

*А. Э. ГЕГЕР, Ю. А. ЧУПАХИНА, С. А. ГЕГЕР*

## **КОМПЬЮТЕРНЫЕ ПРОГРАММЫ ДЛЯ АНАЛИЗА КАЧЕСТВЕННЫХ И СМЕШАННЫХ ДАННЫХ<sup>1</sup>**

В статье рассматривается проблема компьютерной поддержки анализа качественных и смешанных качественно-количественных данных. Систематически анализируются возможности шести компьютерных программ, среди них две отечественные разработки — «Vega» и TextAnalyst, а также четыре зарубежных аналога — MaxQDA, NVivo, Ethnograph, Atlas.ti.

*Ключевые слова:* компьютерные программы, качественные данные, смешанные данные, текстовые массивы, классификация, кодификация.

### **Введение**

До 90-х годов XX века большинство отечественных социологов, практикующих эмпирические исследования, ограничивались использованием методов, основанных исключительно на положениях количественного статистического подхода.

С начала 1990-х годов ситуация в отечественной социологии несколько меняется: все больше исследователей обращаются к качественным методам. Ведь именно эти методы позволяют глубже понять изучаемое явление и предложить множественную интерпретацию.

Однако можно констатировать, что все более широкое использование в исследовательской практике текстовых исследований не сопровождается разработкой и применением адекватного специализированного аналитического инструментария. В целом трудоемкие текстовые исследования без адекватного компьютерного сопровождения остаются кустарным предприятием, не обеспечивающим достаточно глубокой проработки полученной информации, качественных и надежных выводов. Компьютерная поддержка классификации и анализа качественных данных весьма актуальна. При малочисленности доступных программ отсутствуют описания сравнительных аналитических возможностей, что могло бы служить руководством в выборе

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в проекте проведения научных исследований «Ресурсы компьютерных программ для анализа качественных данных», проект № 13–06–00170.

того или иного метода при ограниченных условиях. Таким образом, в современном исследовательском сообществе складывается парадоксальная ситуация: с одной стороны, существует острая необходимость в эффективных процедурах кодировки и анализа текстовых данных, с другой — исследователи не используют методы, предлагаемые для решения данных задач.

Нами был проведен контент-анализ существующих отечественных публикаций, посвященных анализу качественных данных и использованию компьютерной поддержки для такого анализа. Всего статей, посвященных анализу качественных данных, совмещению качественных и количественных стратегий насчитывается порядка 60. Это статьи, опубликованные в ведущих рецензируемых российских журналах (Социология: 4М; Социологические исследования; Журнал социологии и социальной антропологии; Социология: теория, методы, маркетинг и др.). Однако среди значительного потока публикаций нами было обнаружено лишь несколько работ, касающихся компьютерной поддержки анализа качественных данных. Это, прежде всего, работы Г.И. Саганенко и Е.А. Каневского, посвященные описанию возможностей отечественного программного обеспечения для анализа качественных данных «ДИСКАНТ» и «Vega» (Саганенко, 1997, 2008), а также статья О.Б. Божкова и А.А. Помигалова о программе для оцифровки аналоговых звуковых записей (Божков, 2008) и статья А. Горбачика, в которой описывается авторская разработка — программа «Викинг», дающая дополнительные удобства для кодирования открытых вопросов в пакете ОСА (Горбачик, 2006). Как мы можем судить по результатам этого краткого контент-анализа публикаций, тема использования компьютерной поддержки для анализа качественных данных в отечественной социологии практически не разрабатывается. Об этом же говорит и авторитетный методолог в области анализа качественных данных Г.В. Каныгин: «Социальный исследователь, который уже в полной мере на практике ощутил преимущества работы с текстовым редактором, все еще склонен относиться с подозрением к более радикальным компьютерным нововведениям в области гуманитарных наук. Как показывает опыт первого изложения идей качественного исследования на русском языке, конструктивный пафос АКД, выраженный в широком спектре компьютерных инструментов, остается невостребованным отечественными методологами» (Каныгин, 2007; 13).

