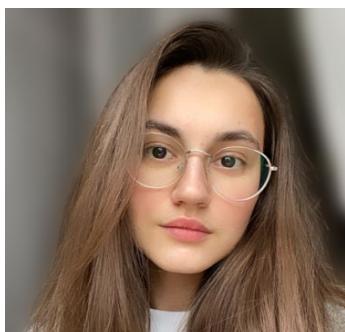


# ЧЕЛОВЕЧЕСКИЙ ФАКТОР

УДК 311.2

## ИЗМЕРЕНИЕ СОЦИАЛЬНОГО САМОЧУВСТВИЯ ГОРОЖАН ПО ЦИФРОВЫМ МАРКЕРАМ: АПРОБАЦИЯ МЕТОДИКИ



**Дымова Полина Игоревна**

Тульский государственный университет, Тула, Россия;  
Dymova.pi@gmail.com



**Домбровская Анна Юрьевна**

доктор социологических наук, профессор кафедры политологии,  
директор Центра политических исследований Финансового университе-  
тата при Правительстве РФ, Москва, Россия; an-doc@yandex.ru

**Аннотация.** В статье на основе данных городских сообществ социальной сети ВКонтакте рассматривается способ измерения социального самочувствия с помощью цифровых маркеров. В рамках исследования использованы публикации онлайн-групп Тулы и Тульской области за 2023 г. Полученные данные обработаны с помощью языка программирования Python 3 и дополнительных библиотек (pandas, spaCy, dostoevsky). В результате получена информация о принадлежности текстовых документов к категориям социального самочувствия, а также определена их тональность. В заключение рассмотрены ограничения используемого метода и возможные пути их пре-одоления для совершенствования методики измерения социального самочувствия с использованием методов киберметрического анализа.

**Ключевые слова:** социальное самочувствие; городское сообщество; онлайн-сообщество; ки-  
берметрия; цифровые маркеры.

**Для цитирования:** Дымова П.И., Домбровская А.Ю. Измерение социального самочувствия горожан по цифровым маркерам // Социальные новации и социальные науки. – 2024. – № 3. – С. 94–107.

URL: <https://sns-journal.ru/ru/archive/>

DOI: 10.31249/snsn/2024.03.07

Рукопись поступила 30.05.2024

Принята к печати 22.08.2024

## **Введение**

Социальное самочувствие является одним из многочисленных индикаторов состояния общества, которое регулярно измеряется как крупными центрами мониторинга общественного мнения, так и отдельными исследовательскими группами. В научной литературе существует большое количество работ, посвященных попыткам изучить социальное самочувствие в стране или определенном ее субъекте, в социально-демографических или профессиональных группах.

Анализ имеющихся исследований социального самочувствия позволяет отследить некоторые проблемы, связанные с его изучением. Во-первых, в социологической науке не существует единого определения и набора индикаторов для измерения вышеупомянутого показателя. В общем смысле, под социальным самочувствием подразумевается субъективная оценка индивидом собственного состояния и положения в обществе: как есть сейчас и на что есть надежды в будущем. Для измерения самочувствия используется различный набор объективных и субъективных показателей. В качестве примера объективных показателей выделяются: уровень материального положения и состояние здоровья индивида [Симонович, 2003, с. 5–6], реальный социальный статус в различных сферах жизнедеятельности [Михайлова, 2010, с. 47] и уровень властных полномочий [Душацкий, 2004, с. 67–69]. Примерами субъективных показателей являются: уверенность индивида в завтрашнем дне [Грачев, Русалинова, 2007, с. 9–10], удовлетворенность реализацией жизненной стратегии [Петрова, 2000, с. 53], восприятие, оценка и отношение к своему социальному статусу [Михайлова, 2010, с. 47] и др.

Приведенные примеры показателей, как объективных, так и субъективных, безусловно позволяют оценить социальное самочувствие индивида в частности и конкретной группы в целом, однако мало говорят о причинах той или иной оценки. Один респондент, обладая крепким здоровьем и имея высокий уровень зарплаты, может ощущать неуверенность в завтрашнем дне и быть неудовлетворенным тем, как складывается его жизнь. Другой, имеющий проблемы со здоровьем и меньшее количество материальных благ, может оценивать свою жизнь как достаточно счастливую и не тревожиться о будущем. Также стоит отметить, что большая часть измеряемых показателей, особенно субъективных, нуждается в дополнительной операционализации и может быть интерпретирована каждым участником исследования по-своему [Рогозин, 2007, с. 111].

Вторая проблема, связанная с исследованиями социального самочувствия, проявляется в существовании большого количества синонимичных как по форме, так и по содержанию социологических понятий: социального благополучия и неблагополучия, качества и уровня жизни, социального настроения и др. С одной стороны, это позволяет обнаружить интересные для исследователя

методы и подходы, которые можно транслировать на область изучения социального самочувствия, с другой – возвращает к первой, ранее озвученной трудности.

И, наконец, третья проблема – это используемый для оценки социального самочувствия метод исследования. Самым часто используемым инструментом является опрос, проводящийся лично или при помощи телекоммуникационных средств. Использование опросной методики не только не позволяет выяснить, что стоит за ответом, который дал респондент, но и не исключает влияние на ответы фигуры интервьюера.

