20]. С другой стороны, разметку данных можно организовать внутри исследовательской команды, предварительно обучив привлеченных специалистов. Наконец, ряд

размеченных данных можно найти в открытом доступе, особенно если дело касается популярных классов (например, политика, экология, экономика).

Анализ данных и моделирование

В зависимости от уточненных целей анализа, последующая работа с данными может включать в себя визуализацию данных, анализ тональности текста (анализ эмоциональной окраски текста), классификацию текста, построение информационных панелей (дашбордов).

Классификация и визуализация текста могут выявить трендовые тематики в том или ином регионе, в том числе по частоте встречаемости слов или фраз, по доминирующей тематике и росту публикаций по этой тематике (классу) с течением времени, а также связь тематик между собой через проведение сетевого анализа.

Анализ эмоциональной окраски текста может быть, как самостоятельной частью одной из поставленных задач, когда необходимо оценивать фон (позитивный, нейтральный или негативный), настроений в обществе рассматриваемого региона в динамике за фиксированные периоды. С другой стороны, автоматическая разметка текста по тональности может стать шагом фильтрации данных, когда в качестве цели проводится работа по более глубокому анализу именно негативных новостей, постов или комментариев. Ряд инструментов уже содержат в себе модули оценки тональности текста (например, Orange Data Mining, PolyAnalyst), а также существуют модели оценки эмоциональной составляющей текста с применением методов машинного обучения [21, 22].

Приведем в качестве практического примера небольшой сценарий (рабочий поток) анализа текстовых данных, проведенного с использованием системы PolyAnalyst. Сценарий анализа постов некоторого сообщества показан на рис.4.

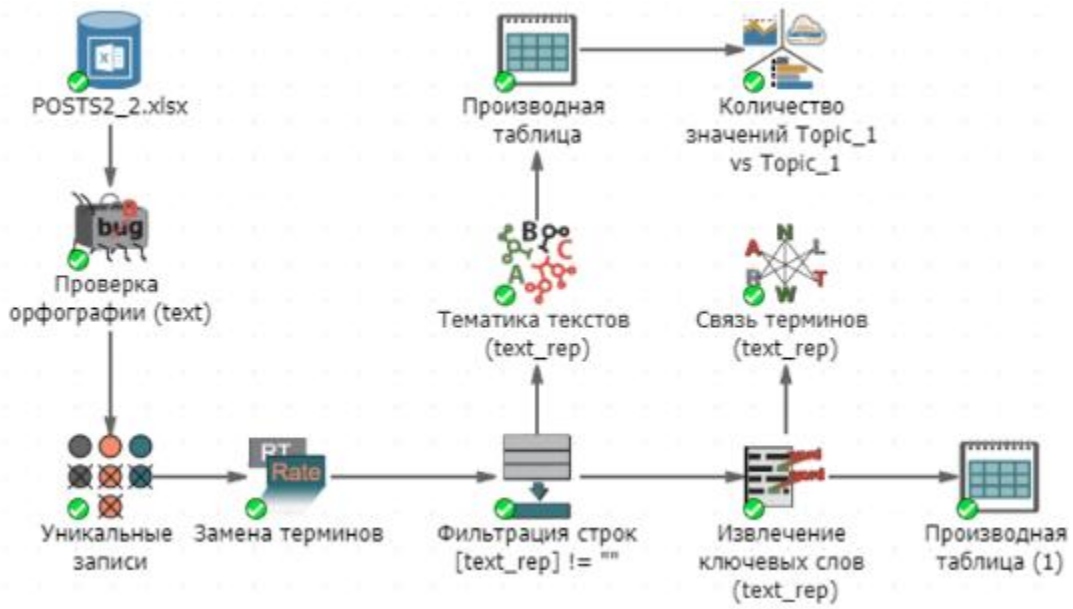


Рис. 4. – Сценарий анализа постов сообщества

Данный сценарий состоит из нескольких этапов. Процесс начинается с загрузки данных в модуль Excel, в котором хранятся ранее собранные данные. Назначаем формат данных каждой ячейки и проверяем правильность их отображения. Затем проводится предобработка данных по нескольким модулям, а именно:

- проверка орфографии и пунктуации текстов;
- извлечение только уникальных записей со страницы сообщества;
- очистка от хештегов и лишних символов;
- очистка данных от пустых строк (так как на странице сообщества содержится не только текстовая информация, но и фото- и видео-контент, то такие ячейки остаются пустыми).