## Отечественные разработки для анализа смешанных и качественных данных: Vega и TextAnalyst

Компьютерная программа «Vega» вышла в свет в 2007 году. Это совместная отечественная разработка СИ РАН и ЭМИ РАН. Разработчиками являются Г.И. Саганенко, Е.А. Каневский и К.К. Боярский. По сути, это усовершенствованная и адаптированная под операционную систему Windows программа «ДИСКАНТ» («Диалоговая система контент-анализа текстов»), которая была выпущена еще в 1994 году. Программу «Vega» отличает дружелюбный интерфейс и некоторые усовершенствования, которых не доставало в «ДИСКАНТЕ».

«Vega» — это диалоговая система классификации и анализа текстов, использующая принципы контент-анализа текстов, словари и классификаторы. Система в основном предназначена для обработки структурированной, прежде всего анкетной, информации, представляющей собой ответы респондентов на открытые и полужакрытые вопросы социологических анкет. Кроме того, система позволяет выполнять некоторые элементы анализа текста: составление словарей, подсчет встречаемости слов, поиск слов по словарю и по тексту и т.д. Система обеспечивает статистический анализ ответов на закрытые и полужакрытые вопросы.

Программа «Vega» — это инструмент не только качественного, но и количественного статистического анализа. Это в выгодную сторону отличает данную разработку от аналогов. В меню программы находится блок, названный статистическим анализом. Этот блок является достаточно мощным инструментом для программ текстового анализа.

С помощью данного блока пользователь может получать:

- одномерный статистический анализ;
- двухмерный статистический анализ;
- трехмерный статистический анализ;
- таблицу;
- многоальтернативный статистический анализ;
- шкалирование (что очень удобно, например, при работе с ответами о возрасте и уровне доходов).

В окне статистического анализа автоматически выводится среднее по полю, хи-квадрат, коэффициент Пирсона и коэффициент корреляции.

Основное же назначение данного программного продукта — анализ текстовой информации. Работа со словарями представляет началь-

ный ресурс для анализа текстовых полей. В «Веге» можно получать разнообразные словари, при этом в качестве текстовых единиц могут выступать слова, ключевые слова, фразы и полные ответы на один вопрос. Наиболее полезный ресурс — это частотный словарь, который выводит все слова, встречающиеся в поле, с указанием их частоты встречаемости. Этот инструмент позволяет найти основные темы, на которые приходится фокус внимания респондентов. Словари могут быть получены по одному текстовому полю, по нескольким или по всем текстовым полям исследования. При этом словари могут быть получены на весь массив анкет или на сформированный по одному или нескольким значениям признаков конкретный подмассив.

При **пермутационном выводе текста** дается полный перебор всех фраз из заданного текстового массива в алфавитном порядке по каждому слову, содержащемуся в соответствующем словаре. **Пермутационный словарь** очень полезен для получения исследователем начальных представлений, о чем же говорят респонденты в своих ответах, при поиске тематических классов для процедуры классификации массивов ответов на открытые вопросы.

После первичного ознакомления с текстом с помощью разного вида словарей, как правило, следующая задача исследователя — классификация суждений для выхода на более высокий уровень обобщений. При этой процедуре основная аналитическая нагрузка лежит на исследователе, программа же становится надежным партнером. В нашем исследовании ценностных ориентаций молодежи респондентов просили написать свои ценности, и обычно получаемые ответы были достаточно короткими — не более пяти слов на описание одной ценности (Гегер 2010; Саганенко 2011). Это обстоятельство помогло нам применить процедуру **автоклассификации** текста. При этой операции машина «складывает» все однокоренные слова в одну папку, а исследователь затем группирует эти папки по смыслу. Так, в нашем случае был выделен класс «Семья», в который в качестве подклассов были введены категории «родители» и «дети». После окончания процедуры классификации можно получать различные статистики (распределения и пересечения) по выбранным классификаторам.

