

Нижегородский государственный университет им. Н.И. Лобачевского

Национальный исследовательский университет

Учебно-научный и инновационный комплекс
«Социально-гуманитарная сфера и высокие технологии:
теория и практика взаимодействия»

Основная образовательная программа

040100 «Социология», общий профиль, квалификация (степень) бакалавр
Учебно-методический комплекс по дисциплине
«Методы анализа документов»

Основная образовательная программа

040400 «Социальная работа», общий профиль, квалификация (степень) бакалавр
Учебно-методический комплекс по дисциплине
«Методы анализа документов»

Зернов Д.В., Иудин А.А.

ПРИМЕНЕНИЕ НОВЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ АНАЛИЗЕ ДОКУМЕНТОВ В СОЦИАЛЬНОЙ И ГУМАНИТАРНОЙ СФЕРЕ

Электронное учебно-методическое пособие

Мероприятие 1.2. Совершенствование образовательных технологий, укрепление материально-технической базы учебного процесса

Нижний Новгород
2012

ПРИМЕНЕНИЕ НОВЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ АНАЛИЗЕ ДОКУМЕНТОВ В СОЦИАЛЬНОЙ И ГУМАНИТАРНОЙ СФЕРЕ. Зернов Д.В., Иудин А.А. Электронное учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2012. – 60 с.

В учебно-методическом пособии рассматриваются основные приемы анализа документов. Рассмотрены подходы к классификации видов документа и основные направления работы с ними. Излагаются теоретические принципы анализа, а также приводятся пошаговые инструкции работы со специальным программным обеспечением, в том числе, с программными разработками кафедры прикладной социологии ФСН ННГУ им. Н.И. Лобачевского.

Электронное учебно-методическое пособие предназначено для студентов ННГУ, обучающихся по направлению подготовки 040100 «Социология», изучающих курс «Анализ документов», и по направлению 040400 «Социальная работа», изучающих курс «Анализ документов».

ОГЛАВЛЕНИЕ

Введение	4
Раздел 1. Документальные источники в социологии	5
Раздел 2. Основные приемы анализа документов	13
Раздел 3. Контент-анализ документов	17
Раздел 4. Процедура проведения контент-анализа	32
Раздел 5. Факторный анализ	48
Литература	55
Приложение. Зарубежные компьютерные пакеты контент-анализа	59

Введение

Анализ документов является одним из наиболее распространённых методов научно-практического анализа. Существуют науки, которые целиком или в основном опираются именно на анализ текстов. Анализ документов позволяет эффективно работать не только с текстами, но и с другими видами документов (например, визуальными, ауди-визуальными и т.д.). Роль и значение этого метода растет в связи с тем, что текстовой (документальной) информации становится все больше, она все более доступна и может быть проанализирована с использованием компьютерных методов и программ.

Работать с большими массивами текстов, а также с другими видами документов без специального программного обеспечения чаще всего не представляется возможным. Предпринимаются попытки осуществления анализа текстов в универсальных пакетах, таких, например, как word, excel, но эти попытки малоэффективны. Специального программного обеспечения для автоматизированного анализа документов немного и это, преимущественно, зарубежные разработки. Методические рекомендации по работе поставляются только с самими программами, которые очень дороги.

На кафедре прикладной социологии ФСН ННГУ было разработано специальное программное обеспечение для работы с текстовой информацией, которое хорошо себя зарекомендовало. С использованием этой программы, носящей наименование лексико-семантический текстовый анализатор – Лекта (Lekta) уже проведено более пятидесяти полноценных социологических исследований, подготовлено и защищено не менее десяти диссертаций по социологии (в том числе две докторские). С этим программным пакетом активно работают студенты при подготовке выпускных дипломных проектов.

Данная учебно-методическая разработка преследует цель не только дать знания по концептуальным подходам к проблеме анализа документов (а именно на этом делается акцент в современной отечественной учебно-методической литературе по анализу документов), но и позволить студентам приобрести навыки самостоятельной работы с использованием традиционных методов и специального программного обеспечения.

Учебно-методическая разработка «Применение новых информационных технологий при анализе документов в социальной и гуманитарной сфере», прежде всего, предназначена для дисциплины «Анализ документов», но также позволяет лучше усвоить комплекс дисциплин в рамках программ бакалавриата и магистратуры.

Бакалавриата:

- «Методология и методы социологического исследования»;
- «Качественные исследования».

Магистратуры:

- «Современные методы социологических исследований»;
- «Комплексный анализ социальной ситуации»;
- «Инновационные методы в социологических исследованиях».

Разработка связана с решением задачи формирования у студентов *общекультурных компетенций* ОК-11, ОК-12, ОК-13, ОК-14 и *профессиональных компетенций* ПК-1, ПК-2, ПК-3, ПК-8, ПК-10, ПК-11 (*прим.*: номера компетенций соответствуют профилю социология).

Раздел 1. Документальные источники в социологии

Традиционно в отечественной социологии документ определяется как «специально созданный человеком предмет, предназначенный для передачи или хранения информации».¹ В учебной литературе это понятие конкретизируется и дополняется различными аспектами. Например, В.А. Ядов уточняет, что «документальной в социологии называют любую информацию, фиксированную в печатном или рукописном тексте, на магнитной ленте, на фото-, кино- или видеопленке, дискете».² Автор расширяет понятие документ от «специально созданного человеком предмета» до «любой информации», подчеркивая тем самым, что в социологии документ понимается шире его общеупотребительного контекста, как правило, только официальных материалов. Кроме того в своем определении В.А. Ядов не затрагивает область назначения документов, но акцентирует внимание на способах его существования. В учебнике В.И. Добренькова и А.И. Кравченко можно встретить другое определение понятия документ как средства «закрепления различным способом на специальном материале информации о фактах, событиях, явлениях объективной действительности и мыслительной деятельности человека».³ Как видно из определения, приоритеты отданы содержательному наполнению документальных источников информации. Перечисление различных определений документа можно продолжать и далее: «любая знаковая или образная информация, зафиксированная на каком-либо материальном носителе»⁴, «кодированная информация, вбирающая в себя цели и намерения коммуникатора, его видение, интерпретацию реальности»⁵ и т.д. В учебниках по маркетинговым исследованиям метод анализа документов не упоминается. Однако целые разделы посвящены такому методу как анализ вторичной информации (или вторичных данных), под которым подразумевается работа с информацией существующей независимо от исследователя.⁶ Важным моментом является то, что в документе исследователь имеет дело с информацией, отчужденной от ее создателя. Если во время личного общения человек может изменить свою точку зрения, что-то поправить или дополнить, то в документальном источнике мысль уже зафиксирована.

¹ Рабочая книга социолога, 2-е изд. – М.: Наука, 1983. – С. 276.

² Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. 3-е изд., испр. – М.: Омега-Л, 2007. – С. 185.

³ Добреньков В.И., Кравченко А.И. Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2004. – С. 556.

⁴ Оперативные социологические исследования: методика и опыт организации / Под ред. Д.Г. Ротмана, А.Н. Данилова, Л.Г. Новиковой. – Мн.: БГУ, 2001. – С. 148.

⁵ Готлиб А.С. Введение в социологическое исследование: Качественный и количественные подходы. Методология. Исследовательские практики: Учебное пособие. – Самара: Изд-во «Самарский университет», 2002. – С. 353.

⁶ См., напр.: Малхотра Нэреш К. Маркетинговые исследования. Практическое руководство, 4-е изд. – М.: ООО «И.Д. Вильямс», 2007. – С. 173-229.; Черчилль Г., Браун Т. Маркетинговые исследования, 5-е изд. – СПб.: Питер, 2010. – С. 164-179; Токарев Б.Е. Маркетинговые исследования: учебник, 2-е изд. – М.: Магистр : ИНФРА-М, 2011. – С. 21-133; и др.

Одним из возможных определений понятия документа в социологии может стать следующее:

Документ – это ...

- 1) любая информация,
- 2) созданная ранее
- 3) с той или иной целью, и
- 4) зафиксированная на определенном материальном носителе.

Отличительными признаками документа являются: 1) наличие смыслового семантического содержания, 2) стабильная вещественная форма, 3) предназначенность для использования в социальной коммуникации и 4) завершенность сообщения.



Рис. 1.1. Отличительные признаки документа

Документ состоит из двух неотъемлемых слагаемых – информационной и материальной. Лишившись материального основания, документ превращается в устную речь, а лишившись информационного – также перестает быть документов в полном смысле этого слова.

Документ полифункционален. К наиболее общим (характерным для всех видов документов) функциям принято относить:

- 1) информационную – способность служить источником информации;
- 2) коммуникативную – способность быть средством передачи информации;
- 3) кумулятивную – способность накапливать и упорядочивать информацию.

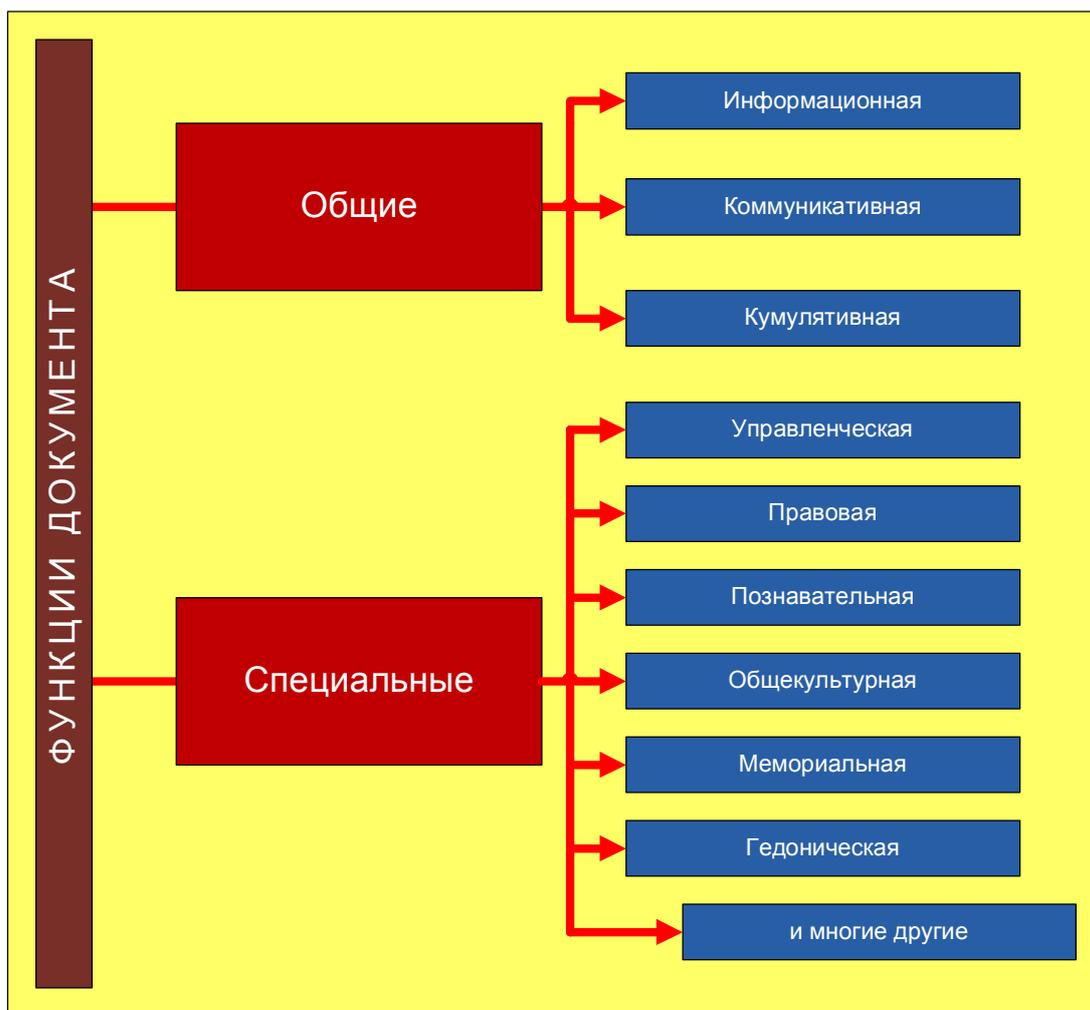


Рис. 1.2. Основные функции документа

К специальным функциям (характерным отдельным видам и типам документов) относятся:

- управленческая (регулятивная) – способность управлять теми или иными процессами (законы, постановления, предписания и др.);
- правовая – это способность служить средством доказательства, подтверждения каких-либо фактов (конституция, законы, указы и др.);
- познавательная (когнитивная) – способность служить средством получения и передачи знаний (учебники, методические материалы и др.);
- общекультурная – способность содействовать развитию культуры, выступать средством закрепления и передачи культурной традиции (литературно-художественные или научные издания, фильмы и др.);
- идеологическая – способность формировать общественное мнение и воздействовать на массовое сознание;
- мемориальная (памятниковая) – способность сохранять информацию и передавать ее от одного поколения к другому (редкие книги или рукописи и др.);
- гедоническая – способность служить средством отдыха, развлечения (художественные книги, фильмы и др.)

К специальным функциям документа можно также отнести статистиче-

скую, источниковую, научно-концептуальную и другие. Количество специальных функций документа очень и очень много.

Это не единственный подход к выделению функций документа.

В частности, метод анализа документов исходит от таких общих функций документа как *трансляция и сохранение сведений о социальной реальности*. Работа с документами позволяет получить информацию о прошедших событиях, лучше понять текущие и даже сделать прогноз о будущих.

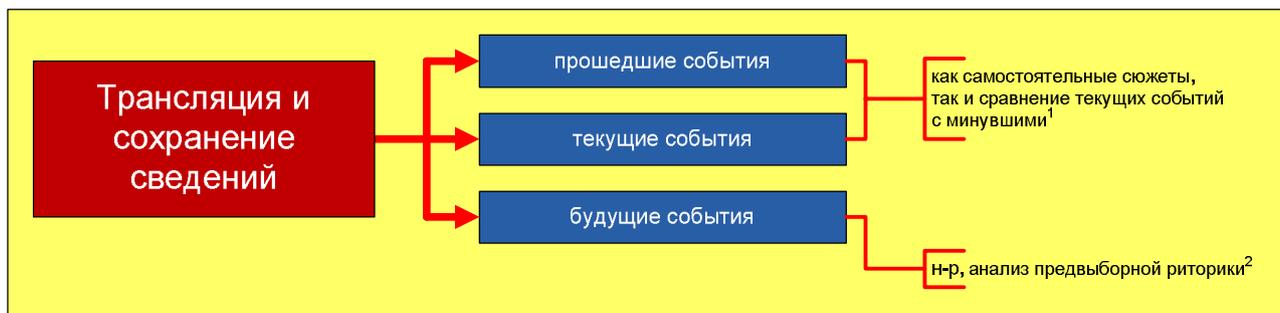


Рис. 1.3. Области применения документа

Коммуникативная функция документа тесно связана, например, со *смыслообразующей функцией*, согласно которой документ обладает семантической автономией от его автора.¹ Интерпретация содержания документа как адресатом, так и исследователем будет происходить с позиций их собственного мировоззренческого опыта. Следовательно, коммуникативное намерение, всегда имеющее в своей основе знаковый характер, может столкнуться с непониманием или с ошибочной интерпретацией.²

Прежде чем приступать к анализу документа необходимо дать ему характеристику, в частности, определить к какому классу и виду он принадлежит. Классификация документов может осуществляться по разным основаниям. Огромное количество самых разнообразных документов приводит к тому, что и оснований для классификаций также немало. Существует отдельная научная дисциплина, такая как документоведение, занимающаяся классификацией документов.³ Но даже в ее рамках озвучивается проблема, что разработка общей

¹ Так, по мнению П. Рикера, «письмо является в этом отношении неким значимым рубежом: благодаря письменной фиксации совокупность знаков достигает того, что можно назвать семантической автономией, то есть становится независимой от рассказчика, от слушателя, наконец, от конкретных условий продуцирования». См.: Рикер Поль Герменевтика и метод социальных наук // http://www.i-u.ru/biblio/archive/riker_germ/

² Подробнее, напр.: Лотман Ю.М. Семиотика культуры и понятие текста // Избранные статьи. Т. 1. – Таллинн: Александра, 1992. – С. 129-132. (<http://www.philology.ru/literature1/lotman-92b.htm>) или Лотман Ю.М. Внутри мыслящих миров // Семиосфера. – СПб.: «Искусство-СПб», 2000. – С.150-390.

³ Бардаев Э.А., Кравченко В.Б. Документоведение: учебник для студентов высших учебных заведений. – М.: Академия, 2010. – 301 с.; Куняев Н.Н., Уралов Д.Н., Фабричнов А.Г. Документоведение: учебник для вузов. – М.: Логос, 2008. – 348 с.; Кушнарченко Н.Н. Документоведение: учебник. – М.: Знание, 2008. – 459 с.; и др.

классификации документов является недостаточно разработанным теоретическим направлением.

Классификация документов осложнена очень большим числом признаков документов, что соответственно приводит к большому числу оснований для деления. Оптимальной классификацией считается дифференциация документов по одному наиболее существенному признаку.



Рис. 1.4. Классификация видов документа (фрагмент)

Так, например, если за основание для деления взять **материальную составляющую документа**, то можно будет выделить группу документов, различающихся по **материальному носителю**, и группу документов, различающихся по **материальной конструкции**. По материалному носителю документы делятся на классы документов на **естественной** и **искусственной** материальной основе. Естественные материалы в качестве носителей информации (глина, камень, а также папирус, пергамент, кожа) сейчас используются очень редко. Основными разновидностями документов на искусственном носителе являются **бумажные и полимерные** (полимерно-пленочные и полимерно-пластиночные) документы. **Бумажный**

документ выполнен на бумаге или на других материалах, по своим свойствам схожих с бумагой (папирус, пергамент или современные синтетические заменители бумаги). К классу бумажных документов относят книги, журналы, газеты, фотографии, нормативные документы и т.д. **Пленочный документ** создан на пленочной основе (кино-, диафильм, диапозитив, микрофиша, магнитная фонограмма, документ с записью информации для ЭВМ). **Пластиночные документы**

созданы на пластмассовой основе (грампластинки, диски). В отдельных случаях они создаются на металлической или стеклянной основе.

