

УТВЕРЖДЕНО

Решением общего собрания членов
Ассоциации «Объединение исследователей рынка
и общественного мнения»
(Протокол от 29.08.2024г. №69)

**СТАНДАРТЫ КАЧЕСТВА ОИРОМ
ПО ИССЛЕДОВАНИЯМ СОЦИАЛЬНЫХ МЕДИА**

Оглавление

СТАНДАРТЫ КАЧЕСТВА ОИРОМ ПО ИССЛЕДОВАНИЯМ СОЦИАЛЬНЫХ МЕДИА	1
1. Введение.....	2
1.1 Общие определения и термины	3
1.2 Платформы социальных медиа	7
1.3 Предметная область исследований социальных медиа	8
1.4 Типы задач, решаемых с помощью исследований социальных медиа	8
2. Основные принципы.....	9
2.1 Разграничение маркетинговых и иных видов исследований	9
2.2 Конфиденциальность.....	10
2.3 Соблюдение законодательства и правил индустрии.....	10
2.4. Работа с разными источниками данных	11
3. Сбор данных социальных медиа	12
3.1 Корректность сбора данных	12
4. Чистка, обработка и анализ данных.....	12
4.1 Чистка данных.....	12
4.2 Анализ данных.....	13
5. Элементы, которые должны содержаться в предложении для клиента	15
5.1 Обязательные элементы в предложении	15
5.2 Опциональные элементы, которые могут содержаться в предложении для клиента	16
6. Элементы, которые должны содержаться в отчетных документах	16
6.1 Обязательные элементы в отчетных документах	16
6.2 Опциональные элементы, которые могут содержаться в отчетных документах.....	17
Приложение 1.	18
П1.1 Типы контента	18
П1.2 Структура данных.....	18

1 Введение

Социальные сети и другие площадки в интернете, где пользователи могут высказать свое мнение, сформировали новое коммуникационное пространство – социальные медиа. Соцмедиа стали важнейшим инструментом для общения – в них пользователи выражают свои мысли и чувства, делятся историями, фотографиями, видео и другими контентом, касающимся всех сфер своей жизни.

Это мощный инструмент для социального взаимодействия, отражающий культурные, социальные и потребительские тенденции. Соцмедиа – цифровой слепок общества, показывающий, чем оно живет сейчас и что происходило с ним в недавнем прошлом, с момента появления соцсетей.

С учетом глубокого проникновения социальных медиа в жизнь всех слоев общества их изучение предоставляет новые масштабные возможности для маркетинговых и социологических исследований.

Следует различать разные подходы к использованию социальных медиа: безопросный подход, использование платформ социальных медиа для опросов, использование соцмедиа для прямых коммуникаций и другие. Безопросный подход предполагает наблюдение и изучение уже опубликованного пользователями публичного контента. Использование возможностей платформ социальных медиа для фактически опросных подходов, включает в себя, например, формирование панелей с пользователями соцмедиа. Использование соцмедиа для других, в том числе прямых коммуникаций с пользователями может быть полезно, например, для решения проблемы у клиента бренда, если он обратился к бренду через социальные медиа.

Данное руководство посвящено именно безопросному методу исследования соцмедиа – исследованию спонтанно высказанных мнений пользователей, являющемуся по-настоящему новым видом исследований и потому требующим отдельных рекомендаций и стандартов.

Объектом таких безопросных наблюдений являются публичные сообщения пользователей. Сообщения с ограниченным доступом не должны анализироваться.

Помимо профилей пользователей, созданных и ведущихся реальными людьми, в соцмедиа есть также сообщества, профили организаций (в том числе СМИ) профили ботов и т.д. Безопросные маркетинговые исследования соцмедиа изучают, в первую очередь, публичный контент, созданный профилями пользователей на своих страницах и на страницах сообществ. Объектами отдельных видов исследований может быть также контент, созданный другими видами профилей.

Задачей исследования может являться типизация и последующее изучение групп пользователей, но не конкретных профилей.

Все промежуточные данные и финальные результаты исследования содержат только неперсонализированные данные. Ни на каком этапе исследования не может ставиться задача выявления реальных физических лиц, так или иначе связанных с контентом или профилями.

1.1 Общие определения и термины¹

Социальные медиа

[Соцмедиа]

В более узком и прикладном смысле под социальными медиа подразумеваются компьютерные технологии и/или технологические платформы, которые предоставляют пользователям (индивидуумам, компаниям/брендам, организациям/учреждениям) различные инструменты для создания, комментирования и публикации контента, публичного взаимодействия с сообществами/брендами/пользователями/избирателями, а также облегчают создание и обмен информацией, идеями, интересами и другими формами выражения через виртуальные сообщества и сети.

Ряд исследователей в определении термина акцентируют внимание на одном из наиболее важных аспектов социальных медиа, а именно на том факте, что социальные медиа – это прежде всего **публичный пользовательский контент**, размещенный на интернет-площадках и созданный или размещенный пользователями; соцмедиа – это цифровой слепок общества, показывающий, чем оно живет сейчас и что происходило с ним в недавнем прошлом, с момента появления социальных медиа.