TextAnalyst — программа, созданная отечественными разработчиками. В 1990-х годах российской компанией MegarputerIntelligence, образованной исследовательской группой искусственного интеллекта из МГУ и МГТУ им. Баумана, была создана система PolyAnalyst, на основе которой уже в 1996 году вышла в свет программа TextAnalyst.

Несмотря на то что ее интерфейс полностью на русском языке и она проста в использовании, по количеству команд она значительно уступает «Веге» и для анализа текста применяет совершенно другие принципы.

Суть программы — позволить в кратчайшие сроки ознакомиться с большим текстовым массивом, чтобы исследователь смог быстрее и лучше в нем сориентироваться. При загрузке файла TextAnalyst автоматически проводит реферирование текста и выделяет ключевые слова, выстраивая при этом семантическую сеть и формируя смысловой портрет материала. На выходе получается древовидная структура с ключевыми словами, которые проранжированы с учетом семантического веса слова и частоты встречаемости.

В этой программе нет функции кодирования материала, здесь нет возможности рассматривать материал, учитывая социально-демографические и половые признаки. Количественные данные, получаемые в результате реферирования текста, довольно скудны и оформлены достаточно неудобным для дальнейшей работы образом. Однако TextAnalyst позволяет выгружать данные и преобразовывать их в гипертекст, представляя древовидные структуры в файле формата html. Гиперссылками выступают выделенные ключевые слова, при нажатии на них приводится список всех фрагментов текста, в которых данное слово встречается.

### **Зарубежные программы для анализа качественных данных**

В данном разделе мы рассмотрим четыре зарубежные разработки, которые имеют большую научную популярность за рубежом, причем в совершенно разных областях: MaxQDA, NVivo, Ethnograph, Atlas.ti. В России крайне сложно найти работы, в которых бы описывался опыт применения программ в научных исследованиях, не говоря уже об учебных пособиях, рассказывающих о том, что это за программы и как ими пользоваться. Опираясь на научные работы зарубежных ученых, сайты разработчиков программ по анализу качественных данных и вспомогательные материалы, которые устанавливаются вместе с программами, мы подробно описали каждую из них, представили логичный и последовательный алгоритм работы. Учитывая ограниченный объем данной статьи, мы остановимся только на ключевых моментах функционирования этих разработок. Такие программы позволяют работать с разного формата данными: текстами, аудио-, ви-

деоматериалами. Мы рассмотрели работу именно с текстом, а вернее, как анализируются ответы на открытые вопросы, полученные в ходе смешанного исследования.

Все без исключения описанные программы имеют как общие особенности, так и свои «изюминки». В первую очередь, важно отметить, что чем богаче качественный материал и в объеме, и в своем разнообразии, тем в большей мере удастся использовать ресурсы программ. Кроме того, работа исследователя в большинстве из них может быть описана следующим алгоритмом: запуск программы, создание рабочего проекта, ввод данных (посредством копирования текста) или же загрузка данных с компьютера, кодирование текстового материала, присвоение как конкретным файлам, так и отдельным элементам различных признаков. Далее посредством фильтрации и создания поисковых запросов — отбор материала, построение корреляционных таблиц и матриц — тот этап работы, где качественные данные представляются в виде количественных показателей. В некоторых программах возможно построение графических рисунков. Ну и наконец, по желанию исследователя — если он нашел интересные закономерности и взаимосвязи и готов работать над выводами, можно завершить работу и приступить к написанию аналитического отчета. Если же есть смысл и видится возможным продолжить тщательное изучение материала, то не стоит останавливаться и можно продолжить аналитическую работу.