Как видно из представленного примера, классификация видов документов достаточно трудоёмкий процесс. Однако для конкретного социологического исследования такая подробная классификация не нужна. Внимание необходимо заострить на признаках документа важных для данного исследовательского проекта. Так, классификация документов по материальной составляющей оказывается менее важной, чем **классификация документов по информационной составляющей**.

По **характеру знаковых средств** документы делятся на:

текстовые; иконографические; идеографические; звучащие; матричные; комплексные.

По **каналу восприятия** на:

визуальные; тактильные; аудиальные; аудиовизуальные.

По **способу документирования** (фиксирования) на:

рукописные; печатные; механические; магнитные; фотографические; оптические; лазерные; электронные.

Эти три подхода довольно близки по смыслу и, отчасти, дополняют друг друга. При этом определить вид документа по этим основаниям нам принципиально важно, т.к. это станет отправной точкой выбора программы анализа документов для исследования.



Рис. 1.5. Классификация документов по их материальной составляющей

Вспомогательное значение будут иметь классификации документов по **степени его распространенности** (опубликованный, неопубликованный, непубликуемый) и **по уровню обобщения информации** (первичный, вторичный, третичный). Хотя отметить это также имеет значение. Достаточно вспомнить, что первичный и вторичный анализ – это два главных метода анализа данных в социологии. Первичной информацией называются данные, полученные на основе непосредственной регистрации совершающихся событий, а вторичной – данные, полученные в результате анализа первичных источников.

Классификация документов по обстоятельству их бытования во внешней среде является принципиальной с точки зрения расчета выборки и определения глубины анализа.¹

Документы различаются: **по статусу** (официальные, неофициальные). К официальным относятся правительственные материалы, постановления, заявления, коммюнике, стенограммы официальных заседаний, данные государственной и ведомственной статистики, архивы и текущие документы различных учреждений и организаций, деловая корреспонденция, протоколы судебных органов и прокуратуры, финансовая отчетность и т.п. Официальные документы создаются организациями или должностными лицами по определенным правилам (стандартам); к официальным документам относят и документы, удостоверяющие личность человека и его права, а также содержащие сведения биографического характера (паспорт, удостоверение личности, водительские права и т.д.). Среди неофициальных документов встречаются личные материалы, а также составленные гражданами безличные документы, например статистические обобщения, выполненные другими исследователями на основе собственных наблюдений.

По содержанию документы делятся на научно-технические (книги, патенты, чертежи и т.д.), правовые (постановления, указы, договоры и т.д.), управленческие (приказы, директивы). Документы могут быть первичными (оригинальный текст документа, книги, статьи) и вторичными (реферат, аннотация). Документы можно также классифицировать по следующим признакам: а) уровень иерархии нормативно-методического обеспечения – международное сообщество, страна, регион (республика, край, область), город, село, фирма; б) правовой статус документа – обязательные к исполнению (законы, стандарты, указы, постановления, положения, программы, планы, формальные распоряжения) и рекомендательные (инструкции, методики, рекомендации и т.п.); в) содержание документа – технические (инвестиционные проекты, конструкторско-технологическая документация, методики), экономические (техико-экономические обоснования, финансовые, бухгалтерские, налоговые документы, бизнес-планы), организационные (организационные проекты, устав, организационные структуры, протокол, инструкции).

По целевому назначению документы подразделяются на целевые и наличные. Целевыми называются документы, подготовленные точно в соответ-

¹ См.: Добреньков В.И., Кравченко А.И. Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2004. – С. 558-559.

вии с программой, задачами социологического обследования: ответы на вопросы анкеты и тексты интервью, записи наблюдений, отражающих мнения и поведение респондентов; справки официальных и общественных организаций, выполненные по инициативе, заказу исследователей; статистическая информация, собранная и обобщенная в ориентации на определенное социологическое исследование. С точки зрения целевого назначения выделяются материалы, которые были спровоцированы, но не созданы самим исследователем. Социолог имеет дело и с материалами, составленными вовсе не ради решения научных задач, а для каких-то иных целей. Их принято именовать наличными документами. К ним относят документы, созданные независимо от исследователя. Их существование ни прямо, ни косвенно не обусловлено техникой проведения социологического исследования: официальные документы, статистические сведения, материалы прессы, личная переписка и т.д. Обычно именно эти материалы и называют собственно документальной информацией в социологическом исследовании.

По степени персонификации документы делятся на личные и безличные. К личным относят карточки индивидуального учета (например, библиотечные формуляры или анкеты и бланки, заверенные подписью), характеристики, выданные данному лицу, письма, дневники, заявления, мемуарные записи. В последние годы появился такой интереснейший и важный источник изучения политической жизни, как документы поименного голосования в представительных органах власти. Безличные документы – это статистические или событийные архивы, данные прессы, протоколы собраний.

Раздел 2. Основные приемы анализа документов

Анализ документов – эффективный и широко применяемый метод сбора социологической информации. Используется как самостоятельный метод, а также активно применяется для решения вспомогательных задач в исследованиях всех видов.

Как вспомогательный метод анализ документов применяется:

1. На этапе подготовки к исследованию
 - 1.1. Создание программы исследования
 - 1.1.1. Описание проблемной ситуации
 - 1.1.2. Предварительный системный анализ
 - 1.1.3. Интерпретация понятий
 - 1.1.4. Выдвижение гипотез
 - 1.2. Создание инструментария
 - 1.3. Построение и расчет репрезентативной выборки



Рис. 2.1. Классификация видов анализа документа

Работа с документами необходима и в тех случаях, когда программой исследования предусматривается построение репрезентативной выборки. При этом выбор документов зависит от того, каким образом строится репрезентативная выборка. Это могут быть простые списки (списки избирателей, отделов кадров и т.д.), если выборка рассчитывается математическими методами, например, как

систематическая, или же документы, специально составленные последователем, в которых отражены некоторые существенные характеристики генеральной совокупности, если выборка строится логическим путем.

2. На этапе осуществления исследования и написания аналитического отчета

2.1. Способ проверки данных, в т.ч. сравнение полученных данных с результатами исследований, проведенных другими людьми

Так, данные, полученные с помощью опроса, касающиеся, например, посещаемости лекций, характера читаемой литературы и т. п., могут быть дополнены такими документами, как самодиагностика, библиотечные формуляры и пр.

2.2. Интеграция данных, полученных в результате анализа документа, в собственный отчет

2.2.1. Цитирование

2.2.2. Слияние результатов своего исследования с чужими проектами.

Специфика анализа текстов как метода раскрывается через пары понятий, описывающих основные контрасты метода.

Дедуция или индукция: от общего к частному или от частных к общим закономерностям. Количественная или качественная стратегия: количественная стратегия предполагает более формальный подход и применение статистических методов, тогда как качественная опирается на способность человека понимать и интерпретировать смыслы.

Денотация и коннотация: денотация и коннотация связаны с социальным значением слов, а не с грамматическими правилами их употребления. Денотация – это фиксированное отношение слова к объектам, которые оно описывает (одно слово может иметь несколько денотаций, например, в языке разных социальных групп). Коннотация – это контекстно-зависимое значение слова или ценностная (оценочная) нагрузка. Примеры: окно – отверстие в стене, форточка – часть окна, но также окно компьютерной программы, окна, форточки – сленговое обозначение ОС Windows (вторичная денотация). Слово социализм или демократия будет иметь разные коннотации для молодого и старшего поколения, как в смысле социального опыта, так и в смысле оценки.

Открытое или скрытое значение, смысл слов – описание и интерпретация: различие открытого, непосредственно выраженного смысла и скрытого смысла, значения. Данное противопоставление напоминает различие открытого и скрытого смысла слов, но относится к текстам в целом. Описательный, дескриптивный анализ концентрируется на таких вопросах, как, например, насколько часто и каким образом слово встречается в тексте, тогда как интерпретационный анализ задается вопросами значения слова и причин его употребления в том или ином контексте.

Важной особенностью этого метода является систематизация большого по объёму тематически связанного, но не структурированного массива (чаще всего текстового). Предварительная систематизация такого материала позволя-

ет сократить время на его обработку. В этой связи существенную важность имеет грамотный подбор источников получения информации, например печатных и электронных СМИ нужных тематических групп, ориентированных на определённые целевые аудитории. Роль и функции их варьируются в зависимости от особенностей освещения экономико-политических аспектов общественной жизни, идеологической, религиозной и многих других её составляющих, по типу социализирующей и образовательной деятельности, по методам воздействия на целевую аудиторию, по степени объективности публикуемой информации. Для исследователя крайне важно чётко идентифицировать позиции медиа при отборе массива, в противном случае противоречивость, размытость результатов могут помешать в полной мере, убедительно подтвердить либо опровергнуть гипотезы исследования.

Под традиционным, классическим анализом документов понимается все многообразие умственных операций, направленных на интерпретацию сведений, содержащихся в документе, с определенной точки зрения, принятой исследователем в каждом конкретном случае. Проведение традиционного анализа означает преобразование первоначальной формы информации, заложенной в документе и далеко не всегда отвечающей запросу исследователя, в необходимую исследовательскую форму. Фактически это интерпретация содержания документа, толкование сути анализируемого материала. Он позволяет улавливать основные мысли и идеи, проследить их генезис, выяснять логику их обоснования, взвешивать следствия. Основной слабостью этого метода является субъективность. Как бы ни был добросовестен исследователь, как бы ни старался он предельно беспристрастно рассмотреть материал, интерпретация всегда в большей или меньшей степени будет субъективна.

Анализ документов в рамках каждого исследования есть самостоятельный творческий процесс, зависящий от формы и содержания самого документа, целей и условий проведения исследования богатства опыта и творческой интуиции исследователя. Однако можно выделить и некоторые общие требования.¹

При проведении традиционного анализа социолог должен ответить на следующие вопросы:

Что представляет собой документ? (Задаваясь этим вопросом, следует обратить внимание на вид, форму документа, решить, потребует ли его анализ применение каких-либо специальных методов и т. п.)

Каков его контекст?

Кто его автор?

Каковы цели, с которыми был создан документ?

Какова надежность самого документа?

Какова достоверность зафиксированных в нем данных?

Каково общественное действие, общественный резонанс документа?

(Этот вопрос вытекает из того обстоятельства, что документы могут представлять собой только свидетельства о каких-либо общественных явлениях, напри-

¹ Рабочая книга социолога, 2-е изд. – М.: Наука, 1983. – С. 290-291.

мер записки политического деятеля о событиях, а могут и сами быть этими явлениями, например публичное выступление того же политического деятеля.)

Каково оценочное содержание документа? (Результатом ответа на этот вопрос должна быть полная ясность относительно того, что именно в документе можно считать выражением оценок мотивов, установок — вообще ценностных представлений, получивших в нем отражение.)

Какие выводы можно сделать о фактах, содержащихся в документе?

Какие выводы можно сделать об оценочных установках, содержащихся в документе?

Какие выводы можно сделать относительно автора документа? Этим вопросом задаются далеко не всегда, а лишь в тех случаях, когда целью исследования является получение представления о личности автора, о ценностных представлениях социальной среды или эпохи, к которым принадлежит автор.

Достаточно ли полны сведения, содержащиеся в проанализированном документе, насколько эти сведения отвечают целям исследования, есть ли необходимость привлекать для дальнейшей работы дополнительный материал?

Перечисленные вопросы служат лишь ориентировочными пунктами для постановки более конкретных вопросов.

В традиционном анализе различают внешний и внутренний. **Внешний анализ** — это анализ исторического контекста документа, т. е. анализ контекста документа в собственном смысле этого слова и всех тех обстоятельств, которые сопутствовали его появлению. Цель внешнего анализа — установить вид документа, его форму, время и место появления, автора, инициатора, цели его создания, насколько он надежен и достоверен, каков его контекст.

Пренебрежение таким анализом во многих случаях грозит ошибочным истолкованием содержания документа. Определенная политическая ситуация может способствовать тенденциозному освещению событий в прессе; требования цензуры могут заставить автора скрыть свои взгляды или выразить их в завуалированной форме; внутренние трения между членами руководства какой-либо организации могут отразиться на содержании документов, описывающих деятельность этой организации.

Внутренний анализ — это исследование содержания документа.

Традиционно считалось, что по существу, вся работа социолога направлена на проведение внутреннего анализа документа, включающего выявление различия между фактическим и литературным содержанием, установление уровня компетенции автора документа в делах, о которых он высказывается, выяснение его личного отношения к описываемым в документе фактам. Однако новые многомерные комплексные методы анализа документов позволяют получить результаты, связанные с внешним анализом. Так, использование этих методов позволяет определить не только позицию автора или издания по исследуемой проблеме, но и обосновывать гипотезы, связанные с обстоятельствами появления документа, его социально-политического наполнения, увидеть цели создания совокупности документов, их достоверность и реальный социальный контекст.

Раздел 3. Контент-анализ документов

Существует множество определений контент-анализа, но при этом большинство из них едва ли полно отражает его сущность. Приведём наиболее часто употребляемые определения контент-анализа.

Количественный анализ текстов и текстовых массивов с целью последующей содержательной интерпретации выявленных числовых закономерностей.¹

Методика объективного качественного и систематического изучения содержания средств коммуникации.²

Систематическая числовая обработка, оценка и интерпретация формы и содержания информационного источника.³

Систематическое изучение объектов (артефактов) или событий посредством их пересчета или интерпретации содержащихся в них тем.⁴

Качественно-количественный метод изучения документов, который характеризуется объективностью выводов и строгостью процедуры и состоит в квантификационной обработке текста с дальнейшей интерпретацией результатов.⁵

Исследовательская техника для получения результатов путем анализа содержания текста о состоянии и свойствах социальной действительности.⁶

Контент-анализ состоит в нахождении в тексте определенных содержательных понятий (единиц анализа), выявлении частоты их встречаемости и соотношения с содержанием всего документа.⁸

Наиболее компактное формальное определение контент-анализа звучит так: «Любая систематическая редукция потока текста (или других символов) к стандартному набору статистически обрабатываемых символов, отражающих

¹ Добренчиков В.И., Кравченко А.И. Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2004. – С. 566. *Отметим, что авторы предварительно не забывают упомянуть о том, что контент-анализ представляет собой перевод в количественные показатели массовой информации (текстовой, аудиовизуальной, цифровой).*

² Джери Д., Джери Дж. Большой толковый социологический словарь. В 2-х томах. Том 1. – М.: «Вече», «АСТ», 1999. – С. 326.

³ Мангейм Джарол Б., Рич Ричард К. Политология. Методы исследования. – М.: Изд-во «Весь Мир», 1997. – С. 270.

⁴ Романов П.В., Ярская-Смирнова Е.Р. Методы прикладных социальных исследований. Учебное пособие. Изд. 2-е, дополненное. – М.: ООО «Вариант» ЦСПГИ, при участии ООО «Норт Медиа», 2008. – С. 76.

⁵ См., напр.: Иванов В.Ф. Контент-анализ как формализованный метод исследования документов // Философская и социологическая мысль. – 1994. – №3. – С. 223-230.

⁶ См., напр.: Таршис Е.Я. Перспективы развития метода контент-анализ // Социология 4М. – 2000 – № 15. – С. 71-92.; *Он же* Исторические корни контент-анализа: Два базовых текста по методологии контент-анализа. – М.: Либроком, 2012. – 160 с.

⁸ См., напр.: Мешков П.Я. Политический мониторинг и контент-анализ в политическом исследовании. // Общая и прикладная политология: Учебное пособие / Общ. ред.: Жуков В.И., Краснов Б.И. – М.: МГСУ; Изд-во «Союз», 1997. – С. 829-837.

присутствие, интенсивность или частоту характеристик, значимых для социальной науки».¹

Эти определения дают фрагментарное представление о методе и не учитывают новых возможностей многомерного статистического анализа. Все эти определения могут быть сгруппированы следующим образом: статистическая семантика; техника для объективного количественного анализа содержания коммуникации; техника для разработки обобщений при помощи объективного и систематического установления характеристик сообщений. Более подробное определение контент-анализа можно встретить в психологической литературе:

Контент-анализ (англ. *content* – содержание) – метод выявления и оценки специфических характеристик текстов и других носителей информации (видеозаписей, теле- и радиопередач, интервью, ответов на открытые вопросы и т.д.), при котором в соответствии с целями исследования выделяются определенные смысловые единицы содержания и формы информации. Затем производится систематический замер частоты и объема упоминаний этих единиц в определенной совокупности текстов или другой информации. Контент-анализ дает возможность выявлять отдельные психологические характеристики коммуникатора, аудитории, сообщения и их взаимосвязи. В отличие от элементарного содержательного анализа, контент-анализ, как научный метод, используется для получения информации, отвечающей некоторым критериям качества (объективность, надежность и валидность). Заметную роль в повышении качества контент-анализа играет возможность использования методов многомерного статистического анализа данных. Особенно широко используется факторный анализ, способствующий выявлению скрытых факторов, определяющих содержание текстов.²

Такое определение несколько громоздко и, по сути, представляет собой описание исследовательской техники. Тем не менее, оно позволяет отойти от представлений о контент-анализе как простом пересчете слов в текстах.