Данные социальных медиа состоят из сообщений и метаданных. Сообщения могут быть мультимодальными, в том числе текстовыми, визуальными, аудио и так далее. К метаданным сообщения могут относиться: вовлеченность, просмотры, ссылка, дата и время публикации, место публикации, язык, автор, тональность, география сообщения. К метаданным авторов могут относиться: пол, возраст, география, интересы, семейное положение, никнейм, кол-во подписчиков/друзей и т.д.

Мониторинг упоминаний объекта / бренда

[Мониторинг бренда, Social Media Listening, Сбор/парсинг массива данных с упоминаниями бренда, медиамониторинг, мониторинг упоминаний в соцмедиа, мониторинг упоминаний объекта в социальных медиа]

Базовая услуга (базовый инструмент) в области управления репутацией и оценки коммуникаций, маркетинговой аналитики, состоит в своевременном и систематическом отслеживании сообщений пользователей и/или журналистов и/или блогеров/инфлюенсеров/ЛОМов, упоминающих бренд/компанию/определенную тему (отслеживание по ключевым словам) среди миллиардов других (не содержащих упоминаний ключевых слов) сообщений в социальных медиа и/или СМИ (в имеющих возможность индексирования источниках сети Интернет). Процесс мониторинга характеризуется систематическим характером наблюдений за состоянием объектов, явлений, процессов, количественных показателей, медиасобытий в социальных медиа и СМИ.

Позволяет устранять «слепые зоны» (а именно проверять упоминания в социальных медиа, которые иными средствами обнаружить объективно невозможно), вовремя реагировать на упоминания, требующие какого-либо реагирования, фиксировать и устранять возможные репутационные угрозы, принимать к сведению и упреждать возможные негативные последствия от сливов конфиденциальной информации, предотвращать возможные финансовые потери (мониторинг рисков оттока клиентов и/или кадров, взлома программного обеспечения, продажи данных на чёрный рынок, сбоя в работе инфраструктуры и т.п.).

¹ Полный набор терминов и их определения можно найти в "Глоссарии индустрии аналитики социальных медиа"

Объект мониторинга

[Ключевик, ключевые слова]

Компания и/или бренд, персона, продукт или словосочетание, по которому выполняется поиск и анализ упоминаний в сети Интернет.

Информационно-аналитическая система

[Мониторинговая платформа, брендтрекер, social listening tool, Система мониторинга и анализа упоминаний]

Программное обеспечение (чаще онлайн-сервис), позволяющее собирать, хранить, обрабатывать по определенным критериям и передавать информацию. Система мониторинга и анализа упоминаний - это программный комплекс для анализа больших данных (Big Data) в сети Интернет, позволяющий отслеживать, анализировать и размечать упоминания компаний и брендов и использовать их для управления их репутацией, а также проведения маркетинговых исследований безопросного типа.

Инфлюенсеры

Инфлюенсером считается пользователь (также возможно виртуальный персонаж), у которого есть некоторое число фолловеров, но важным является то, что он способен влиять на своих подписчиков. Вместе с тем инфлюенсером может являться и виртуальный персонаж, созданный с помощью компьютерной графики, который имеет определённый жизненный бэкграунд и ведёт активную жизнь в соцсетях.

Коммуникационная стратегия

Стратегия построения коммуникаций с целью продвижения бренда, товаров и услуг. Включает три составляющие: маркетинговую, креативную и медийную стратегии.

Инфополе (в глобальной сети)

[Информационное поле]

Информационное поле глобальной сети - пространство Интернета, где существуют гипертексты и ссылки на информационные поводы, осуществляется многоаспектная дискурсивная медиакоммуникация. Чаще всего стимулом к формированию информационного поля в медиа являются либо информационный повод - актуальное событие, либо информационно-медийная личность, формирующая цепь реакций на себя и свои высказывания.

Контент

Любое информационно значимое наполнение ресурса (например, веб-сайта или аккаунта в социальных медиа): тексты, графика, элементы интерфейса, аудио и видеофайлы, файлы в любом формате и т.д. Контент может быть либо постоянным, либо пополняемым.

Мессенджер

Приложение или программа для мгновенного обмена текстовыми сообщениями, аудиозаписями, фотографиями и другими мультимедиа между пользователями. Мессенджеры устанавливаются на компьютер, смартфон, планшет и работают через интернет.

Подписчик

[Фолловер]

Пользователь, проявивший интерес к странице или аккаунту в социальных медиа и выразивший активное желание получать обновления контента от этого источника через трансляционные механизмы интернет-ресурса.

Пользовательский контент / UGC

Контент, создаваемый пользователями.

Социальные сети

Интернет-ресурсы, предназначенные для общения пользователей и предоставляющие комплексные возможности для самовыражения, развлечения и получения информации. Ключевой особенностью соцсетей является стремление ухода от анонимности. Примеры социальных сетей: ВКонтакте, Одноклассники, Facebook и Instagram (признаны экстремистскими и запрещены на территории РФ) и др.