После предобработки данных извлекаем ключевые слова и термины из текстов постов. Это наиболее часто встречающиеся слова по определенным тематикам, чтобы выявить основные тематики постов сообщества. Для выявления основных категорий текстов применяем модуль «Тематика текстов» для наших очищенных данных.

После проделанных операций визуализируем данные для отчётности и интерактивности представления выводов, полученных в результате сентимент-анализа. Визуализация дает представление о наиболее обсуждаемых в сообществе тематиках, что можно увидеть, дополнительно построив «облако слов», наиболее употребляемых в текстах сообщества.

Рассмотрим еще один сценарий анализа собранных данных сообщества «Свобода», который нацелен на изучение реакции населения на новости в стране и в мире (рис.5).

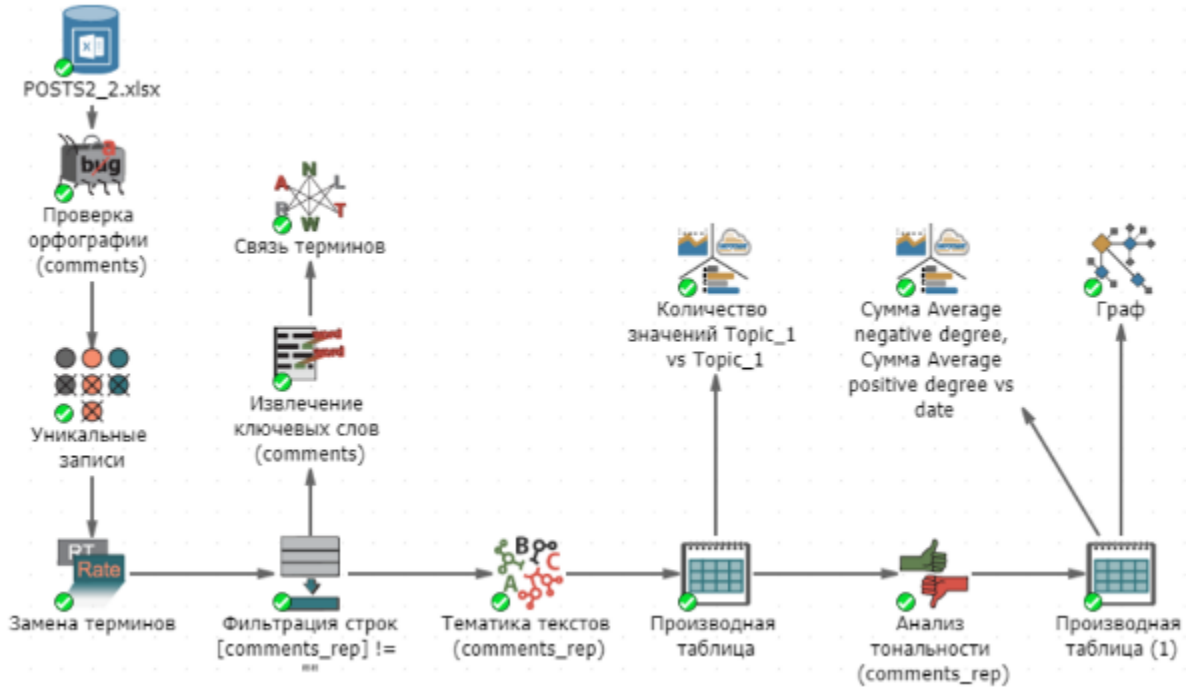


Рис. 5. – Сценарий анализа комментариев к постам сообщества

Этот сценарий, как и предыдущий, проходит этапы загрузки и первичной обработки текстовых данных. Дополнительно выбираются тематики текстов по встроенным словарям используемого инструмента, проверяется выборка и удаляются стоп-слова. Следующим шагом осуществляется анализ тональности текста. Тональность выбрана стандартная: положительная и отрицательная.