Из истории развития метода контент-анализа

Активное применение метода контент-анализ началось с 1940–50-х гг. Однако первые пробы систематического количественного анализа текстов относят к концу XIX – началу XX столетия, а зарождение метода отдельные исследователи относят к XVIII веку, когда в Швеции был осуществлен анализ сборника из 90 церковных гимнов, прошедших государственную цензуру и приобретших большую популярность, но обвиненных в несоответствии религиозным догматам.³

¹ Определение дано Шапиро и Маркоффом (Shapiro and Markoff) в 1997; (Цит. по: Popping, Roel. Computer-assisted Text Analysis. SAGE Publications: London, 2000, p. 7.)

² См., напр.: Словарь психолога-практика / Сост. С.Ю. Головин. – Минск: Харвест, 1998. – 976 с. (или любое другое издание).

³ См., напр.: Почепцов Г.Г. Теория коммуникации. – М.: «Рефл-бук», К.: «Ваклер», 2001. – С. 382.



Рис. 3.1. Основные этапы истории контент-анализа

Одну из возможных периодизаций развития метода контент-анализа предлагает О.В. Попова.¹ Она выделяет пять этапов становления контент-анализа как самостоятельного метода.

Этап 1. Конец XIX – 30-е годы XX в.: Период разнородного и строгого частотного анализа содержания понятий, в ходе которого были отработаны основные процедуры количественного контент-анализа.

Этап 2. 1940-1950 гг.: Классический частотный контент-анализ Г. Лассуэлла и Б. Берельсона. К этому времени относят первый анализ инаугурационных речей президентов США, пропагандистских речей Гитлера и Рузвельта, радиопрограмм в США, скрытым образом проводивших пропаганду идеологии фашизма.

Этап 3. Конец 1950-х – конец 1960-х гг.: Период усовершенствования техники анализа Ч. Осгудом. Появление программы «Универсальный анализатор» («General inquirer»²), разработчиками которой выступили Ф. Стоун и Ч. Бейлс.

¹ Попова О.В. Политический анализ и прогнозирование: Учебник / О.В. Попова. – М.: Аспект Пресс, 2001. – С. 177-179.

² Подробную информацию о программе можно найти на сайте: <http://www.content-analysis.de/>

Этап 4. 1970-е – середина 1990-х гг.: Развитие ЭВМ приводит к снижению трудоёмкости процедур контент-анализа и даёт возможность обрабатывать большие массивы текстов, сокращает время работы и обеспечивает необходимую надёжность результатов.

Этап 5. Вторая половина 1990-х гг. – наше время: Период активной конкуренции с контент-анализом других методов работы с документами.

В СССР метод контент-анализа стал использоваться с середины 1960-х годов. Например, это исследования А.В. Баранова, направленные на изучение степени обращения к субъективным интересам читателей в газете «Известия» или В.З. Когана и Ю.И. Скворцова об информационном воздействии на читателей газеты «Труд».¹

Виды контент-анализа

Существует несколько классификации видов контент-анализа. Например, одну из них предлагает Р. Мертон²:

1. подсчет символов (простой подсчет определенных ключевых слов);
2. классификацию символов по отношению (баланс положительных и отрицательных высказываний по поводу объекта исследования); используется для анализа эффективного расположения символов для пропаганды, для обнаружения контрастных, противоречивых суждений и для определения намерений коммуникатора;
3. анализ по элементам (выбор главных и второстепенных частей текста, определение тем, связанных с основными и периферийными интересами аудитории);
4. тематический анализ (выявление явных и скрытых тем);
5. структурный анализ (выяснение характера соотношения различных материалов: взаимодополняющего, объединенного, сталкивающего);
6. анализ взаимоотношения различных материалов (сочетание структурного анализа с изучением последовательности публикации материалов, объема и времени выхода их в свет).

Контент-анализ также принято разделять на:

- **Количественный и качественный.** Количественный контент-анализ характеризуется частотой появления в тексте определенных характеристик. Качественный контент-анализ, ориентирован на учет сочетания качественных и количественных показателей, наиболее эффективен для выявления явных или скрытых целей субъекта.
- **Ненаправленный и направленный.**³ Ненаправленный контент-анализ основывается на гипотезе, что некоторые слова текста, названные репрезентативными, могут быть репрезентативны по отношению ко всему тек-

¹ См.: Рабочая книга социолога, 2-е изд. – М.: Наука, 1983. – 480 с. – С. 301-302.

² Цит. по: Попова О.В. Политический анализ и прогнозирование: Учебник / О.В. Попова. – М.: Аспект Пресс, 2001. – С. 179.

³ Подробнее с ненаправленным и направленным контент-анализом можно ознакомиться в: Боришполец К. П. Методы политических исследований: Учеб. пособие для студентов вузов. – М: Аспект Пресс, 2005. – С. 51-66.

сту. Эти слова должны часто встречаться в тексте и не носить функционального характера. Направленный контент-анализ предусматривает предварительное составление перечня понятий, конкретизирующих каждую отдельную категорию.

- **Фронтальный и рейдовый.** Фронтальный контент-анализ ориентирован на составление максимально подробного представления об информационном потоке в определенный момент времени или на протяжении некоторого периода с целью выявления содержательной динамики. Фронтальный контент-анализ носит прикладной характер. Рейдовый анализ ориентирован на решение частных исследовательских задач.
- **Тематический и семантический.** Основной отличительной особенностью *тематического* подхода является то, что выводы делаются на основании данных о встречаемости концептов, понятий в тексте. Основное допущение подхода состоит в том, что существует связь между наличием в тексте тех или иных тем и интересом к этим темам у автора текста. Семантический анализ в значительно большей мере, чем тематический, опирается на свойства естественного языка. Поэтому анализ (особенно компьютерный) в значительной степени специфичен для каждого языка. Каждый блок текста, попавший в выборку, суммируется при помощи нескольких кодов, взаимосвязанных в соответствии с общим пониманием предмета изучения.

Существуют и другие деления, например, в зависимости от характера гипотезы контент-анализ делят на поисковый и контрольный, в зависимости от специфики применения – на непосредственный и косвенный.

Современное программное обеспечение контент-анализа позволяет осуществлять сочетание различных методов анализа. Так, полнопоточный анализ текстового массива (например, публикации газет за продолжительный период времени) позволяет осуществлять одновременно и направленный, и ненаправленный, фронтальный и рейдовый, тематический и семантический анализ одновременно.

В настоящее время контент-анализ является междисциплинарным методом анализа информации. К нему регулярно обращаются социологи, политологи, психологи, историки и филологи. Контент-анализ обычно применяется при наличии обширного по объему и несистематизированного текстового материала, когда непосредственное использование последнего затруднено. Эта методика является особенно полезной в тех случаях, когда категории, важные для целей исследования, характеризуются определенной частотой появления в изучаемых документах, а также тогда, когда большое значение для исследуемой проблемы имеет сам язык изучаемого источника информации, его специфические характеристики.

Контент-анализ занимает особое место среди других в силу своей эффективности при анализе больших информационных массивов. Чаще всего он используется при анализе текста и заключается либо в подсчёте наиболее часто встречающихся в нём слов, словосочетаний, самостоятельных тем, выражен-

ных, например, целостными абзацами, и других лексических единиц, либо единицами контент-анализа выступают такие величины как протяжённость текста, численность строк, абзацев, колонок, страниц. Метод также применяется и при изучении видео и аудио материала и единицами анализа становятся графическая составляющая, сопровождающая тексты, метраж аудио и видео плёнки с материалами, интересующими исследователя, объём эфирного времени, время суток, в которое материал транслируется аудитории. С помощью этого метода можно изучать такие материалы как, например, статьи в СМИ, речи политиков, партийные программы, программы общественных движений, видеоматериалы массовых мероприятий, съездов и митингов, нормативно-правовые акты, рекламные сообщения, произведения художественной литературы, исторические тексты, письма и многое другое. Обязательным условием проведения контент-анализа является фиксация материала на материальном носителе. Только при его соблюдении возможно использование этого метода.

С помощью анализа текстов могут быть протестированы три типа гипотез:

1. гипотезы относительно частоты встречаемости тех или иных терминов, понятий;
2. гипотезы о связи понятий в тексте, отдельных частях текста или совокупностях текстов;
3. гипотезы, касающиеся соотношения между текстуально-аналитическим исследованием и другими видами исследований; гипотезы такого типа используются для сравнения результатов исследований, проведенных с помощью различных методов или для установления связей между текстуальными и не-текстуальными явлениями (например, для сравнения высказываний и реальных действий людей).

Часто результаты контент-анализа дополняются использованием других методов. Интересен он также и тем, что не требует больших материальных затрат, несложен в использовании, не подразумевает ощутимых технических и других трудностей при использовании специализированного компьютерного программного обеспечения. Полевой этап исследования более прост, чем при использовании многих других методов. Так, проведение простого (хотя и неглубокого) контент-анализа доступно даже при использовании базовых средств Microsoft Office или его аналогов.

Ограничения анализа текстов как метода:

- для количественного анализа необходимо статистически значимое количество текстуальной информации, он не предназначен для анализа уникальных текстов;
- анализируемые тексты должны поддаваться формализации, поэтому данный метод лишь ограниченно пригоден для анализа художественной литературы и совсем не пригоден для анализа поэзии;
- качественный анализ позволяет глубже понять текст, но он требует значительного количества времени и усилий; таким образом, традиционный качественный анализ малоприменим для исследования больших объемов текста. Последнее ограничение ныне снимается посредством создания

программных средств, осуществляющих лексический анализ текстов. В последние годы предпринимаются попытки и семантического машинного анализа вербальной информации;

- главным ограничением является то обстоятельство, что текст менее сложен, чем индивидуальное или общественное сознание, которыми он порожден; текст является упрощенным, редуцированным отражением социальной реальности.

Основные методологические категории метода контен-анализа

Контент-анализ как метод предоставляет исследователю богатые и разнообразные возможности, но требует тщательного формирования исследовательской стратегии путем выбора из нескольких альтернатив. Рассмотрим эти альтернативы.

Основа контент-анализа – это подсчет встречаемости некоторых компонентов в анализируемом информационном массиве, дополняемый выявлением статистических взаимосвязей и анализом структурных связей между ними, а также снабжением их теми или иными количественными или качественными характеристиками. Отсюда понятно, что главная предпосылка контент-анализа – это выяснение того, что считать; иными словами, определение единиц текста.

Единицы текста. Единица – это отдельная группа слов, рассматриваемая как целое. Выделяется несколько типов единиц.

Единицы анализа – это единицы, составляющие основу анализа, единицы, которые исследователь стремится охарактеризовать. Пример: слово, газетная статья.

Единицы выборки – части наблюдаемой реальности или потока текста, которые рассматриваются как независимые друг от друга. Они имеют ясно различимые границы, им могут быть присвоены уникальные номера и они могут включаться в выборку с заранее известной вероятностью.

Единицы кодирования (также единицы записи или единицы текста) – это отдельные сегменты текста, помещаемые в ту или иную категорию. Для каждой единицы кодирования исследователь принимает решение, имеет ли она те или иные атрибуты, которые интересуют его в данном исследовании, относятся ли они к теме исследования. Пример: идея превосходства мужчин над женщинами (идея, формирующая категорию) может быть выражена в таких единицах кодирования, как слово, смысл слова, предложение, тема, абзац, текст целиком.

Единицы контекста – это та совокупность текстов, которую необходимо принять в расчет, характеризуя единицу кодирования. Они формируют контекст, который определяет значение, смысл единиц кодирования, в том случае, если этот смысл контекстно-зависим. Например, в статье, посвященной финансовым вопросам, слово долг будет иметь другое значение, чем в тексте, посвященном религиозным вопросам. При анализе текстов без применения компьютера контекст обычно легко распознаваем. В компьютерном анализе контекст, как правило, определяется через анализ слов, окружающих в тексте единицу кодирования.

Единицы счета – это те единицы, с помощью которых квантифицируются атрибуты текста. Они совпадают с единицами кодирования, если исследователь

заинтересован в подсчете количества слов или других элементов текста. Другими словами, единицы счета – это именно то, что подсчитывается в процессе исследования, то, к чему относятся числа в матрице данных. Примеры: 5 слов были идентифицированы как относящиеся к агрессии (попадающие в данную категорию). В матрицу ставится число 5 – в данном случае единица кодирования совпадает с единицей счета. Пример несовпадения этих единиц: анализ пространства на страницах газеты, отданного под освещение определенной темы. Статья, идентифицированная как относящаяся к теме – это единица кодирования, а число квадратных сантиметров (в которых измерена площадь статьи и полученный результат занесен в матрицу) – единица счета.

Физические единицы имеют отдельную физическую форму (например, отдельный номер газеты или газетная полоса).

Синтаксические единицы – те, которые являются естественными для грамматики соответствующего средства коммуникации (например, слово во фразе или отдельная новость во фразе выпуска новостей). *Единицы референции* – те, которые описывают разными словами один и тот же объект (например, «глава государства», «президент», «Путин», в определенном контексте – просто «он»). *Пропозиционные единицы* – это части сложных предложений, имеющие собственную структуру, описания конкретных положений дел (ситуаций). Такие единицы используются для того, чтобы избежать сложности естественного языка. Например, фраза «Агрессивный вор угрожает полицейскому» распадается на два простых предложения «Вор агрессивен» и «Вор угрожает полицейскому».

Единицы различного рода могут пересекаться и включать друг друга. Например, при анализе книг первая единица анализа – это книга, вторая – главы в книгах, третья – параграфы или абзацы. В случае если параграф – наименьшая из единиц, на которые исследователь разбил текст, он также служит и единицей кодирования. Однако можно продолжить делить текст дальше вплоть до предложений или грамматических частей предложений. В таком случае единицей выборки может стать абзац. Каждая единица, которая больше, чем составляющие ее части, может служить единицей контекста: фраза для слова, глава для параграфа и т.д.

Удобную схему классификации единиц контент-анализа предлагает О.В. Попова¹:

¹ См.: Попова О.В. Политический анализ и прогнозирование: Учебник / О.В. Попова. – М.: Аспект Пресс, 20011. – С. 181.

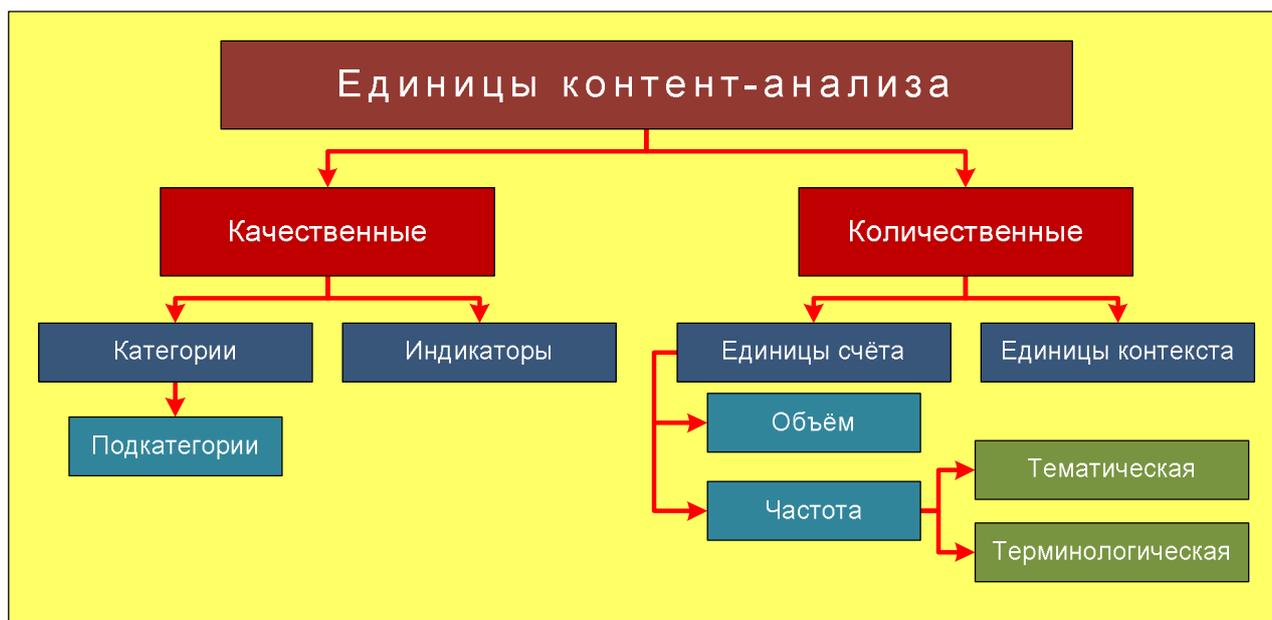


Рис. 3.2. Классификация единиц контент-анализа

Кроме единиц контент-анализа используются и **концептуальные категории**. *Концепт* – это единица смысла, отдельная идея. Концептуальные категории – это агрегации единиц текста, основанные на общей идее, релевантной для теоретической основы исследования. Иными словами, категории – результат операционализации идей с помощью слов и фраз. Концепты могут быть образованы дедуктивно (на основе теории) или индуктивно (на основе исследуемых текстов).

Методологические основания кодирования текста. *Кодирование* – это процесс систематической трансформации и агрегации исходных данных в категории, которые позволяют точно описать характеристики текста, релевантные для исследования. Основная проблема кодирования – неоднозначность текста. Слова могут иметь различные коннотации. Способ избежать проблем неоднозначности – выработка детальных правил кодирования и принятие мер к тому, чтобы избежать неоднозначности, основанной на контексте. Кодирование может производиться как вручную, так и с использованием компьютера. Каждый способ имеет свои преимущества и недостатки. Люди более способны к совершению осмысленного выбора из нескольких значений или вариантов смысла, но они работают медленно и иногда склонны присваивать словам значения, не предусмотренные исследователем. Компьютерное кодирование значительно быстрее, но хуже справляется с неоднозначностью текста.