Маскирование данных

Специальный метод, при котором исходные данные изменяются таким образом, что становится очень сложно их в сети Интернет с помощью поисковых систем, а, следовательно, и установить личность автора. Это эффективный метод, позволяющий сохранить анонимность авторов комментариев в случае необходимости.

Сообщение

[Публикация, материал, пост]

Публичное сообщение (посты, репосты, комментарии) в социальных медиа или СМИ, содержащие упоминание объекта мониторинга. Сообщения могут быть мультимодальными, в том числе текстовыми, визуальными, аудио и т.п.

Источник

Сайт, онлайн-платформа или оффлайн-СМИ, на котором было опубликовано сообщение, содержащее объект мониторинга. Например, ВКонтакте, Telegram, Яндекс.Карты, газета Ведомости, радио Business FM и т.д.

Вовлеченность сообщения

Показатель реакции аудитории, позволяющий выявить наиболее значимые публикации. Рассчитывается как сумма комментариев, лайков и репостов сообщения.

Комментарий

Публичное сообщение в социальных медиа, опубликованное пользователем в формате ответной реакции, либо оставленное в цепочке обсуждения на информационной площадке.

Лайки

[Лайк, Мне нравится, Мне понравилось, Класс!, +1, Like]

Функция в коммуникационном программном обеспечении, предназначенном для социальных сетей, блогов, интернет-форумов, сервисов социальных закладок, новостных сайтов, которая используется для выражения отношения пользователей к тому или иному контенту. Лайк выражает одобрение какого-либо действия, фотографии, публикации или обновления статуса пользователя в социальных сетях, нажатием на соответствующую кнопку (чаще в форме сердечка или поднятого

вверх большого пальца руки). Релевантно в отношении ВКонтакте, Youtube, Одноклассников, Twitter (X), TikTok, признанных экстремистскими Instagram и Facebook. Для выражения антипатии или неодобрения к опубликованному контенту используется кнопка Дизлайк /Dislike/ «Мне не нравится», которая обычно обозначается большим пальцем руки, опущенным вниз.

Репост / дополненный репост

[Shares]

Один из основных видов активности в соцсетях наряду с лайками и комментариями, когда один пользователь или страница делится постом другого пользователя или страницы в своей ленте, со своим комментарием (дополненный репост с пользовательским сопроводительным комментарием) или без.

Аудитория сообщения

Количество потенциальных просмотров сообщения. Аудитория сообщения определяется по месту публикации сообщения. Аудитория сообщений, опубликованных одним и тем же автором у себя в ленте или в сообществе, будет иметь разную аудиторию: для поста в ленте автора - аудитория равна количеству друзей и подписчиков автора, для поста или комментария в сообществе - аудитория равна количеству участников сообщества.

Пост

Сообщение-публикация в блоге/в социальной сети/на информационной площадке в сети Интернет. Запостить — написать сообщение.

Тег / хештеги / хэштеги

Текстовая, интерактивная метка с символом решетки, используемая для агрегации контента в поисковой выдаче интернет-ресурса.

Упоминание

Упоминание бренда, продукта или лица в социальных сетях, публикациях или блогах. Это информативный показатель, который помогает компаниям оценить уровень узнаваемости их бренда.

WOM

[WOM-сообщения, сарафанное радио]

Сообщения от первого лица, в которых пользователи, как правило, выражают собственное мнение, рассказывают о личном опыте, рассуждают, сравнивают, жалуются, благодарят, рекомендуют или задают вопросы. Чаще всего такие сообщения содержат пользовательские мнения о товаре или услуге, основанные на личном опыте, написанные от первого лица и адресованные в диалоге собеседнику, изначально доверяющему автору сообщения.

Бот / фейковый аккаунт

Искусственно созданный профиль, имитирующий человека. Порой создается с целью развития комьюнити-менеджмента или для управления репутацией бренда в соцсетях. Иногда это компьютерная программа, автоматически выполняющая запрограммированные задачи.

Тональность

[Tone of voice]

Качественная характеристика упоминаемых объектов, демонстрирующая позицию автора по отношению к объекту. Учитывает эмоционально окрашенную лексику, специфику информационного повода и то, как его оценивает сама компания/организация/бренд. Обычно измеряется по трехпозиционной шкале (нейтральная, позитивная, негативная), но могут быть и другие варианты.

Автоматический анализ тональности текста

Класс методов контент-анализа, предназначенный для автоматического выявления в тексте эмоционально окрашенной лексики, а также мнений (эмоциональных оценок) автора по поводу объектов, о которых идет речь в тексте.

Атрибуция / кодировка

[Разметка данных, разметка параметров, тегирование]

Процесс определения критериев/параметров для анализа базы публикаций и простановки значений этих параметров для каждой из совокупности публикаций. В число наиболее распространенных кодируемых параметров входят тональность, упоминаемость информационных поводов/тем/продуктов/объектов, степень сфокусированности на объектах, ФИО/должность спикеров и др.