Визуализация данных по второму сценарию показывает тематики текстов и популярные, эмоционально окрашенные слова, которые употребляются пользователями в каждой рубрике.

Для понимания того, как связаны термины рассматриваемых текстов, был построен граф связи ключевых слов, где узлы – это сами слова, а ребра – это их сила связи между собой. На рис.6 представлен пример части графа, который отображает только наиболее часто встречающиеся слова. Обозначение [no] программа записывает в отчёт вместо частицы «не».

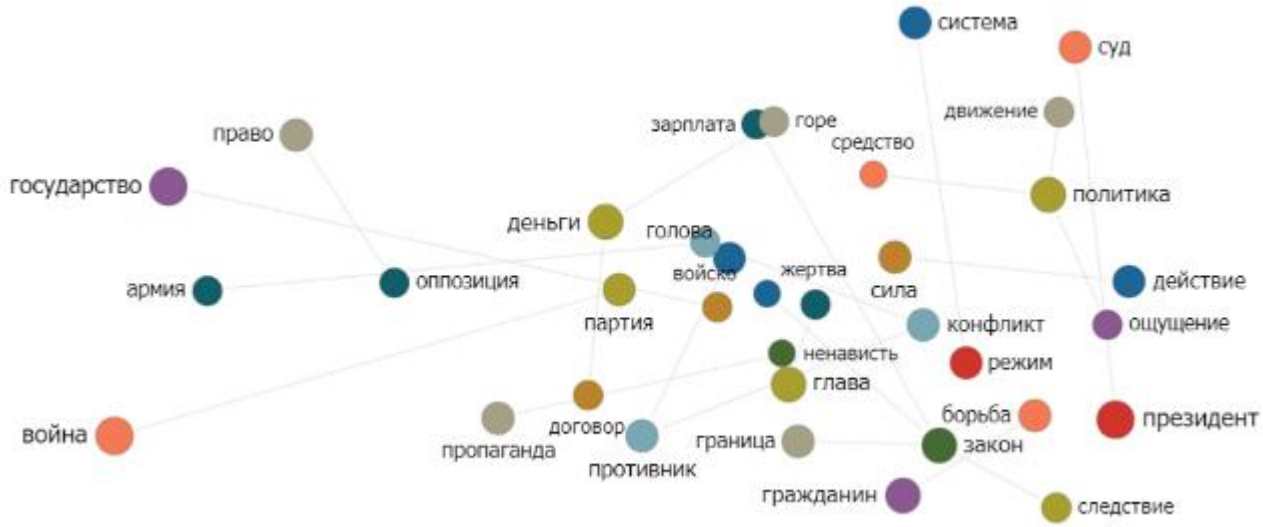


Рис. 6. – Граф часто употребляемых слов в постах сообщества

Расширяет эту визуализацию график частоты употребления слов, распределенный на тематики (рис.7).