Если разрешение первых двух обозначенных трудностей в данный момент не представляется возможным, то с точки зрения выбора метода исследования интерес представляет анализ неспровоцированных высказываний в социальных сетях по цифровым маркерам [Развитие методологии и методики интеллектуального поиска ..., 2017, с. 83–84]. Преимущество киберметрических методов для анализа социальных и политических процессов состоит не только в обилии данных, оставляемых пользователями социальных сетей, но и в наличии готовых инструментов для мониторинга социальных медиа.

Изучение активности пользователей в онлайн-сообществах широко используется в государственном управлении. В 2020 г. в каждом субъекте Российской Федерации появился Центр управления регионом (далее – ЦУР) для оперативного реагирования на проблемы градоустройства. По словам Анатолия Курманова, администратора развития Платформы обратной связи, главной идеей ЦУР является предвидение потребностей граждан. Благодаря анализу обращений аудитории сетевого сообщества появляется возможность проведения глубокой аналитики по регионам, создания «тепловых карт», выявления типовых проблем, унификации и оптимизации способов их решения. Подобные аналитические данные в полной мере могут служить опорой для принятия решений органами региональной власти и создания программ развития регионов с учетом наиболее важных для граждан направлений городского развития [Большакова, Климова, 2022, с. 394].

Удовлетворенность инфраструктурой является значимым фактором в измерении социального самочувствия городского населения [Цветкова, 2017, с. 115]. Деградация городской инфраструктуры, по словам автора, приводит к постепенному снижению социального самочувствия. Продолжительная неудовлетворенность жизнью и общий негативный фон в рамках конкретного города и региона способны привести к оттоку жителей, в попытке найти более подходящее и комфортное для жизни место, а также к снижению психоэмоционального и физического здоровья, росту преступности и распространения девиантного поведения.

Как было обозначено ранее, в социологической литературе существует достаточно большое количество синонимичных социальному самочувствию понятий. В работе Е.В. Щекотина, М.Г. Мягкова и др. используются данные онлайн-активности пользователей социальной сети

ВКонтакте для расчета индекса субъективного (не)благополучия в 43 субъектах Российской Федерации [Субъективная оценка (не)благополучия..., 2020, с. 92–93].

### **Описание методологии и метода**

В рамках проведенного исследования для операционализации понятия социального самочувствия был выбран подход, разработанный коллективом авторов: Е.И. Головаходой, Н.В. Паниной и А.П. Горбачик. Суть данного подхода состоит в измерении интегрального показателя социального самочувствия на основании суждений о достаточности социальных благ в 11 сферах жизнедеятельности индивида [Головахода, 1998, с. 49]. В оригинальном исследовании для оценки достаточности каждого из компонентов социального самочувствия использовалась шкала достаточности, включающая в себя оценки «не хватает», «трудно сказать, хватает или нет», «хватает», «не интересует».

Стоит также упомянуть, что измерение интегрального индекса социального самочувствия предусматривает проведение анкетирования, в расширенном или сокращенном формате. Так как в рамках проведенного исследования не предполагалось применение опросной методики, было принято решение судить о достаточности того или иного компонента социального самочувствия, опираясь на тональность анализируемого документа. Положительная тональность говорит о достаточности, негативная – о нехватке социальных благ, а нейтральная – приравнивается к ответам «трудно сказать» и «не интересует», которые в оригинальной методике также были объединены и оценивались одинаковым количеством баллов [Головахода, Панина, Горбачик, 1998, с. 50–51].

Следующим этапом исследования стал отбор онлайн-сообществ для сбора документов в социальной сети ВКонтакте. В результате ручного отбора, проводимого в марте 2024 г., был сформирован список из 28 сообществ Тулы и Тульской области, из которого далее были исключены две группы, так как не подходили под тип «городское сообщество». Сообщества определялись по следующим критериям:

1. Наличие достаточно большого числа подписчиков в соотношении с населением муниципального образования по данным на 2023 г.
2. Активность, наличие недавно опубликованных записей на момент отбора.

Внимание при отборе сообществ было, в первую очередь, направлено на поиск онлайн-групп с типом «городское сообщество». В отличие от новостных групп, городские онлайн-сообщества содержат больше эмоционально-окрашенных публикаций. Ниже приведена таблица с распределением сообществ относительно административно-территориальных единиц Тульской области.

**Число сообществ, принадлежащих административно-территориальным единицам  
Тульской области**

<b>Название административно-территориальной единицы</b>	<b>Количество сообществ</b>
г. Тула	10
г. Новомосковск	2
г. Ясногорск	2
г. Алексин	1
г. Кимовск	1
г. Ефремов	1
г. Киреевск	1
г. Белев	1
г. Узловая	1
г. Щекино	1
Не определен	5

Следующим этапом стала разработка программного кода на языке программирования Python 3 для автоматизации выгрузки информации о сообществах, их публикациях, а также приведения текста документов к нормальному виду для проведения дальнейшего анализа.