Язык поисковых запросов и поисковые операторы

Набор инструментов и правил, с помощью которых пользователь формирует свой запрос через интерфейс информационно-аналитических систем. В указанных системах можно вводить не только буквы, но и определенные символы/комбинации, которые позволяют найти нужную информацию. Эти символы называют операторами поисковых запросов. Идеальный поисковый запрос удовлетворяет одновременно двум критериям: полнота (покрывает все упоминания искомого объекта) и релевантность (в поток не попадают упоминания, не относящиеся к искомому объекту). Как правило, поисковый запрос учитывает варианты написания, порядок слов, синонимы и ассоциации, возможные опечатки, жаргонизмы и хештеги.

1.2 Платформы социальных медиа

Платформы, которые попадают в анализ социальных медиа – это сайты и площадки, на которых создается пользовательский контент. От типа платформы может зависеть тип собираемого контента. При проведении исследовательского проекта необходимо фиксировать, и по запросу предоставлять клиенту информацию о том, на каких площадках собирались данные.

Типы источников, которые попадают в анализ:

- Социальные сети
- Блоги
- Форумы
- Гео-сервисы
- Видео-хостинги
- Отзовики, включая esom
- Мессенджеры
- Комментарии в СМИ

Для сбора данных чаще всего используются информационно-аналитические системы. Также могут применяться системы собственной разработки (“парсеры”). Заказчик на этапе предложения должен быть проинформирован о способе сбора данных и связанных с ним ограничениях (полнота, сроки, стоимость).

1.3 Предметная область исследований социальных медиа

Речь в данном разделе пойдет об анализе данных, собираемых безопросными методами, - исследованиях социальных медиа. Главная ценность такого рода данных – возможность получить спонтанное пользовательское мнение.

Исследования социальных медиа представляют собой анализ публичного пользовательского* контента, спонтанно созданного и размещенного на интернет-площадках пользователями.

*Пользователи социальных медиа.

1.4 Типы задач, решаемых с помощью исследований социальных медиа

Данный стандарт относится к исследованиям социальных медиа, целью которых, является принятие маркетинговых решений (например создание и развитие продукта / бренда, понимание целевой аудитории, выработка стратегии и тактики коммуникаций).

Типичные объекты и задачи исследования:

Объект	Задачи
Продукты (относящиеся к заказчику и/или конкурентам)	Сильные / слабые стороны продуктов, поиск направлений для развития Пользовательский путь (CJM) по продукту
Категория (например, питьевые йогурты / искусственный интеллект и т.п.)	Неудовлетворенные потребности в категории Пользовательский путь (CJM) в категории
Восприятие бренда /имиджа продукта (в т.ч. в конкурентном окружении)	Сильные / слабые стороны восприятия бренда Ключевые факторы, влияющие на восприятие бренда
Краткосрочные тренды (события / ситуации, вызывающие активный отклик аудитории)	Выявление трендов для подготовки ситуативных коммуникаций
Долгосрочные тренды (долгосрочные изменения в потребительском поведении, категории)	Выявление трендов для планирования коммуникационной и продуктовой стратегии
Коммуникационных кампании / мероприятия (собственных или кампаний конкурентов)	Оценка эффективности элементов коммуникации (акционные механики, посеы, блогеры и т.п.)
Коммуникационный инструментарий в соцмедиа (сообщества, лидеры мнений, механики)	Подбор эффективных инструментов для решения маркетинговой задачи
Аудитория (например, медработники)*	Выявление мнений, потребностей, digital-профиля (выбираемые бренды, предпочитаемые инфлюенсеры, привычки и т.д.) конкретной аудитории

** Возможности анализа аудитории существенно ограничены характером собираемых данных. Базово выявление большей части аудиторий возможно только на основании ключевых слов. Также выявление аудитории может быть организовано на базе метаданных, которую пользователи размещают в своих профилях (пол, возраст, место работы), однако необходимо учитывать ограничения доступности этой информации (не все пользователи заполняют профили, часть социальных медиа не предоставляют отдельные виды информации о пользователях для сбора)*

Возможность покрытия отдельных исследовательских задач может варьироваться с учетом особенностей способа сбора данных и объекта исследования. Некоторые объекты / задачи не могут быть покрыты из-за отсутствия необходимого объема данных (объект исследования не упоминается или слабо упоминается в социальных медиа). Перед принятием решения о реализации конкретного исследования должна быть проведена предварительная оценка объемов и структуры данных (feasibility-check) для формирования адекватных ожиданий заказчика и корректного расчета стоимости. Результаты оценки объемов должны быть в явном виде представлены заказчику на этапе предложения.

Исследовательская задача может быть решена в рамках разового проекта (ad-hoc) или как серия последовательных проектов (трекинг)

2. Основные принципы

Все основные принципы стандартов ОИРОМ применяются и к исследованиям социальных медиа. В нижеследующем разделе объясняются принципы, лежащие в основе подхода к проведению исследований социальных медиа:

- Разграничение маркетинговых и иных видов исследований
- Конфиденциальность
- Соблюдение законодательства и правил индустрии

2.1 Разграничение маркетинговых и иных видов исследований

Разграничение деятельности по исследованию социальных медиа для решения маркетинговых задач от прочих видов деятельности (в том числе продажи, direct-маркетинга, рекламы, PR, GR и других видов деятельности, связанных с направленным формированием общественного мнения или отношения к какому-либо объекту), является условием, необходимым для:

- обеспечения объективности и достоверности результатов исследования, а также качества принимаемых на их основе решений;
- поддержания репутации профессионального исследовательского сообщества как независимой экспертной функции общества;
- защиты интересов авторов публикаций в социальных медиа от направленных на них нежелательных действий без их прямого согласия.

