

# ИНСТРУКЦИЯ ПО ВЕДЕНИЮ НАУЧНОГО ИССЛЕДОВАНИЯ

Вторичные данные





Сегодня мы изучим использование вторичных данных для вашего исследования, как можно использовать различные виды ресурсов в вашей работе, и как проводить анализ над этими данными.

Вторичные данные – это данные, которые существуют независимо от исследовательского проекта. Это данные, которые можно найти в документах, базах данных и в Интернете – ни один из них не был создан исследователем для конкретной цели его исследовательского проекта.

Это существующие данные, которые исследователи просто собирают и анализируют.

## Работа с существующими данными

### *Анализ текста / документов*

Часто ответы на вопросы исследования содержатся в документах, возможно, письмах, журналах, программных документах и отчетах. Метод, который мы будем использовать для анализа таких документов, называется анализом документов. Но вас также могут заинтересовать такие формы общения, как видео, телешоу, радиопередачи и даже веб-сайты. В этом случае, вместо того, чтобы ссылаться на них как на документы, мы используем немного более общий термин «тексты» и занимаемся текстовым анализом.

### *Сбор, просмотр и анализ текстов*

Итак, как вы заставите текст говорить? Как только вы найдете, приобретете и оцените достоверность своих текстов, вы будете готовы «извлечь» данные. Первый шаг – задать себе вопросы по поводу текста. Это относится к вопросам, связанным с автором / создателем, аудиторией, как они были произведены, типом, является ли это типичным или исключительным примером, стиль, тон, повестка дня, политическая цель, содержат ли они факты, мнения или и то, и другое; в основном любая справочная информация, связанная с документом.

Это иногда называют скрытым контентом или «невольным» доказательством. Ответы на эти вопросы могут лежать в самом тексте (например, тип, тон и стиль) или могут потребовать дополнительного изучения (например, информация об авторе или жанре документа). Следующий шаг включает в себя изучение «остроумных» доказательств или содержимого документа.

Есть несколько способов сделать это. Первый заключается в использовании «методики интервью», а второй включает в себя запись событий, метод, сродни формальному структурированному наблюдению:

- **Собеседование.** При «собеседовании» с вашими документами вы, в некотором смысле, рассматриваете каждый документ как респондента, который может предоставить вам информацию, имеющую отношение к вашему запросу. Задаваемые вами вопросы будут зависеть от характера вашего запроса и типа документа. Как и в случае с собеседованием, вам необходимо определить, что именно вы хотите знать, и может ли ваш документ предоставить вам ответы. Затем вам нужно «задать» каждый вопрос и выделить отрывки в документе, которые дают ответ. Организовать ваши ответы можно с помощью цветовой кодировки нужных частей документов, или вы можете обратиться к качественным программам управления данными, таким как Nvivo или NUD \* IST, чтобы помочь вам с индексацией документов.
- **Отмечать частоту.** Отмечать частоту – это процесс, в котором количественно определяется использование определенных слов, фраз и понятий в данном документе. Как и в формальных структурированных наблюдениях, исследователь определяет, что «будет искать», и отмечает количество, частоту и контекст.

Например, вы бы хотели отследить рост изменения климата как точку отсчета в федеральном законодательстве. Сначала вы должны определить, какие законодательные документы вы хотели бы изучить, и убедиться, что у вас есть доступ. Принятие во внимание происшествий будет состоять из поиска фразы «изменение климата» и других связанных терминов, которые вы считаете уместными. Собеседование с документом происходит на месте и включает в себя «задание» соответствующих вопросов документу и изучение их на предмет ответов, чтобы вы могли лучше понять, как и почему используется этот термин.

### *Сбор, проверка и опрос «текстов»*

Чтобы начать процесс опроса текстов, вам необходимо:

1. Соберите соответствующие тексты – большинство текстов, описанных выше, могут быть собраны, но некоторые из них потребуют, чтобы вы вышли в «поле»; Например, посмотреть на граффити, музейные экспонаты или мусор на месте.