В рамках инструментального подхода набор категорий создается на основе теории, которая есть у исследователя еще до начала исследования. В таком случае, целью исследования может быть подтверждение или опровержение теории. Важно, что в данном случае значения, которые не релевантны по отношению к теории, а также варианты значений, возможно подразумеваемые автором текста, не принимаются во внимание.

Выделяется два базовых подхода к кодированию: кодирование открытого, эксплицитно выраженного содержания текста (манифестное или инструментальное) и кодирование скрытых, неявно выраженных смыслов, которые, по мнению исследователя, подразумевались создателем текста (латентное, репрезентационное).¹

Манифестное кодирование. Кодирование содержания текста, лежащего на поверхности, называется манифестным (открытым, явным). Исследователь подсчитывает количество появлений фразы или слова (например, красный) в письменном тексте; или количество определенных действий (например, поцелуев или ударов), представленных на фотографиях или в видеосцене.

Система кодирования включает перечень терминов или действий, помещенных в текст. Именно таким перечнем является, например, словарь. Исследователь может использовать компьютерную программу (например, Lekta) для поиска слов или фраз и, таким образом, переложить на компьютер всю работу по кодированию. Для этого исследователю нужно изучить компьютерную программу, составить список соответствующих слов или фраз и затем представить текст в форме, которая может быть пригодна для компьютера.

Манифестное кодирование весьма надежно, поскольку фраза или слово может либо наличествовать, либо отсутствовать. К сожалению, манифестное кодирование не принимает в расчет коннотации слов или фраз. Одно и то же слово может иметь различные значения в зависимости от контекста. Возможность каждого слова выступать во множестве значений ограничивает валидность манифестного кодирования.

Латентное кодирование. Исследователь, который использует латентное (скрытое) кодирование, ищет скрытые, имплицитные значения содержания текста. Например, он прочитывает весь абзац целиком и решает, присутствует ли в его содержании эротика или же это романтический жанр. Применяемая им система кодирования должна следовать общим правилам, устанавливающим принципы интерпретации текста и определяющим, имеют ли место те или иные темы или жанры.

По сравнению с манифестным латентное кодирование менее надежно. Оно зависит от степени владения исследователем языком и общепринятыми значениями. Здесь очень важной становится позиция автора по отношению к исследуемой проблеме. Повысить надежность могут тренинг, практика и описание правил, но остаются трудности, связанные с правомерностью идентификации тем, жанров и т.д. Латентное кодирование может быть валиднее манифестного, поскольку люди передают значение множеством неявных способов, прежде всего, в зависимости от контекста, а не от тех или иных слов.

С технической точки зрения, исследователь может разработать систему кодов для латентных смыслов, содержащихся в тексте. Затем в конце каждого фрагмента после текста вставляются кодовые обозначения. Например, крайне негативный, критический настрой автора текста, отраженный в отдельном

¹ См., напр.: Ньюман Л. Неопросные методы исследования // Социологические исследования. – 1998. – № 6. – С. 123.

фрагменте, может быть обозначен кодом «НГТ». Степень негативности может быть передана неоднократностью кода НГТ. Коды затем учитываются при обработке текста как отдельные фильтры.

Исследователь может использовать и манифестное, и латентное кодирование. Если в использовании двух подходов нет противоречий, результат будет сильнее; если же манифестное и латентное кодирование используются несогласованно, исследователю может потребоваться перепроверить операциональные и теоретические дефиниции.

Итогом проведения контент-анализа как самостоятельного или дополняющего метода является аналитическая записка, интерпретирующая полученный материал, классифицированный по семантическим группам. Она включает в себя либо только общие выводы, вытекающие из проведённого анализа, либо также подробную информацию, описывающую изучаемые массивы.

Наряду с очевидным достоинством метода контент-анализа – возможностью его применения при анализе больших массивов неструктурированного текста – существуют и недостатки. Чаще всего выделяют высокую вероятность субъективной интерпретации полученных данных, субъективизм при отборе исходных данных, категорий. Обычно проблема субъективной трактовки полученных данных решается коллективным участием в работе нескольких исследователей.

В процессе работы над контент-анализом может вызвать сомнение выбор объёма массива, достаточного для определения зависимости между категориями и индикаторами, транслирующими семантическую нагруженность конкретной части массива. Следует помнить, что контент-анализ неразумно и неэффективно использовать при изучении уникальных несхожих между собой документов, для исследования которых необходимо получить всестороннее и полное их описание, не допуская игнорирования тех или иных уникальных, а потому не встречающихся постоянно в массиве элементов. Это также справедливо и для анализа весьма сложных документов и документов, в которых явно недостаточно материала для проведения контент-анализа, результаты которого не будут репрезентативны.

Отбор источников и построение выборки. Как уже было сказано ранее, этап определения фильтров поиска полезных в исследовании источников информации и отсеивание ненужных, несвязанных с темой, либо не имеющих весомых связей с ней, является очень важным и не должен рассматриваться как формальный, механический и малозначимый этап работы. Ошибки при выборе типологических групп источников могут отрицательно сказаться на всех этапах последующего исследования, включая получение необъективных его результатов. В связи с этим необходимо прежде всего определить круг источников потенциально полезных для исследования, содержащих в себе материалы по заданной теме. Далее важно установить дополнительные рамки отбора материала: определить тип источника (телевидение, пресса, рекламные материалы, радио и др.). Затем нужно определить вид сообщения (публицистические статьи в электронном либо в печатном СМИ, информационные заметки, рекламные пла-

каты) роль участника коммуникации (отправитель или получатель сообщения). Определяются минимальные и максимальные границы объема сообщений, их протяженности, частота, время, место и средство трансляции сообщений целевой аудитории. Существуют и другие критерии отбора сообщений, и их количество и выбор варьируется в зависимости от поставленных задач исследования.

Далее следует этап определения объема выборочной совокупности. В случае ограниченного количества материала по заданной теме, выборочная совокупность может быть эквивалентна генеральной. Это актуально, например, при предварительном проведении интервьюирования на заданную узкую тему, при котором весь массив текстов будет использоваться для анализа. Классическая трактовка метода контент-анализа подразумевает возможность сокращения выборочной совокупности сообщений при их схожести и однородности в соответствии с вышеописанными критериями. Такая редукция допустима, если объем генеральной совокупности очень велик. Выборка при исследовании больших совокупностей данных случайная и производится также в соответствии с заданными вышеуказанными критериями. Безусловно, необходимо рассчитать её объем так, чтобы она оставалась репрезентативной, важно определить допустимую ошибку выборки. Техническое задание исследования должно содержать такие критерии сбора материала, регламентируя этот процесс и не давая нежелательному (ненужному или деструктивному для исследования) тексту проникнуть в массив. Для определения конгруэнтности особенностей представления информации в СМИ, выбранных исследователем для проведения контент-анализа, и конкретных сообщений, потенциально попадающих в выборку, может быть произведен эксперимент. Его целью будет определение степени точности эмпирической интерпретации категорий исследования. Стоит добавить, что зачастую объем выборочной совокупности объясняется исследователями, исходя из понятий здравого смысла, доступности материала, скорости анализа материала в сжатые сроки, а не расчетом допустимой ошибки выборки, не достаточностью массива для сохранения его репрезентативности.

Для того чтобы применение контент-анализа было успешным, источник должен отвечать определенным требованиям. При выборе источника, прежде всего, нужно определить, в какой мере его содержание соответствует поставленной задаче. Необходимо также изучить все существующие источники по данной проблеме и, если понадобится, выявить оптимальный размер репрезентативной случайной выборки. При построении выборки необходимо учитывать уровень исследуемой проблемы, цели и задачи исследования, ресурсоемкость (затраты труда, времени и средств) построения выборки и последующего проведения исследования на ней.

Пример построения выборки.¹ Например, перед исследователем стоит задача узнать, как изображаются женщины и представители меньшинств в американских еженедельных журналах. В качестве единицы анализа избирается ста-

¹ *Подробнее см.:* Ньюман Л. Неопросные методы исследования // Социологические исследования. – 1998. – № 6. – С. 124-128.

тья. Генеральная совокупность (популяция) включает все статьи, опубликованные в Time, Newsweek, U.S. News & World Report между 1969 и 1989 гг. Сначала нужно проверить, издавались ли названные журналы в указанные годы и определить, что понимается под статьей. Например, являются ли статьями обзоры кинофильмов? Можно ли определить минимальный размер текста (например, текст, состоящий из двух предложений), позволяющий квалифицировать его как статью? Если статья состоит из нескольких частей (и печатается в нескольких номерах), следует ли рассматривать эти части как отдельные статьи, или же как одну? Исследование указанных трех журналов показывает, что в среднем каждый номер содержит 45 статей. В год издавалось 52 еженедельных номера. Учитывая 20 лет определенных временных рамок, генеральная совокупность включает приблизительно 140 000 статей ($3 \times 45 \times 52 \times 20 = 140\,400$). Рамочные параметры для выборки задаются перечнем всех этих статей. Затем нужно принять решение об объеме и виде выборки. Допустим, что исходя из размеров бюджета и времени выборка ограничивается 1400 статьями. Таким образом, пропорция выборки составляет 1%. Необходимо также избрать вид выборки. Систематизированная выборка не подходит, поскольку журнальные издания выходят в свет циклично (интервал между выходом каждого из 52 номеров на протяжении каждого года – всегда неделя). Все номера важны для исследования, поэтому используется стратифицированная выборка. Стратификация проводится по журналам: $1400 / 3 = 467$. Выборка стратифицируется также и по годам. Результат – примерно 23 статьи из каждого журнала за год. Наконец, составляется случайная выборка с использованием таблицы случайных чисел, чтобы отобрать 23 номера для 23 выбранных статей из каждого журнала за каждый год.

Организационные моменты проведения контент-анализа на этапе сбора информации. Традиционные методы работы с текстами подразумевают, что организаторы работы весьма тщательно должны работать с бригадой кодировщиков, ибо именно они выполняют весь объем технической работы, от которой зависит качество всего исследования. Контент-анализ часто включает кодирование очень большого круга источников информации. Исследовательский проект может потребовать просмотра содержания нескольких десятков книг, сотен часов телевизионных передач или тысяч газетных статей. В этой работе часто используется помощь ассистентов, предварительно обученных правилам ведения записей и методике кодирования. Кодировщики должны владеть моделями системы кодирования и консультироваться по всем вопросам неоднозначного характера. Исследователь обязан фиксировать все решения, которые он принимает относительно того, как трактовать возникшую при кодировании новую ситуацию.

1. Определение круга кодировщиков. В соответствии с поставленными проблемами, сроками, ресурсами рассчитывается число работников, привлекаемых к кодированию, обработке материала. Содержание в штате исследовательской лаборатории постоянных кодировщиков малоэффективно. Целесообразнее привлекать временных исполнителей. Это могут быть работники биб-

лиографических отделов, в случае анализа статей, ведь именно они разносят карточки по различным направлениям, работники службы отдела кадров, если ведется анализ аттестационных характеристик, работники канцелярий и отделов по связям с общественностью.

2. *Организация кодирования.* Инструкция по кодированию должна быть написана языком, доступным кодировщикам. Инструктаж должен быть проведен как в устной, так и, что особенно важно, в письменной форме. Это позволит устранить элементы недопонимания между методологами, аналитиками и исполнителями. Важны и моральные стимулы: технические работники были в курсе целей, задач, важность и назначение исследования. Поэтому в ходе беседы перед кодировщиками должны быть раскрыты основные задачи и гипотезы исследования. Необходимо наладить тесный контакт с кодировщиками и разработчиками инструментария по вопросам толкования различных моментов текста.

3. *Мотивация труда кодировщиков.* Здесь особенно важны материальные стимулы, их достаточность и справедливость распределения средств. В связи с этим могут использоваться различные варианты оплаты: по затратам времени; сдельная, за каждую найденную статью; пропорции к обработанному материалу. Могут быть введены поощрения в виде премий за скорость обработки информации; качество обработки; творческий подход к делу.

4. *Контроль результатов кодирования.* Необходимо осуществление выборочного контроля. Из бланков первичного кодирования исследователь отбирает несколько бланков, отличающихся от среднестатистического объема заполнения, например тем, что: а) в клеточках преобладают нули; б) все клеточки заполнены.

Как правило, наименования, газеты, даты выхода и наименования статьи заносятся в бланк полностью, что значительно облегчает процедуру контроля.

Исследователь, который пользуется помощью нескольких кодировщиков, должен всегда проверять однозначность кодирования. Для этого он просит кодировщика закодировать текст самостоятельно и затем сопоставить полученные результаты с уже имеющимися по всему тексту. Таким образом замеряется надежность воспроизведения информации, полученной другими кодировщиками, что является типом эквивалентной надежности со статистическим коэффициентом, который передает степень согласованности действий кодировщиков. Этот коэффициент приводится в отчете по результатам контент-анализа.

По прошествии известного времени (например, трех месяцев) исследователь также проверяет, насколько устойчива надежность взаимодействия, для чего каждый кодировщик заново самостоятельно кодирует текст, который он уже кодировал ранее. На основании полученных результатов исследователь делает вывод о том, сохраняется ли стабильность кодирования. Например, шесть часов телевизионных эпизодов кодировались в апреле, а затем подверглись новой кодировке в июле, при этом кодировщик не имел возможности пользоваться полученными ранее результатами. Любое значительное отклонение обязывает переобучить кодировщика и повторно провести кодирование текста.

Следует отметить, что изложенные здесь приемы и методы кодирования в настоящее время несколько устарели. Кодирование текстов с использованием даже самых распространенных компьютерных программ позволяет значительно упростить указанные процедуры. Это обстоятельство тем более важно, что сегодня практически любое издание, любой текст можно найти в оцифрованном варианте. Старые тексты, которых нет в сети, можно легко отсканировать.

Специфика работы с данными на электронных носителях. В том случае, если анализируемые документы имеют электронные копии, процесс анализа значительно упрощается, но необходимо не забывать о разнице между возможностями человека и машины. Так, например, с применением компьютера более уместен анализ документов управления, где не нужны эмоциональные оценки и большее внимание может быть уделено анализу лексики.

Особое внимание здесь следует уделить установлению однозначного синонимического отношения. Например, дать программе команду считать синонимами или индикаторами одного и того же явления в рамках поставленной задачи слова, например, нормативы и тарифы.

Все определяется постановкой задачи; если, например, анализируется стиль управления государственным имуществом, то в рамках гипотезы могут быть рассмотрены два направления – рыночное и директивное, где рыночному соответствуют термины: цена спроса, цена предложения, конъюнктура; а директивному – нормативы отчислений, тарифная база. Если же исследование текста идет с позиций анализа состояния рынка, то термины спрос и предложение не могут быть использованы в качестве единой смысловой единицы.

Раздел 4. Процедура проведения контент-анализа

Процедура проведения контент-анализа

Как уже отмечалось выше, существует немало подходов, методов и методик осуществления контент-анализа, поэтому процедуру лучше всего рассматривать на примере конкретного подхода. Поэтому подробнее остановимся на одном из подходов к анализу документов и рассмотрим его практическую реализацию. Далее речь пойдет о контент-анализе, сочетающем качественную и количественную стратегии (в том числе, применение методики факторного анализа). Единицей анализа является фильтр – семантическая цепочка, состоящая из некоторого количества лексем.¹

В процедуре контент-анализа можно выделить несколько этапов:

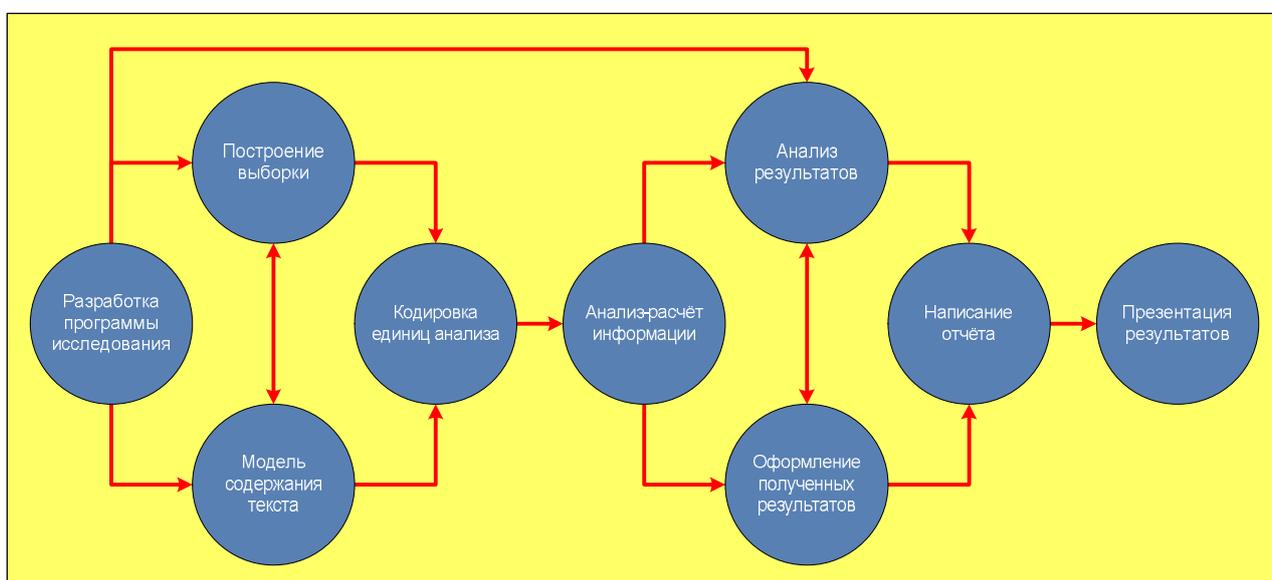


Рис. 4.1. Сетевая схема организации контент-анализа

1. Разработка программы исследования (цели, задачи, гипотезы). Этот этап работы определяет срезы содержания. На этом этапе, как правило, формулируется т.н. эмпирическая теория исследования. То есть, в ходе подготовки к проведению контент-анализа, ученый систематизирует гипотезы, существующие в контексте данной проблематики и отсеивает те из них, которые не поддаются верификации на данных информационного массива.
2. Построение выборки документов на основе определения общей совокупности, какие документы являются носителями необходимой информации.