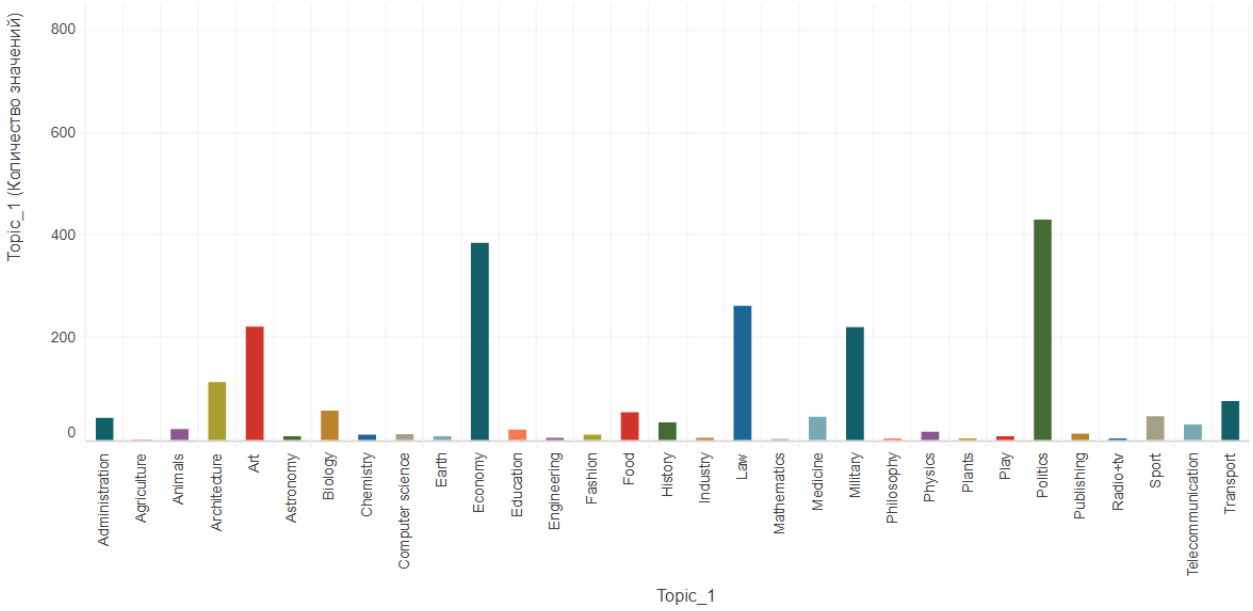


Рис. 7. – Распределение тематик

В данном кейсе использовались словари тем, встроенных в саму систему. По оси «X» отложены тематики, а по оси «Y» частота их употребления.

Для более детального изучения выявленных популярных категорий можно построить более детализированный граф в контексте эмоциональной окраски (эмоциональной составляющей) (рис.8).