2. Организуйте – разработайте и используйте схему организации и управления для собранных текстов.
3. Копируйте – сделайте копии оригинального текста, чтобы делать правки и пометки.
4. Подтвердите подлинность – оцените подлинность и достоверность текста.
5. Изучите повестку дня текста – просмотрите текст и рассмотрите любые присущие ему недостатки.
6. Изучите справочную информацию – извлеките справочную информацию об авторе / создателе, аудитории, цели и стиле в соответствии с исследуемым текстом.
7. Задайте вопросы по тексту – Кто его подготовил? С какой целью его сделали? Когда, где и почему этот текст был изготовлен? К какому виду данных относится данный текст? По сути, вы хотите изучить любую доступную справочную информацию (иногда называемую скрытым контентом или невольным доказательством).
8. Изучите контент – это зависит от типа текста и, как обсуждается в следующем разделе, может включать качественные и количественные процессы. Ключ к успеху здесь – обрисовать то, что вы планируете извлечь из текста заранее.

### Работа с онлайн данными

Очень интересное применение сгенерированных в Интернете данных, это «Интернет, сохраняемый для потомков». С 1996 года Wayback Machine заархивировала более 450 миллиардов веб-страниц и сохранила для потомков два петабайта данных (1 петабайт = 1000 терабайт = 1 000 000 гигабайт). Этот тип данных может быть чрезвычайно полезен для анализа воздействия значительных событий, движений или технологий. Как изменилось финансовое регулирование после 11 сентября или глобального финансового кризиса? Интернет-архивы с открытым исходным кодом содержат резервные копии конкретных веб-сайтов, которые периодически сохраняются в течение последних двух десятилетий. Самое приятное то, что эти веб-страницы не являются статичным снимком экрана, а скорее резервной версией страницы, что означает, что некоторые ссылки по-прежнему могут быть доступны внутри них. Два из лучших источников для интернет-путешествий во времени включают:

- Wayback Machine – Wayback Machine, из интернет-архива, позволяет вам видеть развитие определенного веб-сайта с течением времени, <http://archive.org/web/web.php>.

- Веб-архивы Библиотеки Конгресса – в рамках проекта «Веб-архивы Библиотеки Конгресса» были заархивированы сайты, относящиеся к значимым событиям, таким как атаки 11 сентября, выборы в США и война в Ираке, <http://lcweb2>.

Возможно, эти два сайта могут показаться странными инструментами, правда состоит в том, что возможностей для оригинальных исследований предостаточно. Эти заархивированные веб-страницы могут быть использованы для чего угодно, от аудита маркетинговой стратегии компании (для определения согласованности имиджа бренда и сообщений на их веб-сайте с течением времени) до исследования информации о повестках дня местного сообщества. Архивные веб-сайты с Wayback Machine даже использовались в качестве доказательств в делах коммерческих судов (таких как патентные споры) или для разгадывания загадок. Так же, как детективы могут многому научиться из вашего мусора, так и вы можете извлечь уроки из отброшенных остатков Интернета.

Интернет-архивы, такие как Wayback Machine, являются особенно удобным инструментом для триангуляции. В методах исследования триангуляция определяет местонахождение и проверяет достоверность данных путем перекрестной проверки с другими известными (то есть расположенными) источниками. Например, у вас есть данные из нескольких источников (возможно, из литературы или интервью), которые предполагают, что ранняя маркетинговая стратегия продукта X была откровенно сексистской. Посмотрев на членов совета директоров, перечисленных для компании в этом году, вы можете обнаружить, что женщины не представлены.

### Большие данные

Большие данные, безусловно, имеют большой потенциал. Тенденции, которые они могут отслеживать, значительны во многих отношениях. Например, Австралийское налоговое управление использует анализ больших данных для понимания тенденций уклонения от уплаты налогов. Большие данные также связаны с социальными данными, данными, создаваемыми людьми, которые сознательно и добровольно передают информацию; примеры включают твиты, посты и видео, опубликованные в Twitter, Facebook и YouTube. Беспрецедентное использование социальных сетей означает, что большая часть социальных данных также



является большими данными.