Исследователи не должны допускать использования собираемых ими для исследовательских проектов персональных данных (или синтетических / восстановленных данных, схожих с персональными) в каких-либо иных целях, помимо проведения маркетинговых и социальных исследований, осуществлять какую-либо альтернативную деятельность под видом маркетинговых исследований (скрывая ее прямые цели от объекта воздействия), а также предпринимать какие-

либо действия, направленные непосредственно на лицо, участвовавшее в исследовании, без его прямого согласия.

Эти требования допускают проведение исследований организациями, не являющимися исследовательскими, при условии соблюдения указанных требований и данного стандарта.

2.2 Конфиденциальность

Исследовательские организации в ходе анализа социальных медиа используют открытые данные источников, собранных самостоятельно или с помощью доступных информационно-аналитических систем. Для данных, собранных в публичном пространстве, не применяется маскирование информации, в отчетных материалах возможно давать ссылки на посты и цитаты в неизменном виде.

При этом стоит помнить, что исследовательские организации отвечают конфиденциальность и, как правило, работают с неперсонализированными данными. Кроме тех случаев, когда пользователь самостоятельно разместил в открытом доступе информацию и ее использование необходимо для решения задач исследования.

Рекомендуется применять маскирование данных в следующих случаях:

- если неприкосновенность частной жизни участника или его ожидание неприкосновенности частной жизни может быть нарушено
- при изучение сенситивных тем
- если есть риски или косвенное подтверждение, что пользователь является несовершеннолетним

Также маскирование данных может применяться при отдельной договоренности с заказчиком исследования.

2.3 Соблюдение законодательства и правил индустрии

Исследовательские организации должны соблюдать существующие законы России и международные законы и правила, регулирующие неприкосновенность частной жизни, защиту данных, а также разглашение, получение и использование информации, позволяющей установить личность участника. Также стоит принимать во внимание необходимость соблюдения правовых процедур (например, требований норм о соблюдении конфиденциальности), касающихся трансграничной передачи информации, позволяющей установить личность участника (в тех случаях, когда персональные идентификаторы не были удалены из информации, однако такая информация передается через национальные границы); а также в вопросах, касающихся обеспечения надлежащей защиты данных в стране, куда такие данные передаются.

Получая доступ практически ко всем платформам соцмедиа, исследовательские организации должны быть осведомлены о Правилах пользования сервисами и следовать законодательству. Следует отметить, что в некоторых случаях материалы собираются разных сайтов, и потому может быть неясно, какие Правила пользования регулируют сложившуюся ситуацию. В таких случаях исследователи должны принять необходимые меры по соблюдению прав на неприкосновенность частной жизни или авторских прав.

Например, с Пользовательскими соглашениями популярных социальных медиа можно ознакомиться по следующим адресам в сети Интернет:

- ВКонтакте: <https://vk.com/terms>
- Одноклассники: <https://odhelp.ru/pravila-polzovaniya/>
- Телеграмм: <https://telegram-sss.ru/user-agreement/>

Государственное регулирование исследований социальных медиа развивается по мере развития самих социальных медиа. Исследовательские организации обязаны соблюдать и выполнять правила и законы по мере их появления.

При использовании информационно-аналитических систем следует руководствоваться условиями договора, описывающего порядок сбора данных.

2.4. Работа с разными источниками данных

При работе с социальными медиа необходимо разграничивать 2 вида контента:

- Контент, размещенный в публичном пространстве
- Контент, размещенный в закрытом пространстве

2.4.1 Публичное пространство

Публичное пространство – это область, в которой большинство участников обоснованно полагает, что любой человек может прочитать, процитировать, воспроизвести, дать ссылку, позаимствовать и поделиться их статьями, сообщениями и медиа-файлами. Большая часть социальных медиа подпадает под это определение. В качестве примеров можно привести страницы публичных профилей в социальных сетях, публичные блоги, микроблоги, некоторые сообщества, возникшие естественным образом, комментарии на новостных сайтах и многие форумы (включая те, которые могут потребовать ввода имени пользователя, но при этом автоматически предоставляют доступ к своим ресурсам и не являются модерлируемыми исследователями).

Некоторые социальные медиа позволяют пользователям указывать, какие их материалы будут доступны для широкой публики, а какие – нет. В данном случае термин «публичное пространство» означает разделы сайта, в которых имеется публичный доступ к данным, опубликованным пользователями, в соответствии с настройками конфиденциальности или Правилами пользования ресурсом.

2.4.2 Частное пространство

Частное пространство – это область, доступ к которой предоставляется только конкретным участникам и в рамках которой пользователи считают свои комментарии не предназначенными для посторонних лиц. Несмотря на то, что процедура получения доступа может носить практически мгновенный характер, эти источники становятся доступны только после получения пользователем учетной записи и/или пароля, цель создания которых – ограничить любые формы доступа к ресурсу. В качестве примеров можно привести многие частные форумы, сообщества, разделы чатов, системы мгновенного обмена сообщениями и форумы/группы, куда доступ контролируется администратором или модератором и материалы которых скрыты от широкой публики.