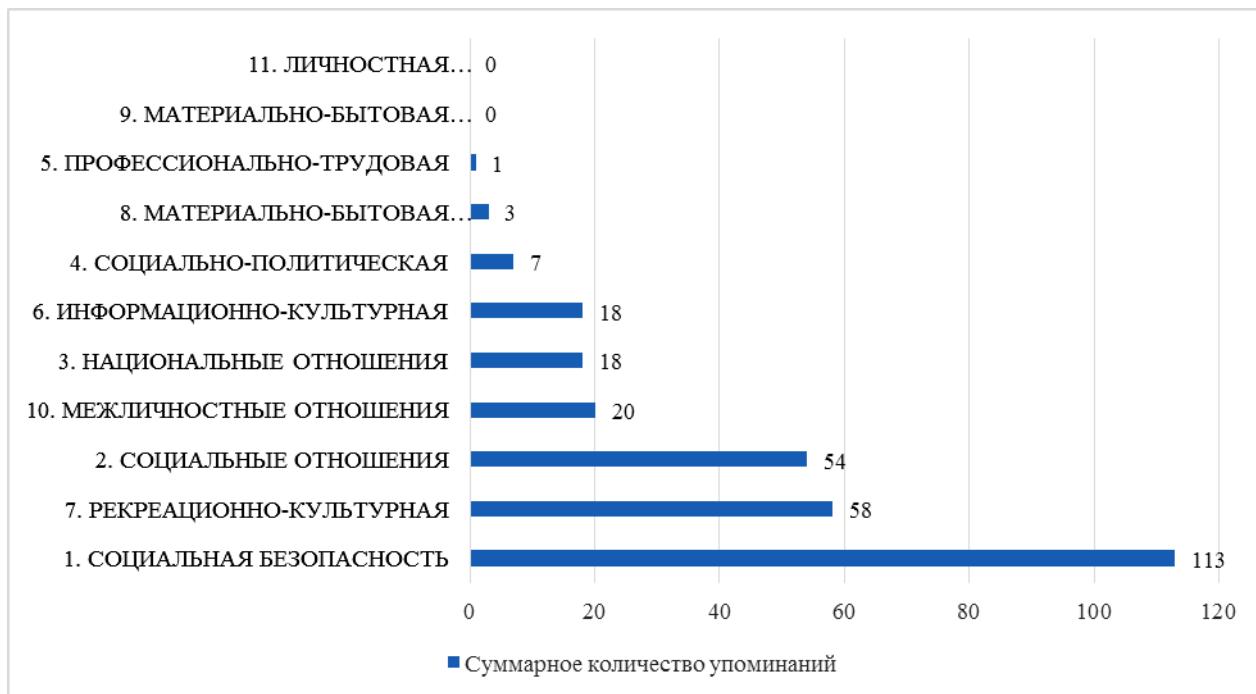
Данные об отобранных сообществах социальной сети ВКонтакте были получены с помощью методов VK API, обращение к которым осуществлялось посредством модуля requests. Для получения таких характеристик, как количество участников сообщества, идентификатор, город и др., использовался метод groups.getById. Для формирования списка публикаций по каждому отобранныму сообществу использовался метод wall.get, позволяющий получить: дату и время публикации, текст и вложения (фото- и видеоматериалы), идентификатор автора, количество просмотров, комментариев, лайков и репостов и многое другое. Полную информацию о возвращаемых полях можно прочитать в документации на официальном сайте социальной сети.

Общий массив данных, выгруженный из отобранных сообществ, составил 472 855 публикаций: самый поздний документ датирован 14 апреля 2024 г., самый ранний – 3 марта 2009 г. Для вычисления индекса социального самочувствия были выбраны публикации за 2023 г., общее число которых составило 52 810 записей.

В рамках предварительного анализа из всего массива данных было отобрано 1197 документов, опубликованных в период с 2024 по 2021 год. Задачи предварительного анализа состояли в следующем.

1. Соотнести текст каждой публикации с категорией социального самочувствия в рамках выбранного подхода.
2. Определить две категории, публикации из которых встречаются чаще всего.
3. Составить словари маркеров социального самочувствия для конкретной категории и дополнить их собственными ключевыми словами при необходимости.

В результате, среди исследуемых документов: 581 публикация маркирована как «мусорная» в связи с тем, что содержит неподходящие высказывания, 268 публикаций – новости, 25 – объявления о поиске работы, 14 – рекламные тексты, 13 – объявления о поиске пропавших, четыре – дубли тех записей, что встречались ранее. Оставшиеся документы распределились по категориям социального самочувствия следующим образом (рис. 1).



**Рис. 1. Распределение публикаций по категориям социального самочувствия\***

\* В интегральном индексе социального самочувствия есть две материально-бытовые категории: 1-го и 2-го уровней. 1-й уровень включает в себя: хорошее жилье, необходимую одежду, возможность приобретать необходимые продукты и необходимую мебель, 2-й уровень: автомобиль, возможность питаться в соответствии со своими вкусами, модную и красивую одежду, садовый (приусадебный) участок.

Несмотря на то что по количеству публикаций на втором месте оказалась рекреационно-культурная компонента социального самочувствия, было принято решение проводить дальнейший анализ по категориям «социальная безопасность» и «социальные отношения», так как документы, попавшие в категорию «рекреационно-культурная», носили нейтральный характер и скорее представляли собой анонсы культурных мероприятий.

Для поиска цифровых маркеров в целевом массиве документов текст публикаций был подвергнут стандартным процедурам обработки естественного языка, которые включали в себя:

1. Избавление текста от знаков препинания, переносов, эмоджи и любых других дополнительных символов.
2. Приведение текста к нижнему регистру.
3. Токенизацию и лемматизацию.

Процедура токенизации представляет собой разделение текста на отдельные слова и знаки пунктуации, если прежде он не был от них очищен. Процедура лемматизации – это приведение

слова к его начальной морфологической форме. Подобная обработка упрощает поиск цифровых маркеров в текстовых документах, так как при составлении словаря пропадает необходимость использовать разные формы одних и тех же слов, например склонения существительных, чтобы учесть все возможные вариации словоформ.