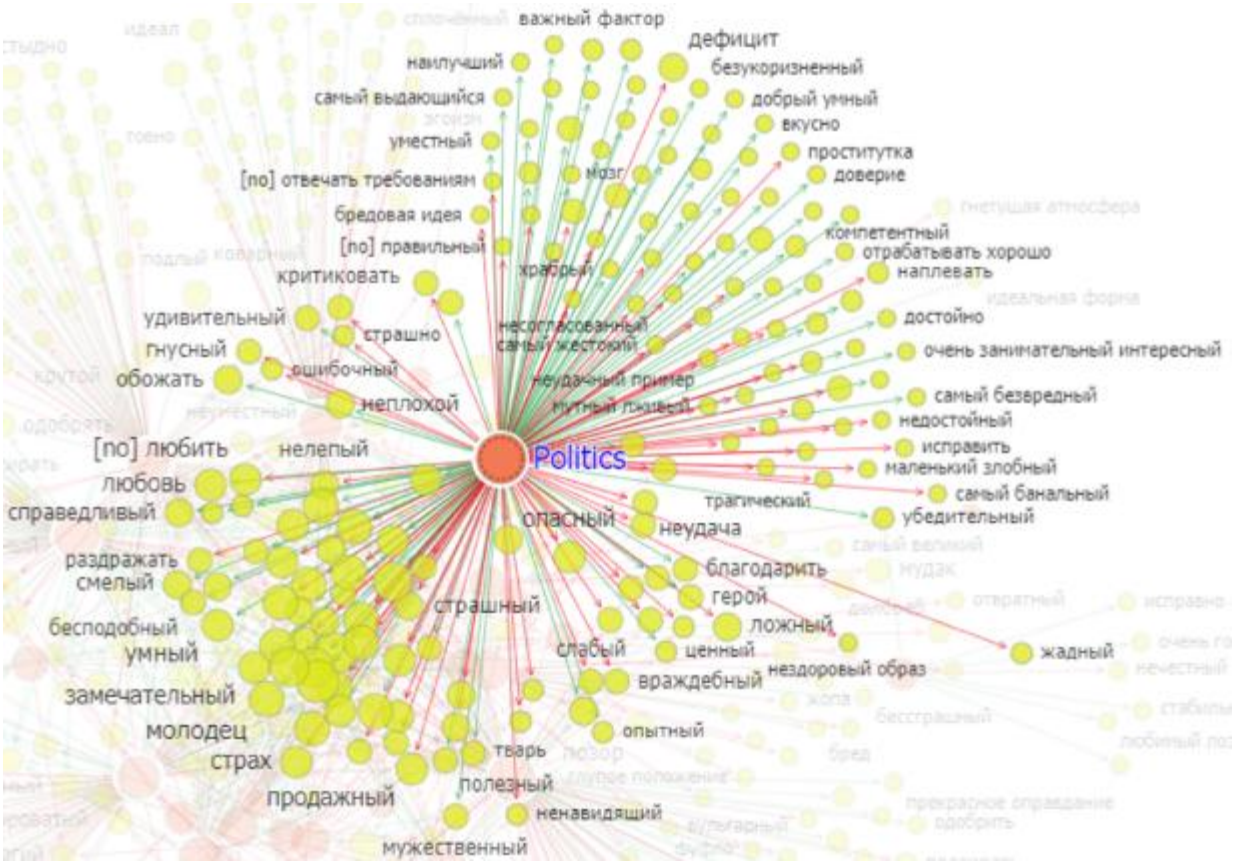


Рис. 8. – Граф тематики «Politics» (Политика)

На представленном рис.8 основным узлом графа выступает категория «Политика». Остальные узлы являются часто употребляемыми словами, связанными с тематикой исследования. Частотность определяется размером узла: чем больше узел, тем выше частота употребления слова или словосочетания. Ребра графа указывают на эмоциональную окраску в политическом контексте, а именно, ребра зелёного цвета указывают на позитивно окрашенные связи, а красные ребра определяют негативную окраску.

Можно заметить большое количество негативно окрашенных слов, но также присутствуют и положительно окрашенные фразы. На основе экспертной оценки можно провести интерпретацию данного исхода. Для того, чтобы увидеть динамику изменения тональности в текстах, был проведен анализ тональности текстов за определенный период (рис.9).

Рис. 9. – График тональности постов

Показатель по шкале «Y» – это среднее значение количества комментариев. Период интервалов на графике - полгода. На оси «X» расположены позитивные и негативные тональности анализируемых текстов. На графике оранжевый цвет – количество положительно окрашенных отзывов, а зелено-синий – негативных высказываний в комментариях.

Заключение

Таким образом, мониторинг социальных медиа и текстовых новостей СМИ как исследовательский инструмент предоставляет возможность оперативного сбора и анализа данных в контексте оценки социальных настроений по рассматриваемым тематикам. В рамках данной работы были обсуждены аспекты сбора релевантных текстовых данных из различных источников, подходы к их предобработке, а также анализу с применением механизмов визуализации и машинного обучения для классификации по выделенным тематическим рубрикам. Важное значение стоит придавать качеству и объему собираемых данных как на этапе подбора источников, так и этапах предобработки и фильтрации.

Рассматриваемые подходы могут быть объединены в единый пайплайн (рабочий процесс), который в дальнейшем может использоваться многократно, начиная с автоматизированного мониторинга и загрузки новых данных и заканчивая представлением результатов в визуальной форме.

Немаловажна здесь и роль специалиста, который принимает активное участие в интерпретации результатов и принятии на их основе решений.

Исследование выполнено за счет гранта Российского научного фонда № 22-18-00301 «Процесс конструирования новых идентичностей в Каспийском макрорегионе в контексте социетальной безопасности».

Литература

1. Аманов М.Э., Акмурадова К.К. Основные экологические проблемы Каспийского региона // Каспий и глобальные вызовы : материалы Международной научно-практической конференции. Астрахань: АГУ. 2022. С. 9-14.

2. Baeva, L.V., Romanova A.P., Challenges to frontier allegories: The Caspian Sea region in Southern Russia, Cultura. International Journal of Philosophy of Culture and Axiology. 2015. Vol. 12. No 1. P. 159-172.

3. Han J., Kamber M., Pei J., Data Mining Concepts and Techniques. - USA: Morgan Kaufmann Publishers is an imprint of Elsevier. 2012. P. 703.

4. Kozitsin I.V., Chkhartishvili A.G., Marchenko A.M., Norkin D.O., Osipov, I.A. Uteshev S.D., Goiko V.L., Palkin R.V., Myagkov M.G. Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte, Mathematical Models and Computer Simulations. 2020. Vol. 12. No 2. P. 185-194.