В рамках практики маркетинговых и социальных исследований изучение такого контента практически не проводится. При необходимости проведения такого анализа, не являющегося формально исследованием соцмедиа, необходимо учитывать и выполнять все требования законодательства РФ, в том числе получение личного информированного согласия всех участников,

посты которых находятся в этом пространстве и будут анализироваться, выполнение условий конфиденциальности, а также другие.

3. Сбор данных социальных медиа

3.1 Корректность сбора данных

Компания несет ответственность за корректность сбора данных социальных медиа. Все необходимые материалы должны сохраняться и по запросу предоставляться клиенту: количество и типы источников, размер выборки, поисковые запросы, и другие.

Исследовательские организации должны всегда проверять, какие условия применяются к материалам, используют ли они социальные медиа и выполняют ли запросы по защите персональных данных. Следует отметить, что исследователи должны воздерживаться от самостоятельных попыток обойти механизмы защиты данных вебсайтов (например, с помощью подмены IP-адреса, поддельных имен пользователя и т.д.).

В тех случаях, когда исследовательские организации используют информационно-аналитические системы, на исследователя возлагается обязанность по уточнению у своего поставщика вопроса о получении разрешений и сборе данных на законных основаниях.

Ниже приведены ссылки на наиболее популярные российские информационно-аналитические системы (сбора данных) и их соглашения:

- Brand Analytics: <https://brandanalytics.ru/>
- Медиалогия: <https://www.mlg.ru/>

4 Чистка, обработка и анализ данных

4.1 Чистка данных

Чистка данных предполагает удаление части собранного контента, нерелевантного для целей анализа.

Методы, используемые для чистки данных, должны быть задокументированы для предоставления клиенту (по запросу). Документация о процессе чистки данных должна быть использована в случае повторения проекта в будущем (для обеспечения сопоставимости результатов). Изменение методов чистки исключает возможность сопоставления результатов

В зависимости от целей и задач исследования чистка может включать в себя удаление различных типов контента. Возможные элементы чистки данных:

- Спам (сообщения не относящиеся к объекту исследования например упоминания со словом “Rich” в прямом значении при анализе бренда Rich)
- Рекламные сообщения
- Дублирующийся сообщения
- Несодержательные сообщения (например, состоящие только из хештега бренда, без дополнительного контента)
- Сообщения, содержащие только контент, анализ которого не предполагается (например сообщения, содержащие только изображение)
- Сообщения, сгенерированные ботами (критерии определения ботов должны быть задокументированы и представлены клиенту)

Чистка может быть разного уровня в зависимости от решаемых аналитических задач (например, верхнеуровневая, глубокая, для расчета метрик и т.д). Также чистка может включать в себя

элементы подготовки данных к анализу (например, нормализацию, удаление отдельных символов).

Чистка может осуществляться несколькими способами:

- Сплошной ручной анализ
- Ручная чистка с применением инструментов информационно-аналитических систем
- С применением моделей машинного обучения

Сплошной ручной анализ представляет собой процесс удаления нерелевантного контента разметчиками / кодировщиками, которые просматривают и анализируют каждое сообщение в массиве на предмет соответствия задачам исследования. Сплошной ручной анализ обеспечивает высокую точность чистки данных, однако может иметь высокую длительность и стоимость. Часто процесс сплошного ручного анализа совмещается с другими видами ручной разметки данных, используемыми при дальнейшем анализе.

Результаты сплошного ручного анализа должны быть проверены при подготовке отчета на документацию о процессе чистки. Если в процессе чистки задействовано более 1 человека, особое внимание следует обратить на однозначность трактовок документации всеми участниками процесса. Применение сплошного ручного анализа не гарантирует 100% точности чистки. Финальный уровень погрешности в данных, которые используются в анализе, не должен превышать 5-7% (с учетом релевантных сообщений, размеченных как нерелевантные и нерелевантных сообщений, размеченных как релевантные)

Ручная чистка с применением инструментов информационно-аналитических систем ориентирована на удаление существенной части нерелевантного контента с минимальными затратами ресурсов. Процесс чистки строится на выявлении распространенных паттернов, характерных для нерелевантного контента (например, ключевые слова, места публикации, авторы) и массовом удалении контента соответствующего паттерну. Такой метод чистки может приводить к удалению существенных объемов релевантного контента. Правильно проведенная чистка с применением инструментов информационно-аналитических систем, обеспечивает удаление до 60-80% нерелевантного контента.

Чистка с применением моделей машинного обучения основывается на применении алгоритмов (например, классификатор, выявляющий рекламный контент). Применяемые модели должны иметь оценки точности, которые должны быть включены в документацию по чистке данных.

