## 2.2 РЕЗУЛЬТАТЫ АНАЛИЗА ТЕКСТОВ ПОСТОВ ПАЦИЕНТСКИХ СООБЩЕСТВ

В предыдущем параграфе на примере небольшой группы врачей и пациентов показано, что врачи и пациенты не разделяют «пациентские организации» и «пациентские сообщества», не выявлено разницы и в отношении к разным типам организаций. Найдем ли мы разницу между текстами пациентских организаций и неформальных пациентских сообществ?

Ценность речи в сетевой экономике внимания все возрастает[[1]](#endnote-1), то, как сообщества организуют самопрезентацию, может дать интересные инсайты. Анализ текстов - один из самых традиционных видов аналитики для исследований Internet Studies. За годы использования специалистами выделенные целые группы методов для решения исследовательских задач.

Цель исследования состояла в том, чтобы выявить разницу в текстах постов пациентских организаций и пациентских сообществ. В данной части исследования задачей является выделить различия между коммерческими и некоммерческими аккаунтами, которые могут быть отражены в текстах сообществ.

Следует учитывать ограничения этой базы сравнения: несмотря на наглядность и обоснованность, количество текстов в среднем у аккаунтов сообществ и организаций – разное.

Для сравнения выделяются следующие характеристики датасетов: дескриптивные текстовые статистики, выделение именованных сущностей, G2 статистика для визуализации расхождения в специфичности слов для двух групп текстов.

Из-за политики платформ социальных сетей, наиболее подходящей для исследования выбрана VK. Таким образом, объектом настоящего исследования являются все русскоязычные сообщества VK (не заблокированные и не удаленные администраторами сети), которые были зарегистрированы на момент формирования базы.

За генеральную совокупность принимаем все сообщества VK, по данным Allsocial.ru на апрель 2022 – это 551 951 группы и паблика. Ко времени исследования VK имеет два типа сообществ – это паблики (Public) и группы (Group). В данном исследовании разнице не придается значения, оба типа рассматриваются вместе.

В социальных сетях видно несколько явных тенденций:

- Сбор средств на лечение. Эта тенденция не специфична для определенных заболеваний, есть примеры обращения за помощью вследствие получения травм и др.

- Пациентские чаты – самая закрытая и «личная» группа сообществ. Некоторые НКО создаются участниками таких чатов, чаты могут создаваться самими НКО.

На момент сбора информации (апрель 2022 г.) простой поиск по сообществам выдает 1 415 сообщества (0.26% от общего числа сообществ) на запрос «редкие заболевания». Из собранных 121 сообщество – закрытые.

Для анализа не отбирались сообщества, которые посвящены нетрадиционным практикам лечения, ветеринарные клиники, группы посвященные зависимостям, протестные сообщества (противники абортов и др.), целая группа сообществ о вспышке менингита в Калининграде (2010-е), дубли сообществ, попавшие в выборку, рекламные аккаунты препаратов. В подборку попало большое количество групп о болезни Эбола, их контент также не использован в анализе.

За целевые сообщества принимаем только некоммерческие неформальные сообщества полностью посвященные или затрагивающие тему редких заболеваний[[2]](#footnote-1), сюда вошли: студенческие научные общества, группы взаимопомощи, тематические сообщества, группы по сбору средств на лечение и некоторые другие. Всего целевых сообществ: 344. Общее число участников, включенных в данные сообщества, — 94 862 подписчиков. Среднее количество участников: 295 человек.

Целевые сообщества – группы с 1000 и более участников, часто «закрытые» (чтобы видеть «стену», нужно вступить в сообщество). Сообщества модерируются, но при вступлении не нужно заполнять анкету, то есть можно сделать вывод, что администратор делает внешний осмотр аккаунтов вступающих и только на основании внешних данных принимает решение о допуске аккаунта в сообщество.

Наибольшее количество подписчиков – у больших благотворительных организаций: их посты собирают больше всего лайков.

Во всем датасете - всего два рекламных поста, оба релевантные тематике сообществ, посвящены лечению.

Орфография и пунктуация в названиях групп сохраняется. По названию можно сделать вывод, что решение проблем со здоровьем, особенно детским, мыслится как коллективное действие. Преобладают формы обращения в третьем лице множественного числа, такие как «Поможем …» , «вместе», «давайте».

Часть из этих сообществ посвящена помощи отдельным пациентам, некоторые такие группы со «срочными сборами» переросли в небольшие семейные блоги-дневники либо просветительские блоги, публично рассказывающие о лечении. Крупнейшие из них: petya.belov, cmailikioskarandkatysha, terebovkolyahelp.