Процедуру лемматизации часто упоминают либо вместе, либо вместо процедуры стемминга. В отличие от лемматизации, в стемминге базовая форма слова выделяется за счет удаления суффиксов и окончаний. Ниже (табл. 2) приведен пример, иллюстрирующий различия процедур стемминга и лемматизации на одной из отобранных публикаций.

Таблица 2

#### **Сравнение результатов процедур стемминга и лемматизации**

Очищенный текст	Результат стемминга	Результат лемматизации
масочный режим в учреждениях здравоохранения тульской области введен с сегодняшнего дня для посетителей и медицинского персонала	масочн реж в учрежден здравоохран тульск област введ с сегодняшн дня для посетител и медицинск персона	масочный режим в учреждение здравоохранение тульский область ввести с сегодняшний день для посетитель и медицинский персонал

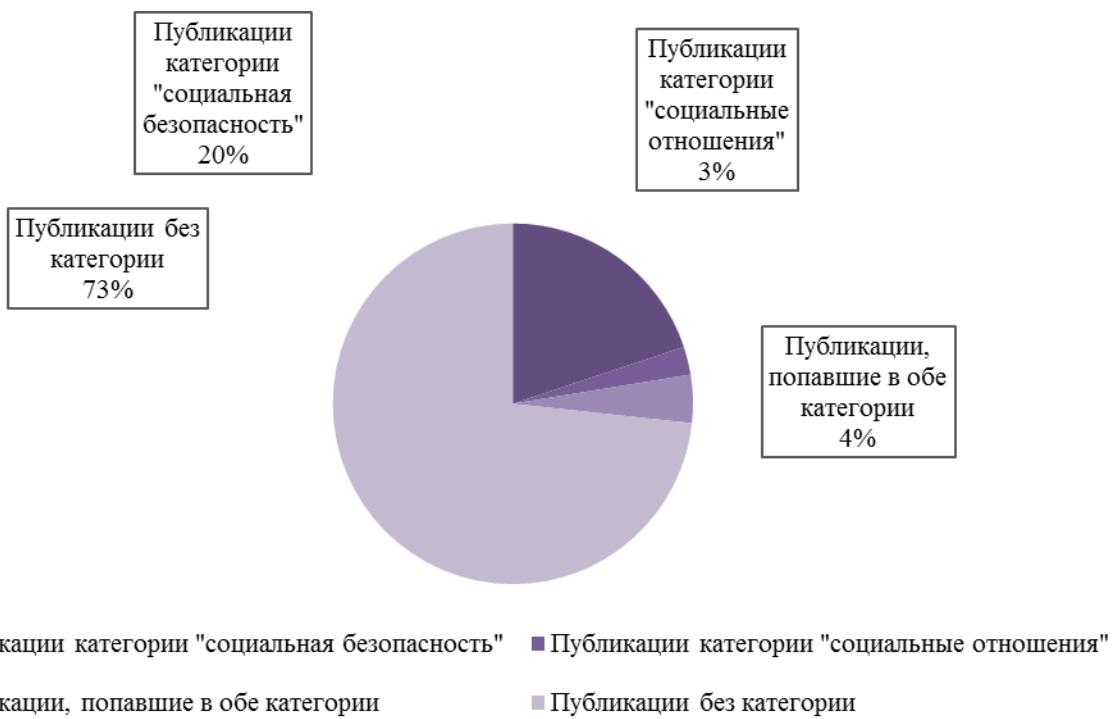
В данной работе текст публикаций, полученный в результате лемматизации, представляется наиболее удобным для поиска маркеров, и именно он в дальнейшем используется для отнесения документа к той или иной категории социального самочувствия.

В результате поиска маркеров из составленных для категорий «социальная безопасность» и «социальные отношения» словарей данные распределились следующим образом (рис. 2): 20% публикаций за 2023 г. попали в категорию «социальная безопасность», 3% – в категорию «социальные отношения», 4% относятся к обеим категориям, а 73% не классифицированы. Публикации, не попавшие ни в одну из категорий социального самочувствия, могут оказаться как «мусорными данными», так и теми, которые, теоретически, могли попасть в девять оставшихся категорий.

Совмещение категорий в рамках одного документа может быть вызвано тем, что публикация содержит маркеры, относящиеся к обеим категориям социального самочувствия. Далее этот набор данных не будет исключаться из анализа.

Для определения тональности сообщений использовалась библиотека *dostoevsky*, предназначенная для сентимент-анализа русскоязычных текстов. Модель, использующаяся в данной библиотеке, обучена на наборе данных RuSentiment, состоящем из открытых публикаций социальной сети ВКонтакте [RuSentiment..., 2018]. В результате обработки текста модель возвращает вероятность, с которой текст относится к той или иной тональности. Разработчики модели выделяют пять категорий тональности: позитивную (positive), негативную (negative), нейтральную (neutral), форму речи (speech) и пропуск (skip). Пропускаются моделью те публикации, в которых трудно

определить тональность, а также публикации-шутки, бессмысленные высказывания или тексты не на русском языке.