5. Чудинов С. И., Сербина Г. Н., Мундриевская Ю. О., Сетевая организация скулшутеров в социальной сети "ВКонтакте" на примере фанатского сообщества "керченского стрелка" // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 4(164). С. 363-383.

6. Богданов А. Л., Дуля И. С. Сентимент-анализ коротких русскоязычных текстов в социальных медиа // Вестник Томского государственного университета. Экономика. 2019. № 47. С. 220-241.

7. Boldyreva A., Alexandrov M., Koshulko O., Sobolevskiy O. Internet Queries as a Tool for Analysis of Regional Police Work and Forecast of Crimes in Regions, Advances in Computational Intelligence. 2016. Vol. 1. No 15. P. 290-301.

8. Харковчук, А. Э., Корзун Д. Ж., Составление цифрового профиля человека на основе поиска информации по его фотографии из открытых источников в сети интернет // Цифровые технологии в образовании, науке, обществе: Материалы XIII всероссийской научно-практической конференции, Петрозаводск, 17–20 сентября 2019 года. Петрозаводск: Петрозаводский государственный университет. 2019. С. 199-202.

9. Справочник API // VK для разработчиков. URL: dev.vk.com/reference/roadmap (дата обращения: 06.06.2022).

10. Зимова Н. С., Фомин Е. В., Смагина А. А. Социальные сети как новый канал взаимодействия общества и власти // Научный результат. Социология и управление. 2020. Т. 6. № 2. С. 159-171.

11. Villar G., Arciuli J., Paterson H. Linguistic Indicators of a False Confession, Psychiatry, Psychology and Law. 2013. Vol. 20 (4). P. 504-518.

12. Бабенко Л. Г. Лингвопсихология на методологической базе когнитивистики: лексикографический аспект // Известия Уральского федерального университета. Серия 2: Гуманитарные науки. 2020. Т.22. № 3(200). С. 264-278.

13. Kotelnikova A., Kotelnikov E., SentiRusColl: Russian collocation lexicon for sentiment analysis, Communications in Computer and Information Science. 2019. Vol. 1119. P. 18-32.

14. Карабак И. И., Зорин К. А., Ажмухамедов И. М. Парсинг телеграм-каналов как элемент системы автоматизированного анализа информации, полученной из сети интернет // Прикаспийский журнал: управление и высокие технологии. 2022. № 1(57). С. 9-17.

15. Shchekotin E., Myagkov M., Goiko V., Kashpur V., Digital methods of analysis of subjective quality of life: Case of Russian regions, Administratie si Management Public. 2021. Vol. 2021. No 36. P. 25-48.

16. Bird S., Loper E., NLTK: The Natural Language Toolkit, Proceedings of the ACL Interactive Poster and Demonstration Sessions: Proceedings of the ACL Interactive Poster and Demonstration Sessions, ACL. – Barcelona: Association for Computational Linguistics. 2004. P. 214-217.

17. Demšar J., Curk T., Erjavec A., Gorup Č., Hočevar T., Milutinovič M., Možina M., Polajnar M., Toplak M., Starič A., Štajdohar M., Umek L., Žagar L., Žbontar J., Žitnik M., Zupan B., Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research. 2013. № 14. P.2349-2353.

18. Kiselev M. V., Arseniev S. B., Ananyan S. M., Polyanalyst data analysis technique and its specialization for processing data organized as a set of attribute values, Lecture Notes in Computer Science. 1998. Vol. 1510. P. 352-360.

19. Berthold M.R., Cebron N., Dill F., Fatta G.D., Gabriel T.R., Georg F., Meinl T., Ohl P., Sieb C., Wiswedel B., KNIME: The Konstanz Information Miner, Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. 2008. P. 319-326.

20. Chapkovski, P. Interactive experiments in Toloka, Munich Personal RePEc Archive. 2022. № 111980. P. 1-39.

21. Гурин А. А. Определение тональности сообщений с помощью методов машинного обучения // Инновации. Наука. Образование. 2020. № 19. С. 471-477.

22. Майорова, Е. В. О сентимент-анализе и перспективах его применения // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал. 2020. № 4. С. 78-86.