Одна из замечательных особенностей социальных данных – это то, насколько их много и насколько они актуальны. Большие данные могут дать нам моментальный снимок настоящего и помочь заглянуть в будущее. Понимание планов поездок, вспышек гриппа, политических убеждений, реагирования на социальные катастрофы и, фактически, всего, что в настоящее время обсуждает интернет-сообщество, представляет потенциал для использования чрезвычайно большого объема данных. Это, конечно, экономит ваше время и затраты, связанные с вводом в эксплуатацию исследований, которые основаны на сборе первичных данных

## Веб майнинг

Веб-майнинг – это процесс обнаружения закономерностей в больших сетевых наборах данных, включающих методы, которые требуют искусственного интеллекта, машинного обучения, статистики и систем баз данных.

Когда дело доходит до Интернета, на самом деле можно выделить три разные вещи: использование Интернета, которое смотрит на историю пользователей и говорит нам, что люди смотрят в Интернете; веб-контент, который извлекает и интегрирует данные из содержимого веб-страницы; и веб-структура, которая анализирует структуру соединений веб-сайта путем изучения гиперссылок. Из-за структурного разнообразия и постоянно расширяющихся источников становится все сложнее эффективно использовать современные инструменты поиска. Это привело к поиску интеллектуальных веб-агентов, в основном, сложных систем искусственного интеллекта, которые могут автономно извлекать и организовывать веб-информацию.

Преимущества здесь включают информацию, которая может помочь целевому маркетингу. Правительственные учреждения также используют эту технологию для выявления преступной деятельности и даже классификации угроз терроризма. В государственном секторе веб-майнинг может помочь определить общественное восприятие и потребности. В сфере здравоохранения он может выявить как вспышки заболеваний, так и опасности для здоровья.

Однако в качестве исследовательской стратегии существуют методологические проблемы, связанные с качеством и репрезентативностью данных, для которых крайне важна прозрачность; а также проблемы, связанные с этикой. Например, использование майнинга – это вторжение в личную жизнь.

## Изучение интернет-сообществ

Так что же представляет собой онлайн-сообщество и как мы можем его исследовать? Вы знаете, что обычное старое сообщество в те времена было географически связанной группой людей – район, деревня или местная церковная группа. Люди в сообществе собрались в социальной среде и общались лицом к лицу. Общины предлагали отдельным лицам общие ориентиры, поддержку, немного сплетен, дружбу, принадлежность, а также нормы, суждения и даже социальные санкции. Онлайн-сообщества могут предлагать аналогичные социальные «преимущества», но они не связаны географически. Физические локации, такие как городские площади, местные парки или церковные территории, заменяются чатами, группами новостей, электронной почтой, досками объявлений, форумами и иногда, как в случае игр, виртуальными мирами. Это означает, что люди в онлайн-сообществах могут находиться где угодно, и физическая близость больше не является главной общностью. Онлайн-сообщества могут формироваться на основе общих интересов. Так что, будь то Facebook, сайт знакомств, учебная среда или игровой мир, есть новые виртуальные сайты человеческого взаимодействия, которые составляют растущую часть нашего социального мира. И с постоянно совершенствующимися инновациями в онлайн-коммуникации, интернет-сообщества, вероятно, будут расти и приобретать новые измерения.

### *Как лучше всего изучить это явление?*

Онлайн этнография – это попытка понять онлайн-сообщества с точки зрения членов сообщества, основанные на классических антропологических предположениях. Отдельные исследователи разрабатывают методологические подходы, которые обычно включают наблюдение и участие в веб-сайтах, блогах, форумах и социальных сетях.

Netnography – это попытка понять уникальную природу компьютерных интернет-сообществ с точки зрения членов сообщества с помощью новых согласованных стандартизированных подходов к наблюдению



Книга: Инструкция по ведению научного исследования

Лекция: Вторичные данные

Автор лекции: Альмира Абилова

---

и участию в веб-сайтах, блогах, форумах и социальных сетях.

### **Анализ предыдущих исследований**

Систематические обзоры – обзор первичных исследований по конкретной теме, основанный на прозрачных воспроизводимых методах поиска, критической оценки и обобщения результатов и выводов достоверных исследований.

Мета-анализ – статистический анализ и обобщение результатов двух или более первичных исследований, которые одинаково рассматривают одну и ту же гипотезу.