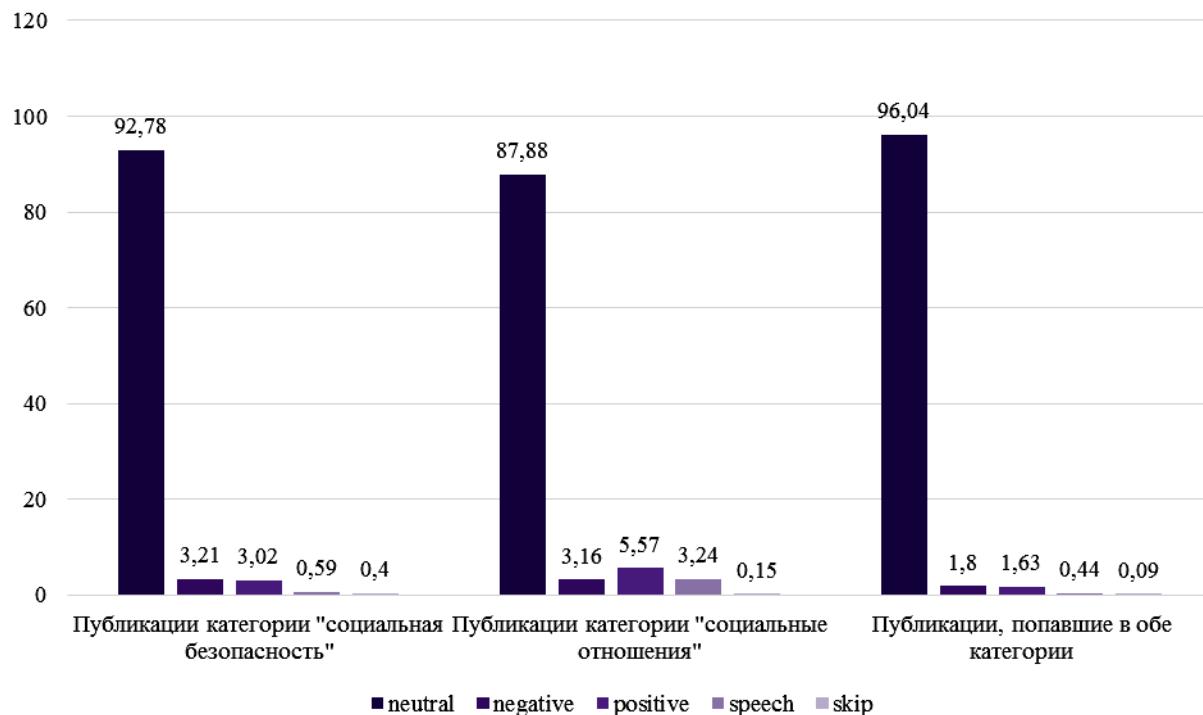


**Рис. 2. Распределение публикаций по категориям социального самочувствия**

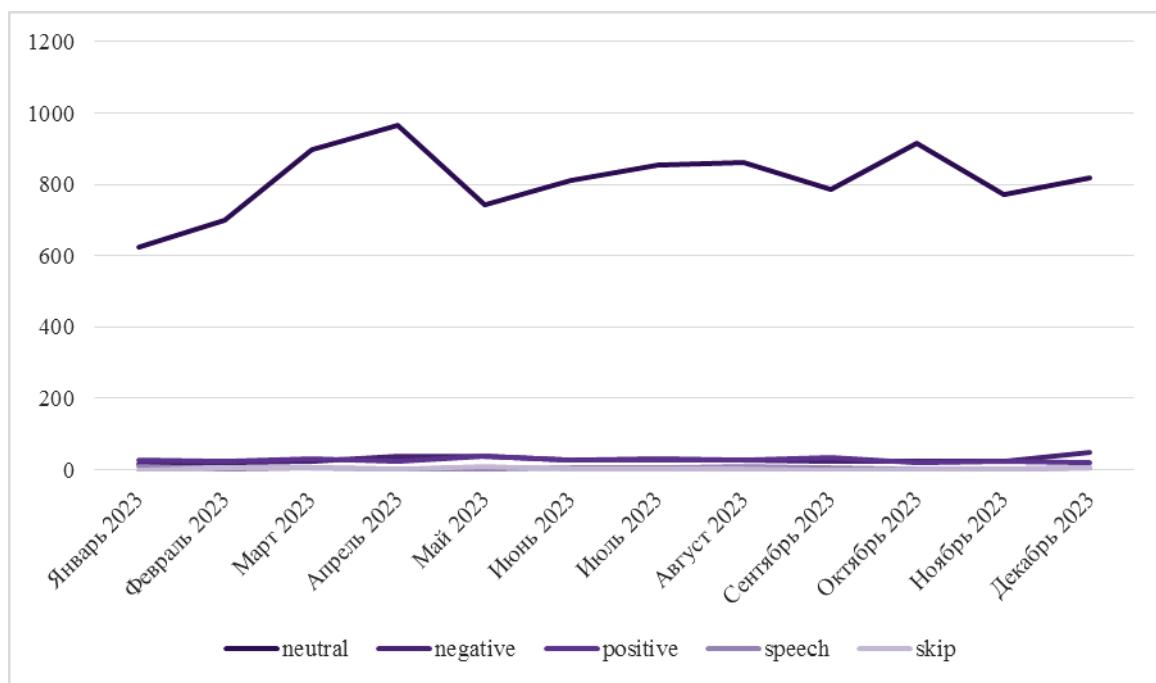
При обработке текста моделью необходимо указать параметр  $k$ , принимающий значения от 1 до 5 и показывающий «уверенность» модели в предсказанной тональности. В рамках данного исследования использовалось  $k=1$ , чтобы получить одно, наиболее вероятное, определение тональности. Увеличение  $k$  при расчетах не ведет к тому, что вероятность тональности пересчитывается. При  $k=2$  в результате обработки текста выдается второй по вероятности вариант тональности, аналогично при увеличении  $k$  появляется третий, четвертый и пятый.