References

1. Amanov M.Je., Akmuradova K.K. Kaspij i global'nye vyzovy: materialy Mezhdunarodnoj nauchno-prakticheskoj konferencii. Astrahan': AGU. 2022, pp. 9-14.

2. Baeva, L.V., Romanova A.P. Cultura. International Journal of Philosophy of Culture and Axiology. 2015. Vol. 12. No 1. pp. 159-172.

3. Han J., Kamber M., Pei J. USA: Morgan Kaufmann Publishers is an imprint of Elsevier. 2012, p. 703.

4. Kozitsin I.V., Chkhartishvili A.G., Marchenko A.M., Norkin D. O., Osipov S. D., Uteshev I.A., Goiko V.L., Palkin R.V., Myagkov M.G. Mathematical Models and Computer Simulations. 2020. Vol. 12. No 2. pp. 185-194.

5. Chudinov S. I., Serbina G. N., Mundrievskaja Ju. O. Monitoring obshhestvennogo mnenija: jekonomicheskie i social'nye peremeny. 2021. № 4(164). pp. 363-383.

6. Bogdanov A. L., Dulja I. S. Vestnik Tomskogo gosudarstvennogo universiteta. Jekonomika. 2019. № 47. pp. 220-241.

7. Boldyreva A., Alexandrov M., Koshulko O., Sobolevskiy O. Advances in Computational Intelligence. 2016. Vol. 1. No 15. pp. 290-301.

8. Harkovchuk, A. Je., Korzun D. Zh., Cifrovye tehnologii v obrazovanii, nauke, obshhestve: Materialy XIII vserossijskoj nauchno-prakticheskoj konferencii, Petrozavodsk, Petrozavodsk: Petrozavodskij gosudarstvennyj universitet. 2019, pp. 199-202.

9. Spravochnik API, VK dlja razrabotchikov [API reference, VK for developers]. URL: dev.vk.com/reference/roadmap (accessed : 06/06/2022)

10. Zimova N. S., Fomin E. V., Smagina A. A. Nauchnyj rezul'tat. Sociologija i upravlenie. 2020. T. 6. № 2. pp. 159-171.

11. Villar G., Arciuli J., Paterson H. Psychiatry, Psychology and Law. 2013. Vol. 20 (4). pp. 504-518.

12. Babenko L. G. Izvestija Ural'skogo federal'nogo universiteta. Serija 2: Gumanitarnye nauki. 2020, T.22. № 3(200). pp. 264-278.

13. Kotelnikova A., Kotelnikov E. Communications in Computer and Information Science. 2019, Vol. 1119. pp. 18-32.

14. Karabak I. I., Zorin K. A. Prikaspijskij zhurnal: upravlenie i vysokie tehnologii. 2022. № 1(57). pp. 9-17.

15. Shchekotin E., Myagkov M., Goiko V., Kashpur V. Administratie si Management Public. 2021. Vol. 2021. No 36. pp. 25-48.

16. Bird S., Loper E. ACL. – Barcelona: Association for Computational Linguistics. 2004, pp. 214-217.

17. Demšar J., Curk T., Erjavec A., Gorup Č., Hočevar T., Milutinovič M., Možina M., Polajnar M., Toplak M., Starič A., Štajdohar M., Umek L., Žagar L., Žbontar J., Žitnik M., Zupan B., Journal of Machine Learning Research. 2013. № 14. pp.2349-2353.

18. Kiselev M. V., Arseniev S. B., Ananyan S. M., Lecture Notes in Computer Science. 1998. Vol. 1510. pp. 352-360.

19. Berthold M.R., Cebron N., Dill F., Fatta G.D., Gabriel T.R., Georg F., Meinl T., Ohl P., Sieb C., Wiswedel B. Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. 2008, pp. 319-326.

20. Chapkovski, P. Munich Personal RePEc Archive. 2022, № 111980. pp. 1-39.

21. Gurin A. A. Innovacii. Nauka. Obrazovanie. 2020. № 19. pp. 471-477.

22. Majorova, E. V. Social'nye i gumanitarnye nauki. Otechestvennaja i zarubezhnaja literatura. Serija 6: Jazykoznanie. Referativnyj zhurnal, 2020, № 4. pp. 78-86.