¹ *Лексема* [*гр. lexis* слово, выражение, оборот речи] – *лингв.* единица словаря языка; в одну лексему объединяются разные парадигматические формы одного слова и разные смысловые варианты слова, зависящие от контекста, в котором оно употребляется.

- Определение круга и объема документов, являющихся носителями необходимой информации (наименование, периодичность выхода, период, тиражи).
 - Построение выборки: какие документы по каким критериям будут привлечены для анализа.
 - Анализ правильности построения выборочной совокупности.
3. Моделирование содержательного плана текста.
 - Классификация социальных ситуаций, соответствующих рассматриваемому кругу проблем.
 - Определение набора единиц анализа.
 - Проверка надежности методики.
 4. Кодирование единиц анализа.
 5. Проведение непосредственного анализа-расчета информации – сбор информации.
 6. Анализ результатов.
 7. Оформление полученных результатов.
 8. Написание отчета.
 9. Презентация результатов.

Построение выборки и моделирование содержательного плана текста являются параллельными этапами исследования, взаимообуславливающими друг друга. Следует помнить, что проблемы, возникающие при кодировании, нередко ведут к пересмотру моделей выборки и текста. Анализ результатов и их оформление также могут идти параллельно. Например, при появлении статистических материалов принимается решение об оформлении их в виде графиков или диаграмм и эти графические материалы становятся объектом анализа; оформление статистических данных в виде таблиц приводит в иной форме представления и описания данных.

Следующим этапом проведения контент-анализа является составление словаря. Словарь часто называют классификатором контент-анализа, разработанной категориальной сеткой или таблицей контент-анализа, представляющей собой совокупность систематизированных и субординированных категорий и единиц. Он строится на основе созданной системы категорий контент-анализа. Категориями являются генерализированные ключевые понятия, отражающие цель и задачи исследования. Категориальный аппарат и подчинённые ему единицы счёта, введённые в словарь в соответствии с созданной классификацией, должны идентифицировать общую тематику исследования и его частные особенности, то есть охватывать её полностью и максимально точно. Необходимо избегать крайностей при определении категорий контент-анализа. Так, при включении слишком крупных и размытых категорий, исследователь рискует получить тривиальные результаты, отражающие только общую суть вопросов. При введении слишком узких категорий есть вероятность получить большое количество малозначимой информации, которую крайне трудно будет в дальнейшем интегрировать и обобщить, для того, чтобы дать комплексную, но ём-

кую оценку исследуемой проблеме. Категории должны максимально полно охватывать исследуемую тему, быть взаимоисключающими, не позволяющими включить одни и те же единицы одновременно в несколько категорий. Они должны обладать надёжностью и трактоваться единым образом. От корректного выбора категорий во многом зависят результаты всей работы, и поэтому исследователей уже давно интересует вопрос автоматизации выбора категорий. Решение этого вопроса позволило бы существенно экономить время проведения контент-анализа и также получать более достоверные и объективные результаты. Стоит оговориться, что автоматизированное создание системы категорий возможно только при работе с большими массивами.

При работе с текстом, в категориальную сетку заносится выбранный массив слов, словосочетаний, лексем (форм слов, имеющих сходное значение) в соответствии с поставленными задачами, определяются ключевые единицы, имеющие чётко идентифицируемую семантику, максимально точно соответствующие выделенным категориям контент-анализа, и их синонимический ряд. Важно учесть всю совокупность вариантов единиц, отражающих широту категории, не игнорируя кажущиеся малозначительными, но так же близкие по значению единицы. Часто в словарь вводятся единицы полярные по значению и оценивающим свойствам (например, антонимы), характеризующие своеобразные символично-знаковые поля, в рамках которых и существуют эти оценочные единицы. Такое позитивное и негативное маркирование важно на последующих этапах контент-анализа при расчёте и изучении корреляций единиц. Серьёзным препятствием при определении позиции и дальнейшей оценки корреляции единицы того или иного оценочного символично-знакового поля является различное отношение представителей целевых групп к одной и той же единице в конкретной изучаемой коммуникативной ситуации. Эта особенность исследуемой аудитории ставит дополнительные трудоёмкие задачи по исключению из категориальной матрицы единиц, трактуемых целевыми группами различно, способных привести к получению некорректных результатов исследования. При невозможности исключения таких единиц из словаря определяющую роль при работе с ними играет исследование контекста конкретных элементов выборочной совокупности текстов, что позволяет адекватно трактовать значение использованной единицы.

На основе построенной таблицы создаётся так называемая кодировальная матрица, служащая инструментом квантификации заданных единиц в исследуемом массиве.

Единицами контент-анализа являются наиболее часто встречающиеся в тексте слова, словосочетания, предложения, абзацы, строки, колонки, физическая протяжённость и площадь текста, его доля в общем изучаемом массиве. Квантификации могут быть подвергнуты также и нетекстовые объекты, такие как аудио или видео плёнка, длительность трансляции по радио или телевидению.

Условно можно разделить словари на два вида: частотные и семантические. Первые подразумевают выделение единиц контент-анализа на основе час-

тоты их использования по отношению к суммарному количеству слов в массиве и к другим единицам потенциально подходящим для включения в словарь. Второй вид представляет собой включение в словарь категорий и единиц на основе заранее проработанных текстов, максимально точно описывающих предмет исследования, а потому уже содержащих большинство единиц будущего словаря. Чаще всего в словарь включаются имена существительные, отглагольные имена, реже глаголы, прилагательные и наречия, совсем редко частицы и союзы. При этом важность выбора видов частей речи включённых в словарь варьируется в зависимости от исследуемого массива, поставленных задач, интуиции исследователя.

Стоит отдельно отметить важную особенность проведения контент-анализа художественных текстов. В них основной единицей счёта выступают не лингвистические единицы (предложения, словосочетания, слова), а смысловые единицы. Они могут содержаться, например, в одном словосочетании, в предложении, а могут находиться в одном абзаце, что существенно усложняет процесс поиска данных. Смысловые единицы помимо своей неструктурированности, могут быть имплицитны и неидентифицируемы в рамках поиска синтаксических единиц, а потому упущены. Существует также мнение о том, что смысловые нагрузки художественного произведения не могут быть соотнесены с нетекстовой действительностью, то есть быть подвержены кодированию и квантификации, а в дальнейшем – качественной обработке – интерпретации, в силу чего производить контент-анализ художественных произведений малоэффективно.

Квантификация и интерпретация результатов проведения контент-анализа. По завершению подготовительного периода следует этап работы с единицами подсчёта, выбранными в соответствии с установленной системой категорий, опирающейся в свою очередь на цель, задачи и гипотезы исследования. Здесь исследователь прибегает к помощи таких инструментов, как регистрационная карточка или кодировальная матрица, бланк контент-анализа также называемый протоколом итогов контент-анализа. На основе систематизированного и дифференцированного материала исследователь пишет работу – записку по результатам контент-анализа, опираясь, главным образом, на протокол итогов, полученный в ходе полной дистрибуции категорий текстового массива. Определение тенденций и особенностей функционирования социальной реальности и является итогом проведения квантификации материала и контент-анализа в целом.

Сама процедура подсчёта (квантификации) близка стандартным действиям классифицирования по взаимоисключающим темам. Оперирование данными производится с помощью таблиц, математических формул, шкалирования, выстраивания данных в определённом заранее заданном порядке, специализированных компьютерных программ и т.д. Интерпретация, полученного числового материала и его дальнейшая тематическая градация, построение искомым моделям социальной действительности производится в соответствии с из-

начально установленным категориальным аппаратом, задачами, целями и гипотезами исследования.

Важно отметить, что иногда контент-анализ используется для определения лёгкости чтения конкретного набора текстов. Единицей счёта здесь является слово, имеющее любое значение. Основу квантификации в этом случае составляет длина слов, количество слов в предложении. На основе таких данных высчитывается общий индекс читабельности текста. Безусловно, такие характеристики, как жанр текста, язык, форма шрифта и другие визуальные и лингвистические особенности текста в немалой мере влияют на степень лёгкости восприятия текста. Контент-анализ и здесь не претендует на получение абсолютно точной информации. Такой вид контент-анализа позволяет также узнать приблизительный уровень образования, требуемый для понимания анализируемого текста. С этой целью используется формула Фреча. Применяют её для изучения англоязычных текстов.

Процесс квантификации чаще всего производится при использовании так называемых простых частот, подразумевающих поиск единиц счёта в одном текстовом массиве. Этот подход неприменим в случае сравнения текстовых массивов. Для сравнения используются относительные частоты, отражающие количество упоминаний единицы счёта на заданный фиксированный по объёму массив (например, на 1000 слов или 1000 страниц текста).

При анализе небольших массивов текста чаще всего единицей счёта является слово, входящее в созданную систему категориального аппарата анализа. Но при исследовании больших массивов иногда допускают некоторую редукцию значимости количества слова в рамках заданной по объёму части текста. Так, абзац, в котором искомое слово упоминается 1 раз, будет приравнен к абзацу, в котором оно использовано многократно.

Регламентирует работу исследователя специализированная инструкция кодировщика, призванная определять то, каким образом будет собираться и регистрироваться (кодироваться) информация. Другими словами, эти система норм и правил, в частности устанавливающая определённые рамки, за которые исследователю нельзя выходить при работе над массивом. В ней приводятся конкретные примеры кодирования, алгоритмы работы со спорными случаями, характеристика категорий и единиц анализа. Серьёзные трудности могут возникнуть при квантификации и обработке данных массива, состоящего из художественных произведений, в том случае, если работа над категориальным аппаратом была проведена недостаточно скрупулезно. Интерпретация смысла при работе с такими массивами заключается в идентификации смысловых единиц. Поиск их, как уже было сказано, затруднён, а семантическая нагрузка может также сильно варьироваться и качественно и количественно, что в ещё большей мере усложняет кодирование и может привести к получению необъективных результатов. При этом в большинстве случаев контент-анализ доверяет регистрации лингвистических единиц, оперируя предположением о соответствии в большинстве случаев смысла отрывка текста семантике включённых в него единиц счёта. Такое формальное отношение к художественному тексту допус-

тимо в меньшей степени, чем к другим его видам. Для решения этого вопроса в анализ вводят дополнительную единицу – тему. Это позволяет редуцировать вероятность несоответствия слова искомому в рамках конкретного этапа исследования значению. Другим вариантом преодоления такой трудности является использование «мнения арбитров» – то есть кодировщиков, классифицирующих контекст, в котором была использована единица счёта. Безусловно, и сам исследователь может им являться.

В ходе квантификации также производится анализ взаимодействия (корреляций) единиц счёта и далее интерпретация этих корреляций. Средством их определения может служить включение в контент-анализ других видов статистического анализа. Такое сочетание методов позволяет производить глубокий, разнонаправленный и точный контент-анализ. На примере функционирования этой программы легко убедиться в том, что контент-анализ это качественно-количественный метод, способный идентифицировать не только эксплицитные характеристики текста, определить его тематику, оценочную составляющую, но и проследить имплицитные сюжетные линии, которыми изобилует массив, скрытые для читателя или исследователя не вооружённого таким инструментарием.

Процедура проведения контент-анализа в пакете Lekta¹

Пакет Lekta – лексико-семантический текстовый анализатор – был создан с целью проведения контент-анализа больших текстовых массивов. Программа не только позволяет решать все базовые задачи метода, но и способна производить факторный анализ лексем, выделенных при первичной обработке массива. Это делает возможным идентификацию лексически и тематически коррелирующих текстовых фрагментов, основных и частных, а также наиболее ярких сюжетных линий, лежащих в основе изучаемого текста, что даёт возможность сделать контент-анализ более глубоким и многомерным.

Основные характеристики программы:

- Позволяет работать с большими массивами информации (более 500 тысяч слов);
- Учитывает как отдельные слова, так и словосочетания;
- Рассчитывает частоту встречаемости выделенных семантических единиц в тексте;
- Позволяет дробить текст на однородные смысловые фрагменты;
- Итоговый результат представляет собой матрицу: семантические единицы на фрагменты текста.

¹ При написании инструкции были использованы материалы с сайта разработчика программы (<http://www.nisoc.ru/lekta.html>), а также блога о контент-анализе (<http://content-analysis.ru/>).

Подготовительный этап

Шаг 1. Набор текстового массива производится из текстов одного жанра и стиля, зачастую важно также не выходить за определённые установленные временные рамки создания анализируемых сочинений. Сами тексты помещаются в документы формата “обычный текст” (txt). В принципе возможно размещение всех текстов в одном документе, но это целесообразно делать в случае анализа сравнительно небольших массивов. В дальнейшем весь массив текстов средствами пакета будет разбит на фрагменты в соответствии с установленными критериями. Технические характеристики программного продукта требуют, чтобы текст был особым образом подготовлен. Для этого следует убрать из текста все рисунки. Целесообразно убрать и таблицы.

Шаг 2. Создание единого реестра текстов. Если количество текстов невелико рекомендуется создать реестр материалов, вошедших в массив, лучше всего в формате excel (xls). В такой реестр имеет смысл включить по возможности как можно больше данных о материалах. Например, если исследователь работает с публицистическими статьями, размещёнными в сети Интернет, важно зафиксировать в реестре оригинальное название статьи, закодированное название статьи, дату её написания, URL адрес, тему в соответствии с выбранной градацией (если массив текстов включает несколько очевидных тематических групп), фамилию и имя автора, название издания, и т.д. В дальнейшем – при осуществлении анализа всегда можно обратиться к такому реестру.

Шаг 3. Кодирование исходных данных текстов. Удобным и важным инструментом при работе в программе Лекта является кодирование названий текстов. В таком коде можно коротко отобразить те особенности материала, на основе которых можно идентифицировать конкретную статью во всём массиве. Такие данные в качестве примера были приведены выше при описании единого реестра материалов контент-анализа. Так, например, если исследователю требуется закодировать статью, опубликованную 7 августа 2008 года, вышедшую в Российском СМИ (допустим, что в массиве также используются материалы зарубежных русскоязычных изданий) и описывающую международный военный конфликт на Северном Кавказе, он может закодировать документ следующим образом: 7Cau070808, где: 7 – международный телефонный код России (статья была опубликована в российском издании); Cau – сокращение от Caucasus (Кавказ) эта составляющая кода описывает общую тему исследуемого вопроса; 070808 – дата публикации статьи. Если в этот день в Российских СМИ, включённых в выборку, было опубликовано несколько статей, касающихся изучаемой проблемы, то и их можно упорядочить, пронумеровав. Например: 7Cau070808_01, 7Cau070808_02, 7Cau070808_03 и т.д.

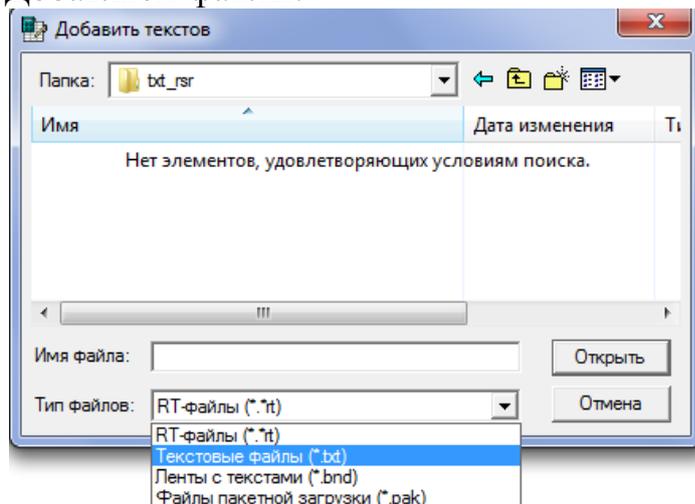
Этап работы с программой

Шаг 1. Загрузка информации

Откройте программу и выберите модуль «Фрагменты». Нажмите кнопку  («Добавить текстов»). В поле «Тип файлов» необходимо выбрать опцию «Текстовые файлы (*.txt)».

Выберите тексты, которые собираетесь анализировать при помощи клавиши «Shift» или «Ctrl» и клавиш прокрутки.

Добавляем файлы:



Файлы загрузились:

##	Фрагмент	Слов	%
1	<input checked="" type="checkbox"/> Пустой	2	--
2	<input checked="" type="checkbox"/> 03BRN60w.txt	374	--
3	<input checked="" type="checkbox"/> 03BRN70w.txt	390	--
4	<input checked="" type="checkbox"/> 03BRN72w.txt	507	--
5	<input checked="" type="checkbox"/> 03BRN73M.txt	589	--
6	<input checked="" type="checkbox"/> 03BRN74M.txt	826	--
7	<input checked="" type="checkbox"/> 3Agr19w.txt	348	--
8	<input checked="" type="checkbox"/> 3Agr22M.txt	387	--
9	<input checked="" type="checkbox"/> 3Agr33M.txt	493	--
10	<input checked="" type="checkbox"/> 3Agr43w.txt	423	--
11	<input checked="" type="checkbox"/> 3Agr45w.txt	478	--

В модуле «Фрагменты» в верхней строке списка текстов по умолчанию стоит пустой фрагмент. После загрузки текстов для анализа, его нужно удалить, выделив левой кнопкой мыши и нажав иконку «Удалить».