В результате обработки данных тональность подавляющего большинства публикаций была определена как нейтральная. На втором месте, но с большим отрывом, находятся публикации негативной тональности, за исключением текстовых документов категории «социальные отношения». Там, следующим за нейтральной тональностью, следует речевое высказывание (рис. 3).

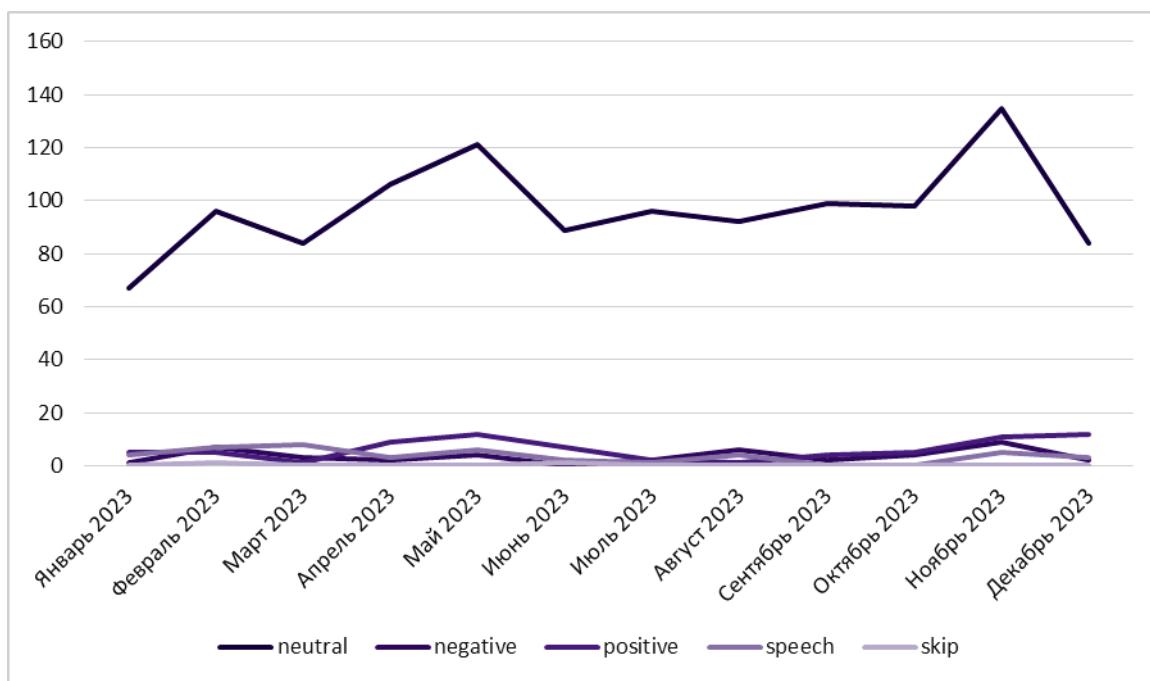
Далее были составлены распределения количества публикаций и их тональности по месяцам 2023 г. Всплески активности в явном виде хорошо определяются на публикациях с нейтральной тональностью. Так, большая часть текстовых документов категории «социальная безопасность» приходится на апрель и октябрь 2023 г. Примерно в то же время наблюдается увеличение числа публикаций категории «социальные отношения»: май и ноябрь 2023 г. Для публикаций, попавших в обе категории, пик активности приходится на март 2023 г.



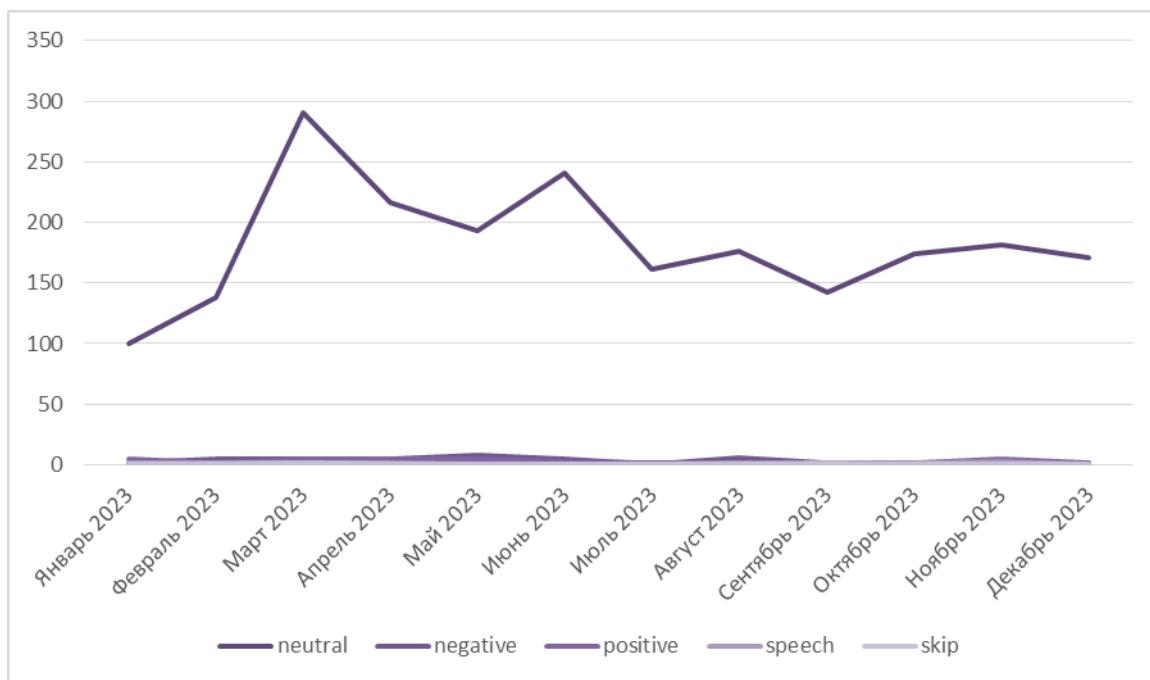
**Рис. 3. Распределение публикаций по тональности внутри категории (относительные показатели)**