4.2 Анализ данных

Выбор методов анализа определяется задачами исследования и должен быть задокументирован и согласован с заказчиком. Методы анализа должны быть экономически и методологически обоснованы, не допускается усложнение дизайна для увеличения стоимости проекта

Изменение методов анализа в процессе исследования допускается, но должно быть задокументировано и согласовано с заказчиком. В случае, если часть анализа уже проведена, исполнитель может потребовать от заказчика компенсации фактически понесенных затрат и пересмотра сметы на исследования с учетом нового подхода к анализу. Рекомендуется избегать методов анализа, которые нельзя будет воспроизвести / перепроверить в будущем.

Анализ может включать:

- Анализ содержания
- Анализ метрик
- Анализ связей

Анализ содержания позволяет выявить ключевые объекты и контексты их обсуждения (например свойства продукта, атрибуты бренда, тренды, особенности восприятия коммуникации и т.п.)

Анализ содержания может проводиться как количественно, так и качественно. Выбор конкретного метода определяется задачами исследования. Для достижения наилучшего результата рекомендуется комбинировать оба подхода

Ключевые категории в анализе содержания должны соответствовать задачам исследования (например если в задачах предлагается построение CJM то категории должны строиться на базе этапов клиентского пути, если предполагается анализ продукта - то свойств продукта). Целевой список категорий, как правило формируется в процессе предварительного качественного анализа. Список категорий рекомендуется согласовывать с клиентом

Для формирования количественных оценок содержания могут использоваться следующие методы

1. Ручная разметка (каждое сообщение анализируется разметчиком / кодировщиком в ручную, для выявления элементов, актуальных для анализа)
2. Семантическая разметка (для целевых категорий подбираются ключевые слова. Отношение к теме определяется по вхождению ключевых слов или их сочетаний, как правило осуществляется на базе информационно аналитических систем)
3. С применением моделей машинного обучения
 - a. Классификация (Top-down) - выделяем предзаданный список категорий
 - b. Кластеризация (Bottom-up) - модель сама определяет категории на основе анализа

Результатом анализа содержания являются доли ключевых категорий. Далее к ним могут быть применены любые методы статистического анализа (например индексы аффинитивности, карты соответствий - correspondence map и др.)

Анализ метрик может включать в себя расчет абсолютных значений, динамики и драйверов метрик упоминаний об объекте исследования в целом и его составных частей, ключевых категорий. Примеры метрик (количество сообщений, вовлеченность, количество лайков на сообщение, share of voice и др.) Базой для расчета метрик являются метаданные анализируемых сообщений. Анализ метрик в большей степени помогает решать коммуникационные задачи (подбирать каналы / лидеров мнений имеющих наибольшие охваты, тематики лучше всего вовлекающие аудиторию и т.п.).

Перечень метрик, используемых для анализа может гибко меняться в зависимости от задачи. Для всех используемых метрик должно быть дано четкое определение и приведено описание расчета. Для авторских метрик допускается концептуально описание без конкретной формулы. Особенности используемых метрик должны быть в явном виде сообщены заказчику для обеспечения возможности корректной интерпретации результатов.

Анализ распространения информации предназначен для понимания специфики развития отдельных тем (например трендов - поиск авторов трендсеттеров и каналов, через которые тренд в дальнейшем распространялся).

Если в процессе анализа используются инструменты / методы (например SaaS-сервисы на стороне третьих лиц такие как ChatGPT), которые могут привести к передаче данных исследования третьим лицам необходимо в обязательном порядке уведомить заказчика.

Источники погрешности должны быть явно указаны в отчетных документах для клиента:

1. Погрешности связанные с первичным отбором данных для анализа (запрос / стоп-слова / фильтры)
2. Погрешности, связанные с чисткой данных (например оценка точности сплошного ручного анализа)
3. Погрешности связанные с построением выборок (если применялись) и / или взвешиванием
4. Погрешности связанные с анализом данных (в т.ч. для моделей машинного обучения, указать основные параметры, влияющие на качество анализа, например точность)

5 Элементы, которые должны содержаться в предложении для клиента

Предложение, которое компания готовит для клиента, должно содержать необходимые (обязательные) элементы, которые позволяют клиенту получить необходимую информацию о предполагаемом исследовании.

5.1 Обязательные элементы в предложении

- **Цели исследования**, его основные потребители и их предполагаемые ожидания от результатов исследования.
- **Методы аналитики, рекомендуемые агентством**. Желательно не только их перечисление, но и описание.
- **Анализируемые источники и период анализа**. Перечисление конкретных источников или указание типов источников с обязательной детализацией по тематике, языку, географии распространения, формату публикуемого контента. Указание типов контента, подпадающих под анализ, и исключенных из него. Политика валидации данных (фильтры и параметры отнесения данных к нерелевантным).
- **Формат предоставления результатов исследования**. Под форматом предполагается как формальное описание ожидаемых продуктов (тип и формат документов, количество и структура контента внутри отчетов, требования к оформлению, визуализации и т.п.), так и функциональное (срок предоставления, наличие или отсутствие консультации по результатам исследования, количество таких консультаций, состав группы консультантов).
- **Стоимость работ**. Смета, детализированная по типам работ. Возможна детализация расчета на основе почасовых ставок специалистов и нормативного количества часов, необходимого для решения задач. Обязательно указывается, включает ли сумма налоги или иные дополнительные коэффициенты или платежи.
- **Указание работ, выполняемых с привлечением подрядчиков, перечисление подрядчиков**. Приведены прямые ссылки на ресурсы, содержащие подробную информацию о конкретных подрядчиках, в т.ч. об их профессиональном опыте и стандартах. Указаны условия, по которым ведутся подрядные работы.
- **Политика использования инструментов на основе ИИ**. Обязательное указание применяемых на проекте технологий, их краткое описание, в том числе методов валидации и проверки результатов работ, проведенных с помощью ИИ. Указание точности для инструментов на основе ИИ.