Цифры, встречаемые в анализируемом материале, все программы воспринимают как текст. Числовая информация, участвующая в статистических расчетах, выводится из общего количества закодированных фрагментов текста, заметок, аннотаций, присвоенных признаков. Чаще всего такие программы сделаны так, чтобы собранную статистическую информацию можно было выгрузить и продолжить анализировать в SPSS или в Excel.

Функционально самой сложной и продуманной является программа NVivo. Работа в ней осуществляется по вышеописанному алгоритму, однако помимо собственных данных в отдельный раздел можно внести интересные находки и полезный материал из других источников. Такой материал называется здесь «экстерналии». Его включение в собственный проект позволяет расширять научную картину и хранить существенную информацию под рукой, имея возможность связать ее с собственными разработками. В процедуре кодирования участвуют узлы (Nodes), которые выступают в качестве кодов,

присваиваемых отдельным фрагментам текста. Доступна и функция автокода, однако в этой программе она применима лишь к тем файлам, в которых обозначены заголовки фрагментов — именно они и участвуют в процедуре автокодирования.

Анализ данных осуществляется благодаря различным типам поисковых запросов. Благодаря им можно искать закодированные одним кодом фрагменты, строить матричные таблицы на основе связи между элементами, сравнивать данные различных признаков.

Один из самых подробных разделов программы — работа с динамическими и статистическими моделями. На основании получаемых кодов, взаимосвязей и классификаций можно создавать графики, обновлять их при изменении данных, применять к ним анимацию, которая пригодится при демонстрации своих разработок.

MaxQDA по принципам функционирования очень похожа на NVivo. Большое ее преимущество состоит в том, что она выпущена на русском языке, поэтому освоить ее можно довольно быстро. Среди всех изученных зарубежных программ MaxQDA — единственная, которая позволяет прикоснуться к анализу смешанной информации. Закодированным фрагментам текста можно присваивать вес (от 0 до 100), а затем сортировать данные по этому признаку. Если, как в нашем случае, в анкете респонденты дают оценочные характеристики, выраженные по 100-балльной шкале, значением веса можно обозначить именно эту оценку и, отталкиваясь от нее, сортировать данные, ну а затем выгрузить их и продолжить анализ в специализированных статистических программах.

Еще одна важная особенность — функция визуализации, обозначающая здесь не построение графиков, а создание матрицы кодов и крестовых таблиц, которые в очень наглядном и информативном виде отображают результаты проделанной работы. Данные в матрице могут быть представлены символично (в виде больших и маленьких точек в зависимости от частоты встречаемости), в частотном и процентном соотношении. Крестовые таблицы отображают связи между кодами и признаками. Среди всех программ MaxQDA лучше всех ориентирована на подготовку данных к дальнейшей статистической обработке в таких программах, как SPSS, Excel.

Разработчики Atlas.ti позиционируют программу как идеально оснащенное рабочее пространство исследователя. Она позволяет работать с огромным количеством разноформатного материала — предоставляется на выбор больше 25 форматов для загружаемого файла,

включая даже географические карты. На данный момент это самая «современная» программа, поскольку она выпущена даже в качестве мобильного приложения в AppStore.

Работа в Atlas.ti осуществляется по тем же принципам, что и в NVivo и MaxQDA. Она также позволяет строить таблицы корреляции элементов между собой, предоставляя возможность использовать обширный список логических функций. Но он будет действенен в том случае, если анализируются большие объемы данных и мест пересечений различных кодов будет много.

Здесь, как и в NVivo, возможно создание графического рисунка, поэтому особую актуальность имеют так называемые networks (связи между элементами), которые исследователь создает самостоятельно и включает в графическую модель. В модели можно обозначить связь двух на первый взгляд не связанных между собой фрагментов (например, если требуется акцентировать внимание на том, что один элемент противоречит другому или абсолютно ему противоположен) при помощи функции гиперссылок и команд Startanchor (начало связи) и Finishanchor (конец связи), которые делают фрагменты взаимосвязанными.