Часть – заканчивается на сообщениях об успешном завершении краудфандинговой акции и переходу к лечению.

Есть группы (matveimalenkov2015), оформленные как сообщества по сбору помощи, но заявляющие своей целью коллаборацию родителей.

Некоторые сообщества курируются фондами и другими некоммерческими организациями, в управлении сообществом указаны контакты сотрудников организаций. Но в текстах постов аффилированность со сторонней организацией не подчеркивается, большинство обновляются родственниками пациентов. В том числе сложности вызвало разделение таких аккаунтов на формально организованные извне или коммерческие и - неформальные. Можно говорить о том, что уже на этапе подготовки данных, гипотеза не подтвердилась: невозможно чётко отследить аффилированность сообщества.

Абсолютное большинство сообществ собирает средства детям, из 1415 сообществ, отобранных простым поисковым запросом, только 13 собирают средства для взрослых пациентов.

Не рассмотренными в теоретической части исследования остались сообщества студенческих научных обществ (dermven, club58968677, snk\_pediatrics, snk\_infectious\_diseases\_msmu). Все они в качестве дополнительной исследовательской работы предлагают студентам изучать редкие заболевания.

По оформлению сообщества достаточно однотипны, большинство используют для оформления семейные, либо студийные фото пациента. Очень редко в оформлении используются православные иконы.

Сообщения о смертях крайне редки. Но есть несколько примеров страниц памяти (public25078667, public26348596, группа сообществ посвященная вспышке менингита в Калининграде в 2010 году, club62312499, club18817269) которые поддерживаются родственниками ушедших людей. Избегание стигматизационных проявлений болезнью – одна из ярко выраженных стратегий, которые применяют некоммерческие организации и проекты, связанные с социализацией инвалидов.

Для анализа использован парсер и код на python3, представленный сотрудником Санкт-Петербургской школы социальных наук и востоковедения НИУ ВШЭ Аленой Пестовой на открытом семинаре (СМ. Приложение 6), используемый с разрешения автора. Часть кода заимствована из книги М. Рассела и М. Классена «Data Mining»[[3]](#endnote-2).

Текстовые данные постов пациентских сообществ собраны при помощи скрипта на Python3 запущенном в GoogleCollab через VK API методом wall.get. Всего собрано посты примерно 100 аккаунтов.

Для удобства анализа и возможности сравнить данные, датасет разделен на две части: Пациентские сообщества и Пациентские организации. Из общего числа аккаунтов выбраны схожие по количеству подписчиков Разделение проведено вручную по внешним признакам, крупные фонды и большие пациентские сообщества, юридически существующие как некоммерческие организации, отнесены к «Пациентским организациям», в эту группу вошли аккаунты: "rarediseasesru", "allergy\_school", "spiporz", "rarediseases", "silneye\_obstoyatelstv", "foprussia2014", "domorphans", "hemophilia\_bashkortostan", "gaucherspiporz". К пациентским сообществам отнесены: "jelezodeficytnaya\_aneviya", "dobroedelo2019", "diro36", "endoferin", "miopija\_blizorukost\_chto\_eto", "prim\_babochki", "mio\_mtm1", "bas\_redkoe\_zabolevanie", "neotchaivaysya", "edvards\_centr".

Корпус сообществ насчитывает 273 строки, корпус организаций – 2774. Полученный датасет имеет следующие признаки:

'group\_domain',

'ads',

'post\_type’,

‘text',

'likes'.

На следующем этапе проведена оценка лексического разнообразия с помощью, предложенного М. Расселом и М. Классеном способа.

Таблица 1.

|  |  |
| --- | --- |
| Оценка лексического разнообразия  (отношение числа уникальных слов к общему числу слов в тексте) | |
| Пациенские сообщества | Пациенские организации |
| 27,703 слов  9,493 уникальных словоформ  lexical\_diversity 0.4972677595628415  Среднее количество слов в предложении: 15.8 | 184,108 слов  30,081 уникальных словоформ  lexical\_diversity 0.659816247582205  Среднее количество слов в предложении: 19.7 |

Подготовка текстов для будущего анализа включала:

- удаление пустых строк;

- удаление рекламных постов;

- вывод и описание распределение длины текстов, удаление слишком коротких и слишком длинных;

- токенизация и лемматизация корпуса;

- выбор существительных для дальнейшего анализа;

- удаление стоп-слов при помощи библиотеки nltk. К стоп-словам добавлены: время, день, месяц, год. Так как в частотных словах всего корпуса выделялась группа обозначений времени.