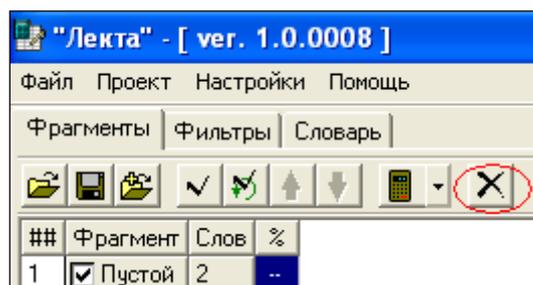
Шаг 2. Фрагментирование

При изучении газетных статей, интервью, научных текстов объектом анализа может быть как каждая отдельная статья, так и ее части (фрагменты). Фрагмент – законченная по смыслу, содержательная часть. Обычно на фрагменты разбиваются большие по объему тексты. Этот прием используется не только при работе с газетными статьями, но и с книгами, монографиями, интервью. Разбить на фрагменты тексты следует в редакторе MS Word до того, как загружать в программу «Лекта».

Основное правило: фрагменты должны быть приблизительно равными по объему (количеству слов). Это означает, что в анализ не должны одновременно включаться фрагменты объемом в две строки и фрагменты объемом в две страницы.

Фрагменты внутри единого текста разделяются между собой специальными символами.

Мы будем использовать в качестве такого символа последовательность знаков: ##SL.



NB: Непосредственно перед и после каждого специального символа должен стоять знак абзаца. Кроме этого ни до, ни после этого знака не должно быть пробелов.

Перед первым и после последнего фрагмента специальные символы не ставятся.

В программе «Лекта» процедура разбиения текста на фрагменты реализуется следующим образом.

Нажмите кнопку  («Обработка») и выберите опцию «Разбить на фрагменты». При работе с текстами, в которых расставлены спецсимволы, необходимо поставить знак «✓» в поле «разбить по строке». После этого нажмите кнопку «ОК».

В поле «Фрагменты» мы видим список текстов. В столбце ## показан порядковый номер фрагмента, в столбце «Фрагмент» - название исходного файла с текстом, в столбце «Слов» - количество слов во фрагменте.

Если текст не разбит на фрагменты при помощи спецсимволов, то можно произвести разбиение автоматически, задав приблизительный объем фрагмента в словах в поле «Настройка фрагментатора». Для этого необходимо поставить знак «✓» в поле «разбить по размеру», а затем задать размер фрагмента от минимального до максимального количества слов. Недостатком автоматического разбиения является то, что не учитывается внутренняя содержательная структура текста. Поэтому автоматическое разбиение применяется обычно только при работе с очень большими текстовыми объемами и/или в условиях ограниченного времени на их анализ.

При автоматическом разбиении текстов необходимым этапом является исправление неудачных фрагментов. Неудачными называются фрагменты, которые не попали в заданный интервал. Например, мы задали интервал 50 ± 20 слов, неудачным будет считаться фрагмент в 80 слов.

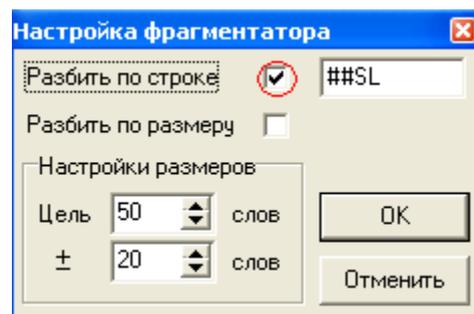
Фрагменты, которые не попали в заданный интервал после операции разбиения отмечены символом «✓». Необходимо просмотреть их. С данными фрагментами можно поступить следующим образом:

- оставить их в том виде, в каком они есть;
- слишком маленькие фрагменты можно присоединить к последующим или предыдущим фрагментам данного текста;
- слишком большие фрагменты можно разбить на части.

Разбить большой фрагмент можно при помощи кнопки .

При помощи кнопок  и  можно присоединить небольшой фрагмент к предыдущему или последующему.

После завершения работы с фрагментами необходимо нажать кнопку  («Отметить все»), чтобы в дальнейшем анализе участвовали все фрагменты.



ИТ	Слов	%
ED08.txt	58	--
ED08.txt	44	--

Кроме того, имеется также кнопка  («обратить отмеченное»), которая позволяет сделать отмеченные фрагменты неактивными, а те, которые не были отмечены, - активными.

Эта операция может также применяться для включения в анализ отдельных интересных нас фрагментов или исключения не нужных фрагментов из анализа.

Шаг 3. Создание словаря

Совокупность всех лексем, использованных для создания модели исходного текста, представляет собой словарь (категорийную сетку). Словарь обладает внутренней структурой. Отдельные лексеммы в нем необходимо объединить в семантические цепочки (категории). Семантическая цепочка – совокупность синонимичных лексем или лексем, относящихся к единой проблематике. Например, в семантическую цепочку «налоги» могут войти лексеммы: налоги, налоговый, фискальный. В ходе математических расчетов семантические цепочки обозначаются термином «фильтры».

3.1. Создание нового словаря

Выберите модуль «Словарь».

Нажмите кнопку  («Создать словарь»). Слева в верхнем и нижнем окнах появятся папки «Базовый словарь» и «Корзинка».

Знак  обозначает выбранные (активные на настоящий момент) папки в левом верхнем и левом нижнем окнах.

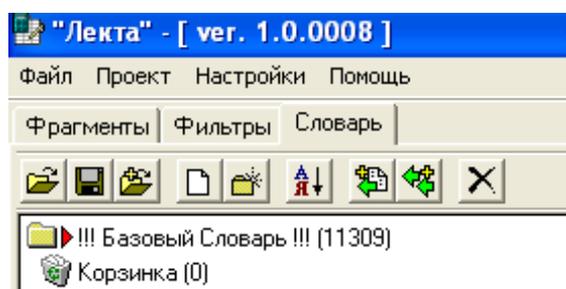
В окнах справа соответственно отображается содержимое выделенных папок (то есть слова, которые в них входят).

3.2. Операции сортировки слов

Лексеммы в каждой папке можно сортировать по алфавиту, частоте встречаемости и размеру. Для сортировки слов в папке по алфавиту необходимо нажать на название столбца

Лексема

, соответственно, сортировка по частоте происходит при нажатии на название столбца "Частота", аналогично и по длине.



Список всех созданных папок (кроме «Базового словаря» и «Корзины») можно сортировать по алфавиту при помощи кнопки . Кнопка  удаляет выделенную папку.

3.3. Поиск слов

В базовом словаре можно осуществлять поиск слов при помощи клавиатуры. В верхнем окне нужно сделать активным «Базовый словарь». Затем на клавиатуре набирается искомое слово. Это слово отображается рядом с названием столбца «Лексема», одновременно оно находится и в базовом словаре. Символ «*» обозначает окончание слова.

3.4. Создание новых папок

Создание семантических цепочек реализуется через создание новых папок  «Создать новую папку». Для каждой семантической цепочки формиру-

ется отдельная папка:  Новая папка (0). Переименовать папку можно при помощи нажатия левой кнопки мыши по выделенному названию папки. Число в скобках указывает количество лексем в папке.

Кроме того, можно создавать папку, непосредственно встав на необходимое нам слово и нажав кнопку  «Создать именованные папки из выделенного». В таком случае папка получает название выбранного нами слова, а само это слово перемещается из базового словаря в созданную нами папку.

3.5. Перенос информации между папками

Перемещение лексем между папками осуществляется при помощи кнопки  «Перенести выделенное». Здесь можно также объединять однокоренные лексемы в единую словоформу (остается единая корневая часть однокоренных слов, а окончания заменяются на знак *).

Например, мы хотим создать папку «Зависимость». Для этого мы выделяем слова «зависимости» и «зависимость» в папке «Базовый словарь», между окнами у нас появляется дополнительная кнопка с корневой формой «зависимост*», которая позволяет нам автоматически объединить лексемы и перенести корневую форму в новую папку «Зависимость».

Мы можем также объединить их со словами «зависеть» и «зависит». Тогда форма, которую мы сможем перенести в новую папку, будет выглядеть, как «завис*». В данном случае мы должны решить, необходимо ли и оправдано подобное сокращение. Если слова сокращаются чрезмерно, то лучше переносить их без сокращения.

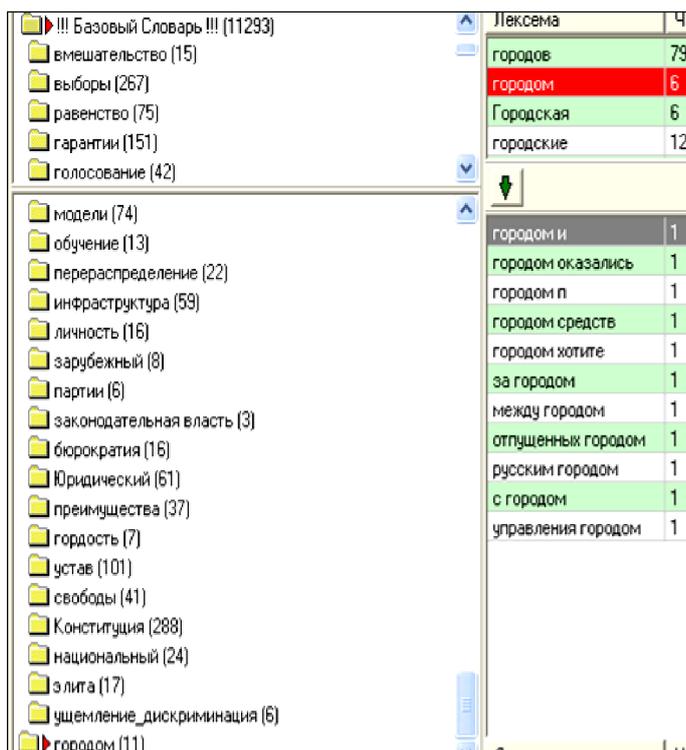
3.6 Работа со словосочетаниями

Лекта позволяет работать со словосочетаниями. Для этого необходимо использовать правую кнопку мыши. В окне, отображающем слова базового словаря, нужно вы-

Лексема (НЕИЗБЕЖНО*)	Частота	Длина
незначительность	1	16
неизбежно	1	9
неизбежного	1	11
неизбежное	1	10
неизбежность	1	12
Неизвестно	2	10

Лексема	Частота	Длина
влияние	14	7
влиять	6	6
воздействия	5	11
заставляет	5	10
зависеть	6	8
зависимости	19	11
зависимость	10	11
зависит	10	7
повлиять	5	8
принуждения	4	11
рычаги	4	6

 **зависимост***



Лексема	Частота
городов	79
городом	6
Городская	6
городские	12
городом и	1
городом оказались	1
городом п	1
городом средств	1
городом хотите	1
за городом	1
между городом	1
отпущенных городом	1
русским городом	1
с городом	1
управления городом	1

делить слово, для которого хочется найти словосочетания. Затем нажать правую кнопку мыши, выбрать опцию «Найти все словосочетания». В конце списка папок появится новая папка, которая и будет содержать все возможные словосочетания для данной лексемы.

Например, насколько часто встречается в тексте словосочетание «управление городом». Мы выделили в базовом словаре слово «городом» и нашли словосочетания. Слово «городом» встречается 6 раз, в результате мы получили 11 словосочетаний. С формальной точки зрения, программа выделяет словосочетания двумя способами:

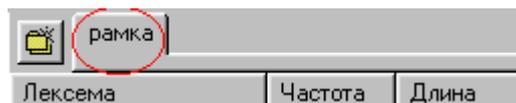
- слово и все слова, которые встречаются непосредственно перед ним;
- слово и все слова, которые стоят непосредственно после него.

Однако невозможно создать запрос на словосочетание, включающее в себя три слова.

Итак, найдя словосочетания для лексемы «городом», мы определили, что словосочетание «управление городом» встречается в нашем корпусе текстов лишь один раз.

3.7. Организация закладок

В процессе работы со словарем в папке «Базовый словарь» правой кнопки мыши ставить закладки для нужных в последующем слов. Эти закладки помогут быстро вернуться к словам в случае необходимости. Закладки отображаются над верхним окном.



3.8. Сортировка папок

Лексемы в каждой папке можно сортировать по алфавиту, частоте встречаемости и размеру. Для сортировки слов в папке по алфавиту необходимо нажать на название столбца **Лексема**, соответственно, сортировка по частоте происходит при нажатии на название столбца «Частота», аналогично и по длине.

Список всех созданных папок (кроме «Базового словаря» и «Корзины») можно сортировать по алфавиту при помощи кнопки . Кнопка  удаляет выделенную папку.

После завершения формирования словаря все папки, содержащие семантические цепочки, преобразуются в фильтры при помощи кнопки  («Добавить все папки как новые фильтры»). Папки в качестве фильтров могут добавляться и по отдельности при помощи кнопки  («Добавить папку как новый фильтр»).

3.9. Критерии качества построения словаря

В столбце  проверьте качество словаря по показателю информативности. Оптимизируйте словарь путем добавления в фильтры важных лексем, которые не были включены в словарь, однако могут повысить объясняющую способность модели и обладают ценностью для анализа.

Основными критериями качества создаваемой модели текста выступают:

- информативность объекта анализа;

- частота используемости фильтров;
- информативность модели.

Информативность объекта анализа представляет собой показатель отношения количества слов, вошедших в состав словаря и содержащихся в данном объекте анализа, к количеству слов, из которых образован анализируемый фрагмент текста. Чем выше информативность, тем более качественно описан данный объект анализа. Информативность объекта анализа, равная 100%, означает, что все слова, из которых состоит анализируемый фрагмент текста, включены в словарь. Это бывает очень редко: любой язык имеет вспомогательные части речи, не имеющие значения для моделирования содержательного плана текста. Хорошим уровнем моделирования является ситуация, когда основная масса объектов анализа (фрагментов) имеет информативность приблизительно равную 30% и при этом доля объектов с низкой информативностью не превышает 10% от общего количества объектов.

Частота используемости фильтра рассчитывается методом сравнения входного потока текста на соответствие заданному логическому выражению. Эмпирически установлено, что наиболее адекватной для моделирования частотой встречаемости фильтра (фильтр - логическое выражение, построенное с помощью логических операций И/ИЛИ из лексем, содержащихся в потоке анализируемого текста) является частота в границах от 5% до 50%. Повышение частоты встречаемости фильтра достигается за счет увеличения логического выражения путем присоединения с использованием операции ИЛИ дополнительных синонимов, содержащихся в исходном анализируемом тексте. Снижение частоты встречаемости, соответственно путем разбиения логического выражения на части.

Информативность модели представляет собой отношение размера словаря к общему числу слов исходного текста. Как показывает практическое использование данной методики, модель текста может адекватно отражать его содержание, если словарь содержит около трети использованных в тексте лексем.

Все три показателя тесно взаимосвязаны между собой. Повышение информативности объекта осуществляется путем просмотра объектов с низкой информативностью (менее 10%) и отбора из них слов, которые еще не включены в созданные фильтры. Найденные слова могут включаться в имеющиеся фильтры и образовывать новые. В случае включения в фильтр нового слова, повышается частота его встречаемости, а значит и качество модели. Соответственно и повышение частоты встречаемости фильтра за счет расширения логического выражения приводит к увеличению информативности объектов, в которых встречаются слова, вновь включенные в анализ.

После оптимизации словаря запустите процедуру применения фильтров еще раз.

Шаг 4. Создание фильтров

4.1. Автоматическое создание фильтров – перенос из словаря

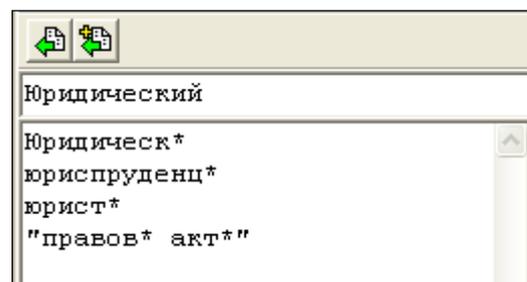
Перейдите в модуль «Фильтры». Удалите фильтр «Пустой» при помощи кнопки .

Модуль «Фильтры» позволяет редактировать как названия категорий словаря (фильтров), так и лексемы внутри словаря.

В левом окне модуля «Фильтры» отображаются фильтры. В столбце ## - порядковый номер фильтра, в столбце «Фильтр» - его название, в столбце «Лекс.» - количество лексем, входящих в данную семантическую цепочку. После того, как будут произведены расчеты, в столбце "%" появится доля, которую данный фильтр составляет в общем массиве слов.

##	Фильтр	Лекс.	%			
116	<input checked="" type="checkbox"/> ушерб	13	--	ушерб*	ущемили	Юридический
117	<input checked="" type="checkbox"/> федерализм	1	--	федерализм*		Юридическ*
118	<input checked="" type="checkbox"/> федеральный	1	--	федеральн*		юриспруденц*
119	<input checked="" type="checkbox"/> финансы_кредит	5	--	финансировани*	финансист*	юрист*
120	<input checked="" type="checkbox"/> фракции	1	--	фракци*		"правов* акт"
121	<input checked="" type="checkbox"/> цели_планы	9	--	цели	цель	
122	<input checked="" type="checkbox"/> централизация	8	--	централиз*	вертикаль	
123	<input checked="" type="checkbox"/> чиновники	1	--	чиновник*		
124	<input checked="" type="checkbox"/> школа	5	--	школа	школ	
125	<input checked="" type="checkbox"/> Экономика	2	--	Экономик*	экономическ*	
126	<input checked="" type="checkbox"/> эффективность	1	--	эфективн*		
127	<input checked="" type="checkbox"/> Юридический	4	--	Юридическ*	юриспруденц*	
128	<input checked="" type="checkbox"/> ЯБЛОКО	3	--	ЯБЛОКО	ЯБЛОКА	

Правое верхнее окно модуля «Фильтры» имеет две части: верхняя отображает название фильтра, который выделен в левом окне, а нижняя – список лексем, составляющих данный фильтр. Здесь можно отредактировать как название, так и лексемы. Вне-



сенные изменения сохраняются при помощи кнопки  («Сохранить взамен текущего фильтра»).