Еще один момент, на который хочется обратить внимание, — работа с признаками. Если в других программах возможно создание признака и присвоение ему нескольких значений одновременно (например, пол, а далее на выбор: мужской/женский), то в Atlas.ti создаются отдельные family-группы для каждого отдельного признака. В случае с мужским/женским полом их было бы создано две. Если бы респондентов пришлось классифицировать по возрасту, месту проживания, социальному статусу, то их было бы намного больше. В этой программе создание классификаций — достаточно трудоемкий процесс.

The Ethnograph — очень простая по оформлению программа, которая устроена емко и логично. Специфика работы в ней состоит в том, что изначально текстовый материал обязательно форматируется в файле формата.doc или.rtf — устанавливаются необходимые настройки абзаца и величина отступов. Об этом нюансе программа обязательно напоминает при загрузке текста. А после того, как правильно оформленный текст загружается в рабочее окно программы, он снова реформатируется в соответствии с наиболее подходящим ему шаблоном, который предлагает Ethnograph.

Кодирование осуществляется по тем же принципам, что и в вышеописанных программах. Функции автокода здесь нет, как и возможности построения графических моделей и матричных таблиц.



Важно отметить, что рабочий язык программы — английский. Имена кодам и названиям файлов следует писать либо на латинице, либо на английском языке.

Самое большое внимание в Ethnograph уделено таким функциям, как поисковые запросы и фильтры. Именно на них выстраивается вся исследовательская работа, результаты этих операций выгружаются для дальнейшей статистической обработки в других программах. В общем и целом нужно признать, что данная программа значительно отличается и уступает NVivo, MaxQDA и Atlas.ti. В сравнении с ними она сильно ограничена в своем функционале.

### **Заключение**

Мы реферативным образом рассмотрели шесть компьютерных программ для анализа качественной и смешанной качественно-количественной информации. В принципе, наличие таких разработок показывает, что производители ПО откликаются на запросы исследователей и создают современные программы для решения самых сложных аналитических задач. Мы убеждены, что потенциал этих программ значителен, используя его, мы можем совершенно по-новому взглянуть на весь исследовательский процесс.

Однако распространению и популяризации этих программ в России мешает один барьер: их высокая стоимость и их доступность. Среди тех программ, которые мы описали, лишь MaxQDA и NVivo есть в свободном доступе в сети Интернет. Что касается других зарубежных программ, на официальных сайтах доступны лишь демоверсии, которые также позволяют осуществлять в них работу, однако с серьезными ограничениями. А платные версии у нас в стране практически не востребованы. Российские разработчики, в свою очередь, не стремятся хоть как-нибудь продвигать свои продукты.

### **Источники**

*Божков О.Б., Помигалов А.А.* Заметки о технических проблемах полевых исследований // Социология: 4М. 2008. № 26. С. 176–189.

- Гегер А. Э.* Выявление индивидуальных и групповых ценностей в группе молодежи. Релевантные методические решения // Социологические исследования. 2010. № 1. С. 132–141.
- Горбачик А. А.* Компьютерная обработка текстов в качественных и количественных социологических исследованиях // Социология: теория, методы, маркетинг. 2006. № 1. С. 124–133.
- Каневский Е. А., Саганенко Г. И., Гайдукова Л. М., Клименко Е. Н.* Диалоговая система классификации и анализа текстов // Социология: 4М. 1997. № 9. С. 198–216.
- Каныгин Г. В.* Инструментальные средства и методологические принципы анализа качественных данных // Социология: 4М. 2007. № 25. С. 70–98.
- Саганенко Г. И., Гегер А. Э., Степанова Е. И.* Потенциал рефлексивных методов в выявлении релевантных ценностей // Социология: 4М. 2011. № 32. С. 75–103.
- Саганенко Г. И., Каневский Е. А., Боярский К. К.* Контексты эмпирического познания в социологии и возможности программы «Vega» // Телескоп. 2008. № 6. С. 43–45.