Итоговый корпус состоит из 7 столбцов: group\_domain, ads, post\_type, text, likes, len\_text, clean\_text.

Таблица 2.

|  |  |
| --- | --- |
| Пациенские сообщества | Пациенские организации |
| значение статистик для длины текстов постов, кол-во символов | |
| count 273.000000  mean 713.172161  std 991.800989  min 6.000000  25% 112.000000  50% 332.000000  75% 888.000000  max 7318.000000 | count 2774.000000  mean 791.876712  std 1233.569459  min 6.000000  25% 103.250000  50% 332.500000  75% 1024.000000  max 14776.000000 |
| частотные слова по группам | |
| ('ребёнок', 175),  ('человек', 116),  ('бабочка', 99),  ('помощь', 83),  ('жизнь', 74),  ('миопатия', 73),  ('заболевание', 64),  ('друг', 62),  ('мама', 56),  ('шт', 53),  ('салфетка', 52),  ('мир', 45),  ('группа', 44),  ('поддержка', 40),  ('лечение', 39),  ('мыло', 38),  ('пациент', 38),  ('акция', 37),  ('родитель', 37),  ('вопрос', 36) | ('заболевание', 2266),  ('пациент', 1958),  ('ребёнок', 1730),  ('препарат', 1302),  ('лечение', 1122),  ('человек', 848),  ('россия', 803),  ('лекарство', 799),  ('болезнь', 775),  ('помощь', 771),  ('врач', 656),  ('организация', 566),  ('жизнь', 552),  ('фонд', 537),  ('центр', 506),  ('вопрос', 499),  ('рф', 478),  ('здравоохранение', 442),  ('минздрав', 438),  ('проблема', 414) |
| визуализация частотных слов в текстах | |
| Изображение выглядит как текст  Автоматически созданное описание | Изображение выглядит как текст  Автоматически созданное описание |
| слова, характерные для обеих групп | |
| ('ребёнок', 175),  ('человек', 116),  ('заболевание', 64),  ('лечение', 39),  ('вопрос', 36)  ('пациент', 38), | ('заболевание', 2266),  ('пациент', 1958),  ('ребёнок', 1730),  ('лечение', 1122),  ('человек', 848),  ('вопрос', 499), |

В целом, по представленному анализу видно насколько разная у выделенных двух групп аудитория:

Чаще всего обращаются к семьям, видно, что такие слова как «ребёнок», в числе самых употребляемых.

Сама болезнь не занимает важного места в контенте, который выбирают организации для своих сообществ. Правовые вопросы больше всего обсуждаются в сообществе пациентов с болезнью Гоше, тематика постов рассчитана на взрослую подготовленную аудиторию.

По лексическому анализу видно, что в текстах отсутствует прямая конфронтация, отсутствие так называемых «хулы» и «хвалы», никаких форм порицания или эмоционального выражения ситуации пациентов. Единственным фактором, показывающем кризисную ситуацию пациентов, можно считать активность по сборам средств.

Также большое разнообразие форм коммуникации может говорить о том, что у медицинских сообществ не выработаны общие позиции[[4]](#endnote-3).

На следующем этапе из данных были выделены имена и названия организаций.

Таблица 3.

|  |  |
| --- | --- |
| Пациенские сообщества | Пациенские организации |
|  | |
| 89 имен  Ляйсан Утяшева  Шекспир  Пьер Огюст Ренуар  Вадим Шефнер  Милорад Павич | 832 имени  В.В. Путин (23)  В. Скворцова (16)  М. Мишустин (12)  М. Мурашко (5)  В. Матвиенко (4)  А.Ю. Кузнецова (3)  Данила Козловский  Агния Барто  Ник Вуйчич  Йозеф Геббельс |