NB: При необходимости замените окончания однокоренных лексем на знак *, а также удалите из фильтров слова, дублирующие корневую словоформу.

4.2. Ввод новых фильтров и редактирование

Иногда необходимо добавить новую лексическую единицу к уже имеющимся фильтрам. Для этого нужно в верхней части правого окна вписать название нового фильтра, в нижней его части – лексемы. После этого новый фильтр можно добавить к уже имеющимся при помощи кнопки  («Добавить как новый фильтр»).

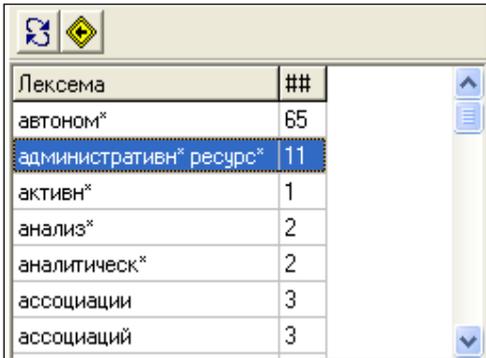
4.3. Сортировка фильтров

При помощи кнопок   можно перемещать выделенный фильтр вверх или вниз в списке фильтров. Нажатием правой кнопки мыши можно выделить команду «Сортировать по алфавиту» или «Сортировать по проценту попаданий». Сортировка по проценту попаданий позволяет увидеть слабо работающие фильтры, которые не будут являться значимыми для дальнейшего анализа. Если с содержательной точки зрения эти фильтры не представляются принципиально необходимыми, их можно исключить из последующего анализа.

4.4. Поиск использованных слов

Нижнее правое окно поля «Фильтры» позволяет осуществлять поиск слов внутри фильтров. Для активизации этого поля необходимо нажать кнопку  «Обновить список слов». После этого в алфавитном порядке отобразится весь список слов, присутствующих в фильтрах. В этом списке можно быстро найти нужное слово. Для этого необходимо встать курсором мыши в нижнее правое окно модуля «Фильтры», затем набрать искомое слово на клавиатуре. Процедура поиска аналогична поиску в базовом словаре.

Кнопка  «Перейти к фильтру» позволяет быстро перейти к фильтру, где находится нужное нам слово.



Лексема	##
автоном*	65
административн* ресурс*	11
активн*	1
анализ*	2
аналитическ*	2
ассоциации	3
ассоциаций	3

К нужному слову можно перейти и с помощью двойного нажатия левой кнопки мыши на нужном слове в данном окне.

4.5. Включение/ исключение фильтров

Кнопка  «Очистить фильтры от дубликатов» позволяет автоматически удалить дубликаты слов, которые были дважды или более раз были использованы в различных папках построенного словаря. Эта опция в первую очередь убирает слова, повторяющие корневые формы тех же слов со знаком *.

Кнопка  позволяет удалить выделенный фильтр.

При помощи кнопки  можно выделить все фильтры для того, чтобы они участвовали в последующем анализе. Активные фильтры отмечены символом «✓». Для того чтобы фильтр не участвовал в последующем анализе можно сделать его неактивным, нажав левой кнопкой мыши по символу «✓». Так же, как и в поле «Фрагменты» здесь присутствует кнопка  «Обратить выделенное».

В модуле «Фильтры» также доступно специальное меню через правую кнопку мыши. Это меню позволяет сортировать фильтры по алфавиту, удалять отмеченные фильтры, копировать весь список фильтров. Скопированные фильтры из буфера обмена можно вставить в программу MS Excel. Можно также редактировать список фильтров в программе MS Excel, а затем вставлять его в модуль «Фильтры» при помощи правой кнопки мыши. Важно помнить, что первый столбец этого списка представляет собой названия фильтров.

Шаг 5. Расчет файла частот

5.1. Применение фильтров

Перейдите в поле «Фрагменты». При помощи кнопки «Обработка»  выберите опцию «Применить фильтры». Сохраните результаты в вашем персональном каталоге.

В столбце  проверьте качество словаря по показателю информативности. Процент в этом столбце показывает, какая доля слов из фрагмента описывается построенным словарем. Необходимо просмотреть фрагменты с нулевым

или невысоким (менее 10-15%) уровнем объяснения. Внутри этих фрагментов могут содержаться важные лексемы, которые по тем или иным причинам были упущены при составлении словаря. Оптимизируйте словарь путем добавления в фильтры (через модуль «Фильтры») важных лексем, которые не были включены в словарь, однако могут повысить объясняющую способность модели и обладают ценностью для анализа.

После оптимизации словаря запустите процедуру применения фильтров еще раз.

5.2. Сохранение результатов

Готовую матрицу результатов необходимо импортировать в программу MS Excel. Загрузите Excel, выберите опцию «Открыть файл», тип файлов «Все файлы». Найдите в вашем каталоге файл результатов и нажмите кнопку «Открыть». Появится диалог импорта. Нажмите кнопки «Далее», «Готово». Файл результатов готов для дальнейшей обработки.

Далее результаты, как правило, обрабатываются с помощью факторного анализа.

Раздел 5. Факторный анализ

Факторный анализ – многомерный статистический метод, применяемый для изучения взаимосвязей между значениями переменных. Главной задачей факторного анализа является переход от *большого* числа признаков изучаемого явления к *меньшему* числу стоящих за ними скрытых (латентных), внешне ненаблюдаемых факторов или комбинаций признаков. Также считается, что выделенные латентные характеристики одновременно являются и наиболее существенными, определяющими изучаемое явление.

Таким образом, можно выделить две цели факторного анализа:

- определение взаимосвязей между переменными;
- сокращение числа переменных

В том виде, в котором используется в настоящее время, факторный анализ оформился в работах американского психолога Тёрстоуна. Хотя обычно рождение факторного анализа как метода исследования связывают с публикацией в 1904 г. статьи Спирмена «Объективное определение и измерение общего интеллекта», а предысторию – с именами английского антрополога Гальтона и английского математика, основателя математической статистики Пирсона¹.



Рис. 5.1. Основные этапы истории факторного анализа

¹ С краткими обзорами истории факторного анализа можно ознакомиться в книгах: Митина О.В., Михайловская И.Б. Факторный анализ для психологов. – М.: Учебно-методический коллектор «Психология», 2001. – С. 3-9; Харман Г. Современный факторный анализ. – М.: Статистика, 1972. – С. 13-16; Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник. – 5-е изд. – М.: Финансы и статистика, 2004. – С. 323; Попова О.В. Политический анализ и прогнозирование: Учебник. – М.: Аспект Пресс, 2011. – С. 364-365.

Теоретические положения метода факторный анализ

Практическое выполнение факторного анализа начинается с проверки его условий. В обязательные условия факторного анализа входят:

- все признаки должны быть количественными;
- число наблюдений должно быть в два раза больше числа переменных;
- выборка должна быть однородна;
- исходные переменные должны быть распределены симметрично;
- факторный анализ осуществляется по коррелирующим переменным.

Исходной задачей факторного анализа является всестороннее и компактное описание объекта исследования.

Процедура факторного анализа состоит из следующих этапов:

1. Подготовка исходной матрицы данных.
2. Вычисление корреляционной матрицы для всех переменных, участвующих в анализе.
3. Факторизация или извлечение факторов.
4. Вращение факторов для создания упрощенной структуры.
5. Подсчет факторных значений по каждому фактору для каждого наблюдения.
6. Интерпретация факторов.

Подготовка исходной матрицы данных

Практически во всех процедурах любой программы факторного анализа в качестве исходных данных используются матрицы данных. Матрица – это прямоугольная таблица чисел, в которой, как правило, горизонтальные линии (строки, ряды) соответствуют наблюдениям (объектам), а вертикальные линии (столбцы) – переменные. Например, при анализе газетных публикаций строка может соответствовать одному газетному сообщению (или его фрагменту), а столбцы – кодам или лексемам, которые были отобраны исследователем.

Вычисление корреляционной матрицы

Факторный анализ основан на взаимодействии переменных. Собственно процедура факторного анализа начинается с вычисления матрицы взаимосвязей переменных между собой (это квадратная матрица, размер которой равен количеству переменных).

Наиболее распространенная мера взаимосвязи – это корреляционная связь. Корреляция (корреляционная зависимость) – статистическая взаимосвязь двух или нескольких случайных величин. При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

Корреляционную матрицу можно вычислять различными способами. Чаще всего вычисляется коэффициент корреляции Пирсона, но может вычисляться и коэффициент ранговой корреляции (тогда говорят о непараметрическом факторном анализе). Коэффициент корреляции удобен тем, что это стандартизированная мера взаимосвязи, не зависящая ни от единиц измерения, ни от диапазона изменения переменных.

Факторизация

Факторами называют гипотетическую латентную переменную, которая одновременно объединяет несколько формально измеренных признаков объекта. Обобщение позволяет выделить связи между исходными переменными, которые ранее не были очевидны, а затем перейти на более высокий уровень понимания явления. При анализе в один фактор объединяются сильно коррелирующие между собой переменные, как следствие происходит перераспределение дисперсии между компонентами и получается максимально простая и наглядная структура факторов. После объединения коррелированность компонент внутри каждого фактора между собой будет выше, чем их коррелированность с компонентами из других факторов.

Извлечение фактора начинается с подсчёта суммарного значения всех участвующих в анализе переменных. И первой задачей является выбор взаимодействующих переменных, чья взаимная корреляция обуславливает наибольшую долю в общей дисперсии. Эти переменные образуют первый фактор. Затем первый фактор исключается, а из оставшегося множества переменных снова выбираются те, чье взаимодействие определяет наибольшую долю оставшейся общей дисперсии. Эти переменные образуют второй фактор. Процедура извлечения факторов продолжается до тех пор, пока не будет исчерпана вся общая дисперсия переменных. В хорошо организованном исследовании суммарный процент объяснённой дисперсии достигает 85-95%. На практике этот процент обычно ниже.

После того как процедура извлекает первый фактор, напротив его номера появляется его собственное значение. Собственное значение фактора пропорционально доле общей дисперсии, определяемой данным фактором. Первое собственное значение всегда является наибольшим и превышает единицу.

В факторном анализе применяется множество методов, позволяющих трансформировать корреляционную матрицу в факторную. Однако на практике чаще всего используется метод главных компонент. Суть данного метода состоит в замене коррелированных компонентов некоррелированными факторами. Другой важной характеристикой метода является возможность ограничиться наиболее информативными главными компонентами и исключить остальные из анализа, что упрощает интерпретацию результатов.

Вращение факторов

Все извлечённые факторы очень редко представляют интерес для исследователя. Поэтому необходимо принять решение, какие из факторов оставить для дальнейшего анализа. Здесь рекомендуется руководствоваться здравым смыслом и оставлять те факторы, которые имеют понятную теоретическую или логическую интерпретацию. Когда выделенным факторам не удаётся дать достаточно чёткую содержательную интерпретацию, применяется метод вращения факторной структуры. Целью вращения является получение простой структуры, которой соответствует большое значение нагрузки каждой переменной только по одному фактору и малое по всем остальным. Нагрузка отражает связь

между переменной и фактором, являясь подобием коэффициента корреляции. Значения нагрузки лежат в диапазоне от -1 до 1. Идеальная структура предполагает, что каждая переменная имеет нулевые значения нагрузок для всех факторов, кроме одного, для которого нагрузка этой переменной близка к 1 (-1). После поворота осей переменные оказываются их близи, что соответствует максимальной нагрузке каждой переменной только по одному фактору. Вращение несколько не влияет на математическую строгость анализа: взаимное положение переменных при повороте осей не меняется.

Выделяют два основных класса способа вращения:

- *ортогональное вращение*, когда при повороте осей координат угол между факторами остаётся прямым, т.е. сохраняется предположение о некоррелированности фактором. Наиболее популярным методом ортогонального вращения является варимакс (Varimax), когда уменьшение количества переменных происходит за счёт группировки вокруг фактора только тех переменных, которые с ним связаны в большей степени, чем остальные.
- *косоугольное вращение*, при котором первоначальное ограничение по поводу некоррелированности факторов не снимается. Наличие непрямого угла между осями факторов означает, что они не являются полностью независимыми друг от друга.

Перед выполнением процедуры вращения задают количество факторов, в пространстве которых производится вращение. Вращение и анализ факторных диаграмм, как правило, производится несколько раз с учётом различного количества факторов. Обычно начинают с избыточного количества факторов, уменьшая их число и доводя его до необходимого предела.

Интерпретация факторов

Путём вращения необходимо ситуации, когда значение нагрузки для рассматриваемого фактора являются большим, а для остальных факторов – малым. Факторные нагрузки – это корреляции между фактором как функционально целым и значениями конкретных переменных, входящих в этот фактор. Обычно требуется, чтобы факторные нагрузки были не ниже 0,40-0,50.

Содержательной интерпретации подвергается матрица факторных нагрузок. Те из них, которые относятся к одному фактору, располагаются в порядке убывания абсолютных значений. Рассматриваются нагрузки, имеющие максимальные абсолютные значения. Далее выясняется общее содержание этой группы параметров, которое объединяет признаки в одну группу. Эта группа и называется фактором.

Проведения факторного анализа с применением пакетов Statistica и EXCEL (пошаговая инструкция)

Ввод матрицы в пакет

Вырежьте числовую матрицу из импортированного в Excel файла результатов контент-анализа и сохраните ее отдельным файлом. Откройте файл в пакете STATISTICA. Сохраните новый файл с расширением .sta в ваш каталог.

Расчет факторного анализа в пакете Statistica

Загрузите модуль .

Нажмите кнопку . Здесь необходимо выбрать те переменные (в нашем случае, номера фильтров), которые будут подвергнуты анализу.

Нам необходимо включить в анализ все переменные. Поэтому нажмите кнопку , после чего нажмите «ОК».

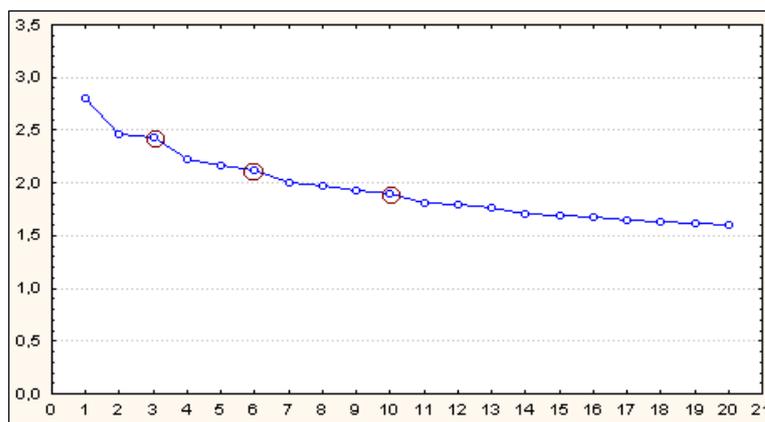
В поле «Factor Analysis» нажмите «ОК» еще раз.

Выберите предварительно ориентировочное число факторов (например, 10) в поле «Maximum no. of factors». Нажмите «ОК».

В открывшемся окне «Factor Analysis Results» нажмите кнопку

 **Scree plot**

Открывшийся график позволяет определить оптимальное число факторов. Оптимальное число факторов определяется точкой перелома графика. На рисунке видно несколько таких



точек: 3, 6, 10. Обычно сначала выбирают максимальное число факторов.

Вернемся в окно «Factor Analysis Results» и нажмем кнопку «Cancel». Мы вернулись в предыдущее окно. Задайте число факторов, равное 10, а затем нажмите «ОК».

Нажмите кнопку  **Eigenvalues**. В итоговой строчке столбца «Cumul. %» вы видите объясняющую способность модели. Если это число меньше 10%, то необходимо выбрать большее число факторов.

Нажмите кнопку  **Factor rotation:** и выберите опцию «Varimax normalized». Полученную матрицу факторных нагрузок скопируйте в Excel.

Нажмите кнопку  **Factor scores**. Полученную матрицу факторных коэффициентов скопируйте в Excel.

Продолжение расчета факторного анализа в пакете EXCEL

Обратимся к *матрице факторных нагрузок*.

Замените номера переменных на названия фильтров.

Теперь необходимо найти максимум по модулю. Встаньте в первый пустой столбец после матрицы и рассчитайте абсолютные значения для каждого столбца.

=ABS(C2)							
C	D	E	F	G	H	I	J
Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 1	Factor 2	Factor 3
-0,217909	-0,025888	0,008035	-0,134536	0,081236	0,555948	=ABS(C2)	0,025888
0,058054	0,487977	0,173903	0,001992	-0,034821			

У вас получилась матрица абсолютных значений.

В следующем столбце найдите максимум для каждой строки абсолютных значений.

G	H	I	J	K	L	M	N
Factor 6	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	
0,081236	0,555948	0,217909	0,025888	0,008035	0,134536	0,081236	=МАКС(H2:M2)

Встаньте в первый пустой столбец после матрицы и задайте формулу: абсолютное значение из верхней строчки первого столбца матрицы абсолютных значений разделить на максимум.

G	H	I	J	K	L	M	N	O	P
Factor 6	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6		Factor 1	Factor 2
0,081236	0,555948	0,217909	0,025888	0,008035	0,134536	0,081236	0,555948	=H2/N2	0,39196
-0,034821									

Поделите на максимум каждое из найденных абсолютных значений.

O	P	Q
Factor 1	Factor 2	Factor 3
1	0,39196	0,046565
1	0,292666	0,911263
1	0,044236	0,036122
1	0,182951	0,685665
1	0,105518	0,080333
1	0,931547	0,777509
0,269259	1	0,212153
0,203494	1	0,026126

Отсортируйте получившуюся матрицу по первому столбцу. После ячеек с единицами вставьте строку. Отсортируйте оставшуюся матрицу по следующему столбцу и снова вставьте строку. Повторяйте процедуру, пока вся матрица не будет отсортирована.