5.2 Опциональные элементы, которые могут содержаться в предложении для клиента

- **Краткое резюме или транскрипт брифа** (брифинга), проведенного перед формированием предложения, с выделением основных требований, озвученных и утвержденных клиентом.
- **Описание опыта и компетенций команды** исследователей, выделяемой на проект.
- **Стандарты, фреймворки и подходы к проектному менеджменту** (управлению проектами), которые будут применяться в ходе проекта. Режим доступности специалистов клиентской поддержки.
- **Стандарты, фреймворки и подходы к контролю качества**, которые будут реализовываться на проекте.
- **Политика работы** с чувствительными темами.

- **Предоставление кейсов** и обоснования, подкрепленное кейсами, в рамках которых решались аналогичные задачи. Список метрик, их описание, методология, расшифровка и формулы для их расчета, иллюстрации по возможным вариантам визуализации этих метрик.

6. Элементы, которые должны содержаться в отчетных документах

Результаты анализа данных представляются клиенту в виде отчета в согласованной сторонами форме. Наиболее распространенные формы отчетности

- Презентация (например в формате PowerPoint)
- База данных / аналитические таблицы (в формате Excel)
- Дэшборд

Отчет должен содержать основные методологические особенности проекта (сбор, анализ данных, источники погрешности) и результаты анализа.

6.1 Обязательные элементы в отчетных документах

- **Цели исследования**, основные задачи, которые решает отчет.

- **Анализируемые источники и период анализа**. Перечисление конкретных источников или указание типов источников с обязательной детализацией по тематике, языку, географии распространения, формату публикуемого контента. Указание типов контента, подпадающих под анализ, и исключенных из него. Политика валидации данных (фильтры и параметры отнесения данных к нерелевантным).

- **Анализ данных**. Подробное описание методов исследования (количественные, качественные), применение кодировки, кодфрейм и т.д., контроль качества

- **Политика использования инструментов на основе ИИ**. Обязательное указание применяемых на проекте технологий, их краткое описание, в том числе методов валидации и проверки результатов работ, проведенных с помощью ИИ. Указание точности для инструментов на основе ИИ.

- **Результаты анализа**. Предоставление результатов анализа данных.

- **Выводы/рекомендации.** Саммаризация результатов исследования, рекомендации на основе данных

6.2 Опциональные элементы, которые могут содержаться в отчетных документах

- **Иллюстрации отчетных материалов** с помощью цитат, визуальных материалов
- **Политика работы** с чувствительными темами.
- **Дополнительная информация**, оговоренная с Клиентом

Приложение 1.

П1.1 Типы контента

Типы контента в социальных медиа:

Есть два типа контента (есть посты без стимула/цели):

- контент-стимул, созданный с какой-то целью
- контент-реакция

П1.2 Структура данных

Данные социальных медиа состоят из сообщений и метаданных.

Все, что может быть получено не нарушая ограничений пользователя, законодательства, может быть использовано в анализе.

Сообщения могут быть мультимодальными, в том числе текстовыми, визуальными, аудио и так далее.

К метаданным сообщения могут относиться:

- Вовлеченность
- Просмотры
- Ссылка
- Дата и время публикации
- Место публикации
- Язык
- Автор
- Тональность
- География сообщения

К метаданным авторов могут относиться:

- Пол
- Возраст
- География
- Интересы
- Семейное положение
- Никнейм
- Кол-во подписчиков/друзей

	Метаданные сообщения	Метаданные автора
Доступные	<ul style="list-style-type: none">● Вовлеченность● Просмотры● Ссылка● Дата и время публикации● Место публикации● Автор● География сообщения	<ul style="list-style-type: none">● Пол● Возраст● География● Интересы● Семейное положение● Никнейм● Кол-во подписчиков/друзей

	<ul style="list-style-type: none"> ● Тип сообщения ● Оценка (отзовики) 	<ul style="list-style-type: none"> ● Ссылка на профиль
Синтезируемые	<ul style="list-style-type: none"> ● Тональность ● География сообщения ● Язык ● Тематики (например, здравоохранение) ● Категории (например, WOM) ● Агрессивность ● Роль объекта 	<ul style="list-style-type: none"> ● Пол ● География ●

Метаданные могут быть представлены как непосредственно платформой, так и синтезированы (полученные в рамках обработки сообщений и/или других метаданных. Например, определение гендера по используемым глагольным формам). Могут относиться как к сообщениям, так и к авторам.