**Рис. 4. Распределение тональности по месяцам для публикаций категории «социальная безопасность»**



**Рисунок 5. Распределение тональности по месяцам для публикаций категории «социальные отношения»**



**Рис. 6. Распределение тональности по месяцам для публикаций, попавших в обе категории**

## **Результаты и обсуждение**

В результате апробации методики обнаружены некоторые проблемы, требующие решения для дальнейшего совершенствования киберметрических процедур в анализе социального самочувствия по цифровым маркерам.

Первая проблема касается неоднозначной классификации текстового документа относительно той или иной категории социального самочувствия. Даже при наличии непересекающихся слов-вариантов маркеров возможна ситуация, когда текстовый документ содержит ключевые слова нескольких из них. Возможно, предположение о том, что каждый текстовый документ можно однозначно отнести к определенной категории социального самочувствия, изначально неверно, и более правильным является подход, определяющий вероятность принадлежности публикации к каждой группе из имеющегося набора.

Вторая проблема связана с определением тональности. Подавляющее большинство документов были охарактеризованы как нейтральные, что не позволяет однозначно определить оценку социального самочувствия в рамках исследуемых категорий. Ранее было отмечено, что в рамках используемого подхода достаточность социальных благ определялась значениями по шкале: «хватает», «не хватает», «трудно сказать, хватает или нет», «не интересует». Предполагалось, что оценки «хватает» и «не хватает» соответствуют позитивной и негативной тональностям, тогда как нейтральная относится к оставшимся двум. Здесь возникает противоречие: с одной стороны, попадание большого количества публикаций в категории «социальная безопасность» и «социальные отношения» на этапе предварительного анализа говорит о важности именно этих тем в повседневной жизни горожан, с другой – трудно определить явную эмоциональную окраску высказываний относительно них. Причиной сложившейся ситуации может являться особенность модели, которая, несмотря на значительную точность, не улавливает небольшие эмоциональные оттенки текста. Попробовать решить эту проблему можно с помощью разработки собственной модели, обученной на материалах, которые используются для анализа социального самочувствия и синонимичных ему индикаторов. Менее радикальным, но интересным представляется другой вариант. Модель для определения тональности, используемая в данной работе, а также модели, используемые в других, предполагают исключение из анализа эмоджи, которыми люди часто пользуются как в личной переписке с близкими, так и в публичном пространстве онлайн-сообщества. Для языка программирования Python существуют библиотеки, позволяющие перевести эмоджи в слово, которое оно обозначает. Так как эмоджи обозначают различные эмоциональные реакции, прямое указание в тексте на смех, радость, гнев или грусть способно повлиять на определение тональности текста. Еще одним расширением методики видится использование комментариев под публикацией, при их наличии. Сейчас они исключены из анализа как единицы текста, и учитываются скорее как показатель активности участников сообщества [Субъективная оценка (не)благополучия..., 2020, с. 92–93].

Третья проблема – необходимость дополнительного исследования по результатам временно-го распределения публикаций. Даже в документах с нейтральной тональностью наблюдаются всплески и падения активности и, как следствие, увеличение и снижение числа записей в онлайн-

сообществах. В рамках проведенного исследования не определены события, повлиявшие на колебания публикационной активности. Необходимо обратить на это отдельное внимание и отследить, какие внутри- и внешнеполитические события на это влияют.

В качестве еще одной проблемы стоит выделить техническое ограничение, связанное со структурой данных о публикациях онлайн-сообщества, получаемых из социальной сети ВКонтакте. Это ограничение связано с особенностями определения автора публикации внутри группы. Большая часть постов публикуется от имени сообщества, а непосредственный автор предложенной записи фиксируется только в том случае, если он не пожелал сохранить анонимность. Среди всех отобранных за 2023 г. данных только 9628 (18,23%) содержат данные автора, оставшиеся 43 182 (81,77%) – опубликованы анонимно. Это затрудняет составление демографического портрета авторов. Однако, существует возможность проанализировать половозрастные характеристики участников сообщества в целом.

