

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**ВОПРОСЫ ИСТОРИИ,
МЕЖДУНАРОДНЫХ
ОТНОШЕНИЙ
И ДОКУМЕНТОВЕДЕНИЯ**

Сборник материалов XI Международной
молодежной научной конференции
(Томск, 08–10 апреля 2015 г.)

Выпуск 11

Т. 1

Научный редактор А.Н. Сорокин

Томск
Издательский Дом Томского государственного университета
2015

Ю.А. Чех

**АКТУАЛЬНОСТЬ ТЕМАТИЧЕСКОГО
КОНТЕНТ-АНАЛИЗА ДЛЯ ОБРАБОТКИ
ОЦИФРОВАННЫХ ИСТОЧНИКОВ (НА ПРИМЕРЕ
ГАЗЕТЫ «СИБИРСКАЯ ЖИЗНЬ» ЗА ПЕРИОД
С 1 ЯНВАРЯ 1905 г. ПО 30 МАРТА 1905 г.)**

Описывается применение тематического контент-анализа для структурирования массивов оцифрованных источников. Статья основывается на практическом применении этого метода. Используются немногочисленные исследовательские работы, посвященные этой тематике.

Ключевые слова: контент-анализ, периодическая печать, крестьянство.

С каждым годом появляется все больше информации, которую, в силу своих особенностей, человеческий мозг не способен усвоить в полном объеме. В большинстве случаев тексты не структурированы, а иногда и не машиночитаемы. В большей степени это касается оцифрованных источников, в нашем случае – газет. Современная пресса является машиночитаемой, и ее легче подвергнуть обработке. Обработку оцифрованных источников нельзя автоматизировать, так как они являются графическим изображением, а компьютер не обладает достаточными навыками, чтобы достоверно структурировать информацию по тематическим блокам. Помимо этого, многие газеты плохого качества: выцвели или затерлись, иногда отсутствуют целые страницы текста. Тематический метод контент-анализа состоит в выявлении явных и скрытых тем (Р. Мертон). Контент-анализ текста (от англ. *contens* содержание) – метод качественно-количественного анализа содержания документов с целью выявления или измерения различных фактов и тенденций, отраженных в этих документах. Структурировать можно вручную, а можно автоматически, но, как говорилось ранее, автоматическое структурирование не подходит для оцифрованных источников. Наиболее оптимальным для анализа оцифрованных источников является создание тематических матриц данных. Создание тематических матриц данных, безусловно, облегчит навигацию по источникам и сэкономит время исследователя. Обработка оцифрованных газет вручную – это трудоемкий процесс, но конечный результат будет востребован многими поколениями ученых.

Для тематического контент-анализа была выбрана газета «Сибирская жизнь» за период с 1 января по 30 марта 1905 г. Было проанализировано 70 номеров данной газеты. Темой анализа являлась крестьянская среда. В каждом номере в среднем 8 тематических разделов (не считая хроники событий на основании полученных телеграмм и объявлений), также присутствуют

статьи вне разделов. В каждом разделе от одной до семи статей, получается, что в среднем в номере 19–20 статей. Однако следует учитывать, что некоторые статьи малы по объему и больше похожи на очерк или справку. Несмотря на это, мною было изучено около 1 400 статей. Крестьянской тематике в газете уделяется достаточное внимание. В матрицу тематического контент-анализа включена 41 статья, которая действительно характеризует крестьян и указывает на особенности их среды. Не во всех номерах присутствует данная тематика. Около 30 номеров не освещали её. Соотношение по месяцам: январь – 12 статей, февраль – 14 статей, март – 15 статей. Так как газета являлась ежедневной, то статьи по нашей тематике появлялись каждую неделю.

Матрица данных включает в себя два блока: источниковедческо-методологический и исторический. **Источниковедческо-методологический блок** включает в себя название и местоположение статьи, ключевые слова (или тематические индикаторы). Индикаторы могут быть контекстно-зависимыми и контекстно-независимыми. Например, слово «манифест» является контекстно-зависимым, так как может распространяться не только на крестьянскую среду, но и на все общество в целом. А вот словосочетание «сельское хозяйство» является контекстно-независимым. Всего был выделен 31 индикатор, в 26 случаях индикаторы были контекстно-зависимы, а в 15 нет.