Вернитесь к исходной матрице нагрузок. Первый блок с названиями фильтров относится к первому фактору, второй – ко второму и так далее. От-

сортируйте каждый из этих блоков по соответствующим факторным нагрузкам. Отрицательные факторы – по возрастанию, положительные – по убыванию.

13	Занимаются аэробикой, фитнесом	0,016666	0,009297	0,570556	-0,132151
14	Соблюдают диету	0,106442	0,058054	0,487977	0,173903
15	Предпочитают вегетарианскую пищу	0,052438	0,135889	0,445750	0,127925
16	Занимаются на тренажерах (бодибилдинг)	0,081277	-0,183080	0,420863	-0,319655
17	Занимаются йогой, оздоровительными практиками	-0,050293	-0,029355	0,385924	-0,091478
18					
19	Принимают лекарственные препараты	0,133760	0,071873	-0,036283	0,689154
20	Регулярно посещают врачей в профилактических целях	0,154551	-0,067730	-0,059119	0,529545
21	Принимают витамины	-0,250184	-0,145855	0,229018	0,420584

Проанализируйте содержание каждого фактора и дайте ему название.

Перейдите к *матрице факторных коэффициентов*. Вам нужно отобрать фрагменты, относящиеся к каждому фактору.

Сначала скопируйте из файла результатов контент-анализа столбец с текстами фрагментов и вставьте после матрицы факторных коэффициентов.

Отсортируйте первый столбец следующим образом: если фактор положительный (см. факторные нагрузки), то отсортируйте по убыванию, если отрицательный – по возрастанию.

Скопируйте фрагменты, факторные коэффициенты которых больше по модулю 0,5 (помните, что к фактору относятся только те объекты, которые совпадают по знаку с нагрузками: если нагрузки отрицательные, то и факторные коэффициенты отрицательные).

Прделайте аналогичную процедуру по отношению к каждому фактору.

У *поляризованного фактора* анализируется как положительная, так и отрицательная часть.

Фрагменты, относящиеся к фактору, следует анализировать с точки зрения смысла фактора, который определяется, фильтрами в него вошедшими. Не следует пересказывать смысл текстов, которые вошли в фактор. Необходимо делать социологические заключения, соотнесенные с целями и задачами исследования, проверять поставленные гипотезы.

NB: Важно помнить, что на этом этапе анализ уже является качественным. Необходимо понять внутреннюю смысловую структуру текстов, их взаимозависимость. Свои мысли нужно подтверждать цитированием оригинальных текстов, однако не следует забывать, что цитата является иллюстрацией, а не самоцелью. Отчет по результатам исследования должен содержать социологические заключения, сделанные на основании факторной модели.

ЛИТЕРАТУРА

Основная литература

1. Боришполец К. П. Методы политических исследований: Учеб. пособие для студентов вузов. – М: Аспект Пресс, 2005. – 221 с.
2. Горшков М.К., Шереги Ф.Э. Прикладная социология: методология и методы: Учебное пособие. – М.: Альфа-М: ИНФРА-М, 2009. – 416 с. – Глава 4 «Метод анализа документов» (С. 126-136).
3. Добренъков В.И., Кравченко А.И. Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2004.
4. Ильин В.И. Драматургия качественного полевого исследования. – СПб.: Интерсоцис, 2006. – 256 с. – Раздел «Анализ текстовых материалов» (С. 206-221).
5. Мангейм Джарол Б., Рич Ричард К. Политология. Методы исследования. – М.: Изд-во «Весь Мир», 1997. – С. 269-328. [Контент-анализ. Источники и применение сводных данных]
6. Оперативное социологическое исследование: методика и опыт организации / Под ред. Д.Г. Ротмана, А.Н. Данилова, Л.Г. Новиковой. – Мн.: БГУ, 2001. – 354 с. – Раздел 4.6. «Анализ документов» (С. 147-164) и раздел 6.7. «Контент-анализ прессы в рамках ОСИ» (С. 306-320).
7. Попова О.В. Политический анализ и прогнозирование: Учебник. – М.: Аспект Пресс, 2011. – 464 с. – Раздел 3 «Работа с документами» (С. 168-228) и Раздел 4, глава 25 «Факторный анализ» (С. 364-379).
8. Рабочая книга социолога, 2-е изд. – М.: Наука, 1983. – 480 с. – С. 275-307.
9. Романов П.В., Ярская-Смирнова Е.Р. Методы прикладных социальных исследований. Учебное пособие. Изд. 2-е, дополненное. – М.: ООО «Вариант» ЦСПГИ, при участии ООО «Норт Медиа», 2008. – 215 с. – С. 76-91.
10. Семенова В.В. Качественные методы: введение в гуманистическую социологию. – М.: Добросвет, 1998. – 292 с.

Дополнительная литература

К разделу 1. Документальные источники в социологии

1. Бардаев Э.А., Кравченко В.Б. Документоведение: учебник для студентов высших учебных заведений. – М.: Академия, 2010. – 301 с.
2. Джери Д., Джери Дж. Большой толковый социологический словарь. В 2-х томах. – М.: «Вече», «АСТ», 1999. – 544с. (I т.)528 с. (II т.)
3. Документ как социокультурный феномен: Сборник материалов IV Всероссийской научно-практической конференции с международным участием / Под общ. ред. Н.С. Ларькова. – Томск: Томский государственный университет, 2010. – 624 с.
4. Документальные системы в управлении / Под ред. Г. Г. Воробьева, М. Г. Гаазе-Раппопорта. – М.: Экономика, 1973. – 207 с.

5. Куняев Н.Н., Уралов Д.Н., Фабричнон А.Г. Документоведение: учебник для вузов. – М.: Логос, 2008. – 348 с.
6. Кушнаренок Н.Н. Документоведение: учебник. – М.: Знание, 2008. – 459 с.
7. Лотман Ю.М. Внутри мыслящих миров // Семиосфера. – СПб.: «Искусство-СПб», 2000. – С.150-390.
8. Лотман Ю.М. Семиотика культуры и понятие текста // Избранные статьи. Т. 1. – Таллинн: Александра, 1992. – С. 129-132.
9. Почепцов Г.Г. Теория коммуникации. – М.: Рефл-бук, 2006. – 656 с.

К разделу 2. Основные приёмы анализа документов

1. Готлиб А.С. Введение в социологическое исследование: Качественный и количественные подходы. Методология. Исследовательские практики: Учебное пособие. – Самара: Изд-во «Самарский университет», 2002. – 424 с. – С. 353-388.
2. Давыдов А.А. Качественные исследования: перспективы развития [http://www.isras.ru/index.php?page_id=922]
3. Ключкина О.Б. Построение теории на основе качественных данных // Социологические исследования. – 2000. – №10. – С.92-101.
4. Ковалев Е.М., Штейнберг И.Е. Качественные методы в полевых социологических исследованиях. – М.: Логос, 1999. – С.317-344.
5. Козина И.М. Особенности применения стратегии «исследование случая» при изучении производственных отношений // Социология: методология, методы, математические модели. – 1995. – №5-6. – С.5-15.
6. Козлова Н.Н. Методология анализа человеческих документов // Социологические исследования. – 2004. – № 1. – С. 14-26.
7. Малхотра Нэреш К. Маркетинговые исследования. Практическое руководство, 4-е издание.: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2007. – 1200 с. – С. 173-229.
8. Масалков И.К., Семина М.В. Стратегия кейс стадии: методология исследования и преподавания: Учебник для вузов. – М: Академический проект; Альма Матер, 2011. – 443 с.
9. Семина М.В. Метод интервью в социологии и маркетинге: учебное пособие. – М.: КДУ, 2010. – 308 с.
10. Татарова Г.Г. Качественные методы в структуре методологии анализа данных // Социология: 4М. – 2002. – № 14. – С. 33-52.

К разделам 3-4. Контент-анализ документов

1. Аверьянова Л.Я. Контент-анализ: Учебное пособие. – М.: Кнорус, 2009. – 456 с.
2. Алмакаева А.М. Категория «качество жизни» в научном и общественном дискурсе: опыт контент-анализа веб-страниц Интернета // Социология: методология, методы, математическое моделирование. – 2006. – №22. – С.19-37.

3. Барсамов В.А. Контент-анализ газетных материалов (события в Беслане) // Социологические исследования. – 2006. – №2. – С. 62-64.
4. Бурганова Л.А., Корнилов П.А. Реконструирование структуры образа военного конфликта (по материалам СМИ) // Социологические исследования. – 2003. – № 6. – С. 56-63.
5. Иванов В.Ф. Контент-анализ как формализованный метод исследования документов // Философская и социологическая мысль. – 1994. – №3. – С. 223 -230.
6. Иудин А.А. Применение методов количественного анализа в социально-политических исследованиях и прогнозировании: на примере Пакистана. – М.: Наука, 1984. – 152 с.
7. Кавевский Е.А., Саганенко Г.И., Гайдукова Л.М., Клименко Е.Н. Диалоговая система классификации и анализа текстов // Социология: 4М. – 1997. – №9. – С. 198-216.
8. Каневский Е.А., Саганенко Г.И. Концептуальное обоснование компьютерного анализа массивов с текстами // Социология: 4М. – 1997. – № 9. – С. 65-81.
9. Клупт А.М. Демографическая политика как предмет контент-анализа // Социологические исследования. – 2003. – №12. – С. 108-117.
10. Мешков П.Я. Политический мониторинг и контент-анализ в политическом исследовании. // Общая и прикладная политология: Учебное пособие / Общ. ред.: Жуков В.И., Краснов Б.И. – М.: МГСУ; Изд-во «Союз», 1997. – С. 829-837.
11. Ньюман Л. Анализ качественных данных // Социологические исследования. – 1998. – №12. – С.101-114.
12. Ньюман Л. Значение методологии: три основных подхода // Социологические исследования. – 1998. – № 3. – С. 122-134.
13. Ньюман Л. Неопросные методы исследования // Социологические исследования. – 1998. – № 6. – С. 119-129.
14. Олейник А.Н. Триангуляция в контент-анализе. Вопросы методологии и эмпирическая проверка // Социологические исследования. – 2009. – №2. – С. 65-79.
15. Рогозин Д.М. Опыт контент-анализа эмотивной лексики молодежных некоммерческих организаций // Мониторинг общественного мнения. – 2004. – № 3 (71). – С. 40-56.
16. Романовский Н.В. НЭП, контент-анализ и князь Кропоткин // Социологические исследования. – 2001. – №12. – С. 136-143.
17. Семёнова А.В., Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения / Под ред. В.А. Мансурова. – М.: Институт социологии РАН, 2010. – 324 с.
18. Степанов А.С. Метод контент-анализа и производные принципы в исследовании актуальных проблем современного общества. – М.: Изд-во МГУ, 1995.

19. Таршис Е.Я. Исторические корни контент-анализа: Два базовых текста по методологии контент-анализа. – М.: Либроком, 2012. – 160 с.
20. Таршис Е.Я. Перспективы развития метода контент-анализ // Социология 4М. – 2000 – № 15. – С. 71-92.
21. Тернер Р. Сравнительный контент-анализ биографий // Вопросы социологии. – 1992. – Т.1. – №1.
22. Тичер С., Мейер М., Водак Р., Веттер С. Методы анализа текста и дискурса / Пер. с англ. – Х.: Изд-во Гуманитарный Центр, 2009. – 356 с.
23. Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации. – М.: Научный мир, 2001. – 214 с.
24. Харченко К.В. Материальная сторона жизни в зеркале субъективных смыслов: опыт контент-анализа // Социология: методология, методы, математическое моделирование. – 2009. – №1(28). – С.129-148.
25. Чураков А.Н. Компьютерный контент-анализ. – М.: Институт социологии РАН, 1996.
26. Шалак В.И. Современный контент-анализ. Приложения в области: политологии, психологии, социологии, культурологи, экономики, рекламы. – М.: Омега-Л., 2009. – 272 с.

К разделу 5. Факторный анализ

1. Бессокирная Г.П. Факторный анализ: традиции использования и новые возможности // Социология 4М. – 2000 – № 12. – С. 142-153.
2. Бююль А., Цёфель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. – СПб.: ООО «ДиаСофтЮП», 2005 – 608 с.
3. Главные компоненты и факторный анализ // Электронный учебник StatSoft – <http://www.statsoft.ru/HOME/TEXTBOOK/modules/stfacan.html>
4. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник. – 5-е изд. – М.: Финансы и статистика, 2004. – 656 с.
5. Иберла К. Факторный анализ. – М.: Статистика, 1980. – 308 с.
6. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ. – М.: Финансы и статистика, 1989. – 215 с.
7. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – 281 с.
8. Лоули Д., Максвелл А. Факторный анализ как статистический метод. – М.: Мир, 1967. – 144 с.
9. Митина О.В., Михайловская И.Б. Факторный анализ для психологов. – М.: Учебно-методический коллектор «Психология», 2001. – 169 с.
10. Наследов А.Д. SPSS 15: профессиональный статистический анализ данных. – СПб.: Питер, 2008. – 416 с.
11. Пациорковский В.В., Пациорковская В.В. SPSS для социологов: Учебное пособие. – М.: ИСЭПН РАН, 2005. – 433 с.
12. Харман Г. Современный факторный анализ. – М.: Статистика, 1972. – 488 с.

Приложение: Зарубежные компьютерные пакеты контент-анализа

TABARI (KEDS) – программа для автоматизированного кодирования данных политических событий. Она использует встроенный анализатор для идентификации слов используемых для проведения контент-анализа. При работе программа обращается к встроенным и загружаемым словарям. Данные могут быть использованы для работы в других специализированных программах, таких как SPSS и SAS.

JFreq создает матрицы частот употребления слов в массиве, используется для проведения контент-анализа и работает с большинством языков мира. Не работает с японским, китайским и тайским языками из за принципиально отличной от большинства языков лингвистической системы этих языков. Программа позволяет исключить из массива нечитаемые символы и знаки не входящие в алфавитную базу данных. Работает на любой операционной системе.

Concordance – программа, используемая для проведения контент-анализа электронных документов, В ней можно создавать списки связанных единиц счёта, индексов, слов, при работе над электронным текстом. Позволяет обрабатывать большие массивы. Даёт возможность просматривать корреляции между словами, входящими в словарь контент-анализа. Результаты работы легко разместить в Интернет с помощью встроенных инструментов программы.

HyperRESEARCH позволяет кодировать, находить и декодировать текстовые, аудио и видео материалы. Позволяет проводить анализ таких форматов данных.

LEXIMANCER – мультязычное программное обеспечение, производящее контент-анализ больших объёмов текста, позволяющее совмещать в массиве тексты разных жанров и стилей, включая диалектические и другие нетрадиционные формы языка.

PROTAN – комплекс 30-ти программ, интегрированных в один блок, позволяющих проводить контент-анализ массивов текста с помощью встроенных словарей и идентифицировать сюжетные линии, определяя корреляции между словами словаря посредством проведения факторного анализа. Обладает большим количеством других функций.

TextQuest – описать (нужен ответ от Harald Klein)

TEXTPACK – кодирует тексты на основе созданных пользователем словарей. Производит сравнение 2-х документов, сравнивая их словарное наполнение, обнаруживает схожие отрывки внутри документов. Данные легко импортируются в такие пакеты, как SPSS или SAS.

QDA Miner является средством качественного анализа текстовых данных, аннотирования, получения и просмотра кодированных данных. Программа позволяет работать большим числом документов, содержащих как текст, так и числовые данные. QDA Miner также предоставляет широкий спектр поисковых средств для выявления корреляций кодированных данных.

WordStat – модуль анализа текста, предназначенный специально для обработки материалов, таких как журнальные статьи, литературные произведения, интервью. Как и другие аналогичные программы позволяет создавать категориальный аппарат и словарь контент-анализа. Дальнейший анализ может быть произведён с помощью создания и расчета перекрестных таблиц, а также KWIS-метода. Пакет позволяет работать и с более сложными методами статистического анализа, такими как кластеризация и многомерное шкалирование. Созданные категориальные аппараты и словари схемы могут быть применены к другим текстовым массивам в дальнейшем.

SALT – программное обеспечение, анализирующее содержимое текстового массива. Поддерживает работу со всеми языками. Определяет среднюю длину предложения, количество искомых слов, общее количество слов. Может создавать алфавитный список слов, кодировать текстовый массив в соответствии с определёнными исследователем кодами. Работает только с операционной системой Windows.

MonoConc – производит поиск единиц текстового анализа, определяет корреляции между ними в массиве.

TROPES – производит хронологически-ролевой качественный анализ текста. Также позволяет получать общую информацию по частотности использования тех или иных единиц счёта.

Qualrus является инструментом проведения качественного анализа данных, кодирующий элементы массива для дальнейшей обработки. Qualrus может быть использован для проведения полного спектра качественных исследований, в том числе в культурологическом анализе, интерпретации методов, семиотике, истории, брекетинге, эмпирическом анализе, анализе рассказов и произведений других жанров.

CAMEO – система, созданная для кодирования и аналитики политических коммуникаций. Включает в себя 20 главных событийных категорий и 200 субкатегорий, обширную базу для кодирования имён политиков в тексте.

AnnoTape – это программное обеспечение для записи и анализа аудио, видео, графических и текстовых данных, предназначенное для качественных исследований, маркетинга, журналистики средств массовой информации, архивных служб. Производит запись звуковых файлов – интервью, бесед, радиопередач напрямую на жесткий диск компьютера. Позволяет сохранять до ста часов звука вместе с текстовыми данными в одной интегрированной базе данных. Производит анализ данных, аннотирование и индексирование оригинальных звуковых и текстовых файлов. Эффективно разбивает на фрагменты массивы аудио-данных