### **Заключение**

В нынешней исследовательской реальности наблюдается существенная диверсификация стратегий и методов сбора и анализа эмпирической информации. Все чаще практикующие эмпиреки обращаются к возможностям автоматизации процесса аккумулирования и обработки базы данных. Социологические исследования – одна из первых областей социогуманитарного знания, в которой все активнее применяется автоматизация, программирование, алгоритмы машинного обучения в изучении цифровых следов социальных настроений, взглядов, ценностей, мнений, в конечном итоге, социального самочувствия. В настоящей статье представлен конкретный эмпирический кейс, показавший возможности и ограничения программного обеспечения в измерении социального самочувствия тульских жителей по его онлайн-сетевым маркерам. Среди ключевых достоинств подобной и других форматов автоматизации сбора и обработки информации – возможность сплошного анализа в пределах необходимого временного интервала и социально-медийной платформы, экономия времени, высокая точность получаемых данных, отсутствие негативного воздействия человеческих факторов (усталость, неточность, ошибки при сборе данных). Ограничениями автоматизированных форматов анализа социальных процессов остается пока неизбежное упрощение в ходе верификации изучаемых признаков, иногда – примитивизация индикаторов измеряемых признаков в связи с необходимостью создать алгоритм извлечения релевантной информации. Перспективы решения этих проблем видятся в совершенствовании методики сбора и обработки данных с применением машинного обучения (к примеру, методики «обучения с учителем»).

### Список литературы

1. Больщакова К.Ю., Климова А.В. Центры управления регионом как новая форма управленческой деятельности // Вестник Российской университета дружбы народов. Серия Государственное и муниципальное управление. – 2022. – Т. 9, № 4. – С. 394.
2. Головаха Е.И., Панина Н.В., Горбачик А.П. Измерение социального самочувствия: тест ИИСС // Социология: методология, методы, математическое моделирование (Социология:4 М). – 1998. – № 10. – С. 45–71. – URL: [https://www.fnisc.ru/index.php?page\\_id=2384&id=3911&l=&j=6&base=ojs3](https://www.fnisc.ru/index.php?page_id=2384&id=3911&l=&j=6&base=ojs3) (дата обращения 10.01.2024)
3. Грачев А.А., Русалинова А.А. Социальное самочувствие человека в организации // Известия РГПУ им. А.И. Герцена. – 2007. – № 30. – URL: <https://cyberleninka.ru/article/n/sotsialnoe-samochuvstvie-cheloveka-v-organizatsii> (дата обращения 01.03.2024).
4. Душацкий Л.Е. Материально-властные ресурсы Россиян в самооценке и социальном самочувствии // Социологические исследования. – 2004. – № 4. – С. 64–70.
5. Михайлова Л.И. Социальное самочувствие и восприятие будущего россиянами // Социологические исследования. – 2010. – № 3. – С. 45–50.
6. Петрова Л.Е. Социальное самочувствие молодежи // Социологические исследования. – 2000. – № 12. – С. 50–55.
7. Развитие методологии и методики интеллектуального поиска цифровых маркеров политических процессов в социальных медиа / Бродовская Е.В., Домбровская А.Ю., Карзубов Д.Н. [и др.] // Мониторинг общественного мнения: Экономические и социальные перемены. – 2017. – № 5. – С. 79–104.
8. Рогозин Д. Тестирование вопросов о социальном самочувствии // Социальная реальность. – 2007. – № 2. – С. 97–113.
9. Симонович Н.Е. Социальное самочувствие людей и технологии его исследования в современной России: автореф. дис.... д-ра психол. наук. – Москва, 2003. – 37 с.
10. Субъективная оценка (не) благополучия населения регионов РФ на основе данных социальных сетей / Щекотин Е.В., Мягков М.Г., Гойко В.Л., Кашпур В.В., Коварж Г.Ю. // Мониторинг общественного мнения: Экономические и социальные перемены. – 2020. – № 1. С. 78–116. (дата обращения 20.04.2024).
11. Цветкова И.В. Факторы социального самочувствия горожан // КНЖ. – 2017. – № 1 (18). – URL: <https://cyberleninka.ru/article/n/faktory-sotsialnogo-samochuvstviya-gorozhan> (дата обращения 20.04.2024).
12. RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian / Rogers A., Romanov A., Rumshisky A., Volkova S., Gronas M., Gribov A. // Proceedings of the 27 th International Conference on Computational Linguistics. – Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. – P. 755–763.

## MEASURING SOCIAL WELL-BEING OF CITY DWELLERS BY DIGITAL MARKERS: TESTING THE METHODOLOGY

**Polina Dymova**

Master of Sociology, Tula State University, Tula, Russia; Dymova.pi@gmail.com

**Anna Dombrovskaya**

Doctor of Sociology, Professor of the Department of Political Science, Director of the Center for Political Studies of the Financial University under the Government of the Russian Federation, Moscow, Russia; an-doc@yandex.ru

***Abstract.** The article examines a method for measuring subjective well-being using digital markers based on data from online city communities of the VKontakte social network. Publications from online groups in Tula and the Tula region for the year 2023 were used in the study. The collected data was processed using the Python 3 programming language and additional libraries (pandas, spaCy, dostoevsky). As a result, information was obtained about the categorization of textual documents into subjective well-being categories, as well as their sentiment. In conclusion, the limitations of the used method are discussed, along with possible ways to overcome them to improve the technique for measuring subjective well-being using cybermetric analysis methods.*

**Keywords:** social well-being, city community, online community, cybermetrics, digital markers.

**For citation:** Dymova P.I., Dombrovskaya A.Yu. Measuring social well-being of city dwellers by digital markers: testing the methodology // Social Novelties and Social Sciences. – 2024. – No 3. – P. 94–107.

URL: <https://sns-journal.ru/ru/archive/> DOI:

10.31249/snsn/2024.03.07