Исторический блок включает в себя тему статьи, контекст (сюжет и фактаж) и интенцию – мнения и выводы автора статьи. Исходя из категоризации индикаторов по тематическим блокам, было выделено 8 тем: 1. Объединение законов о крестьянах с общим законодательством. 2. Крестьянское просвещение и образование. 3. Развитие сельскохозяйственного производства. 4. Товарно-денежные отношения. 5. Жилищный вопрос. 6. Заболевания. 7. Менталитет. 8. Досуг.

Первый тематический блок **«Объединение законов о крестьянах с общим законодательством»** состоит из 11 статей. Все эти статьи вытекают одна из другой. Освещается правовое положение крестьян, расширение их личных и судебных прав. Также подчеркивается общинный и обособленный характер крестьянского сословия, что мешает реформированию этой среды.

Второй блок **«Крестьянское просвещение и образование»**. Это наиболее обсуждаемая тема, состоящая из 17 статей. У крестьян отсутствует мотивация к получению новых знаний, а если и есть заинтересованные в просвещении, то нет возможности. Библиотек мало, книги поступают редко, а периодические издания и вовсе отсутствуют.

Третий тематический блок **«Развитие сельскохозяйственного производства»** состоит из 4 статей. Авторы отмечают, что технологии прочно внедряются в сельскохозяйственное производство. Это приводит к увеличению урожайности и уменьшению ручного труда.

В четвертом блоке освещаются **товарно-денежные отношения**. В данных газетах это единственная статья, но она является весомой для изучения развития товарно-денежных отношений. Из этой статьи мы узнаем, что у крестьян появилось право непосредственной торговли с крупными фирмами. Также стабилизируются цены на хлеб, которые теперь устанавливает биржа.

Пятый тематический блок **«Жилищный вопрос»** состоит из двух статей, которые рассказывают об ужасных жилищных условиях крестьян. Однако есть и положительный момент. В деревне планировали начать огнеупорное строительство.

В шестом блоке речь идет о **заболеваниях**. Этот аспект крестьянской жизни освещен в трех статьях. Подчеркивается неосведомленность сельских жителей об опасных заболеваниях и нежелание пользоваться услугами врачей. В связи с этим возрастает смертность, особенно среди детей.

Седьмой тематический блок о **менталитете**. Две статьи отнесены к этому разделу. Подчеркивается особая роль огнестрельного оружия, которое буквально считалось необходимой принадлежностью для крестьян. Также освещена взаимовыручка, заложенная в крестьянском сознании.

И последний тематический блок, состоящий из одной статьи, **«Досуг»**. В статье раскрывается феномен съезжего праздника у крестьян. Для него характерны «широкий» разгул, драки и увечья.

Наибольшее внимание уделялось объединению крестьянских законов с общим законодательством и просвещению и образованию. Обзорно давались и другие темы.

Проведенный тематический контент-анализ газеты «Сибирская жизнь» за январь – март 1905 г. способствует объемному представлению о крестьянской среде в то время. Структурное изложение данных упрощает исследовательскую работу, что подтверждает необходимость таких матриц данных. Методы контент-анализа незаменимы в структурировании больших объемов оцифрованных источников. Можно предположить, что в скором будущем методы контент-анализа станут ведущими в структурировании исторических источников.

Литература

1. *Бочаров А.В.* Автоматизация обработки неструктурированной текстовой информации и перспективы гуманитарных наук в XXI веке // Человек – текст – эпоха. Томск : Изд-во Том. ун-та, 2011. Вып. 4. С. 15–37.
2. *Бочаров А.В., Скобелев К.В.* Опыт разработки методики контент-анализа отражения крестьянской жизни в сибирской периодической печати начала XX века (на примере газеты «Восточное обозрение» за январь 1905 г.) // Вестник Томского государственного университета. 2013. № 367. С. 58–63.
3. Сибирская жизнь (Томск). 1905.