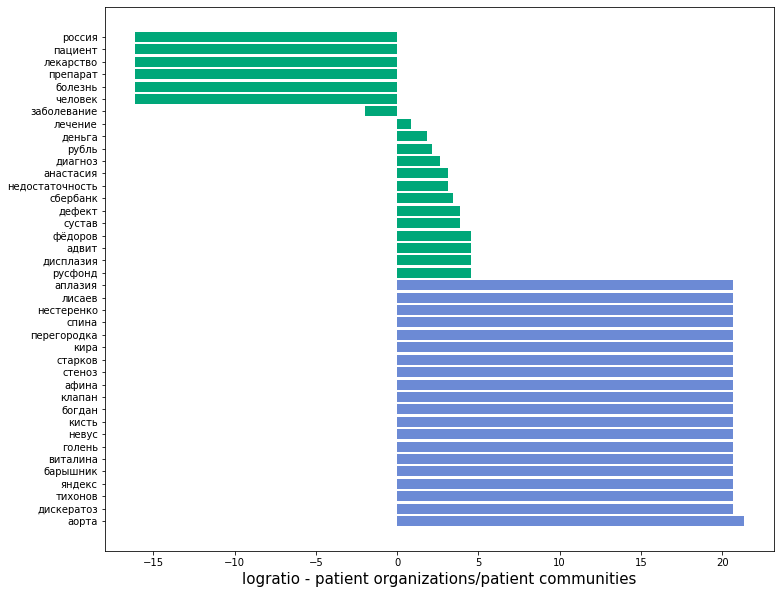
Можно подтвердить замечания из первой главы о важности личных брендов для медицинского сектора социальных сетей. По приведенным именам также можно сделать вывод о лояльности (аффилированности?) пациентских организаций к государственным структурам.

Завершением описания собранного датасета стало построение графика на основе G2 – статистики (Logratio - effect size). Этот метод подсчета коэффициентов для частотных слов в разных группах текстов, позволяет посчитать, насколько большая разница в частотах слов. Чем больше результат по модулю, тем специфичнее слово для определенной группы. Разброс показателей на графике визуализирует числовые данные о специфичности слов для двух групп: пациентских сообществ и пациентских организаций.

Таблица 4.

|  |  |
| --- | --- |
| Выделение слов, наиболее специфичных для каждой группы,  с коэффициентом, указывающим на принадлежность к группе | |
| Пациенские сообщества | Пациенские организации |
|  | |
| Изображение выглядит как стол  Автоматически созданное описание | Изображение выглядит как стол  Автоматически созданное описание |

Таблица 5.



Одним из нетипичных для российского общества особенностью сообществ пациентов с редкими заболеваниями является полное отсутствие дискуссии о нетрадиционных практиках заботы о здоровье. Данный датасет может быть хорошим примером, здесь нет упоминаний, опровержений или поддержки каких-либо методов лечения.

В ходе исследования был проведен сравнительный анализ текстов пациентских сообществ и пациентских организаций. На основании текстового анализа постов сообществ сделаны выводы об особенностях сообществ, их аудитории. Мы может утверждать, что эти сообщества дифференцированы как с точки зрения числа участников, так и с позиции направленности и формата деятельности.

В данном параграфе описаны сообщества, если можно так выразиться, медицинских социальных сетей. Предпринята попытка описать морфологию сообществ (стратификацию, солидарность), мотивы участия пользователей и экспертов и некоторые особенности экспертных и медицинских сообществ.

В будущем интересно оценить публикации (креативы) по нескольким блокам вопросов: ситуативный контекст, в котором воспринимается визуальная информация; информативность; соотношение тематики и лексической стилистики.

Бросается в глаза, что медицинские организации в отличие от организаций «третьего сектора» не используют феминитивы и политкорректные слова, это наблюдение могло бы стать основой для гипотезы следующего исследования.

1. Толентино, Дж. Кривое зеркало. Как на нас влияют интернет, реалити-шоу и феминизм / Джиа Толентино; [пер. с англ.]. – Москва: Эксмо, 2020. – С32; [↑](#endnote-ref-1)
2. Понятие редкое заболевание закреплено в Федеральном законе от 21 ноября 2011 года № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». К редким заболеваниям в Российской Федерации относятся нозологии, включенные в так называемый «большой список», опубликованный на сайте Министерства Здравоохранения и регулярно обновляемый. По состоянию на 23.09.2021 в списке числится 267 заболевания (немного больше чем в год первой публикации – 253 заболевания). [↑](#footnote-ref-1)
3. Рассел, М. Data Mining [Текст] = Mining the social web: извлечение информации из Facebook, Twitter, Linkedin, Instagram, Github / Мэтью Рассел, Михаил Классен ; [пер. с англ.: А. Киселев]. - [3-е изд.]. - Санкт-Петербург [и др.]: Питер, 2020. - 459, [3] с.: ил. - (IT для бизнеса). [↑](#endnote-ref-2)
4. Ванина, Е.Ю. Хула и хвала: коммуникативные действия как точка пересечения гуманитарных дисциплин // Под небом Южной Азии. Хула и хвала: коммуникативные модальности исторического и культурного своеобразия / рук. проекта И. П. Глушкова; отв. ред. Е. Ю. Ванина; Ин-т востоковедения РАН. – М.: Наука. – Вост. лит., 2017. – 927 с.: ил. – С44; [↑](#endnote-ref-3)