

О.Б. БОЖКОВ

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В СОЦИОЛОГИЧЕСКОМ ИССЛЕДОВАНИИ

Новая рутина не намного лучше старой

Когда в начале 90-х годов благодаря поддержке Фонда Дж. Сороса в нашей группе появились “свои собственные” компьютеры (по тем временам мощные профессиональные машины PC IBM 386DX2, HDD 120Mb, 4Mb RAM), мы, не очень сильно задумываясь, “начинили” их тем программным обеспечением, которое видели у других: Windows 3.1, Lexicon, SPSS for DOS, чуть позже поставили Paradox 4.0 for DOS. Машины использовали поначалу совсем незастойливо: текстовые документы (анкеты, разнообразные инструкции, кодировальные листы и т.п.) набирали в Lexicon или в Word for Windows (в зависимости от личных симпатий сотрудников); массивы данных вводили посредством SPSS, в его формате, естественно, здесь же и обрабатывали данные и выводили результаты на печать. Диаграммы при необходимости строили, либо используя графику SPSS, либо аналогичные средства Word for Windows 2.0 (позже – 6.0). Затем вручную строили агрегированные итоговые таблицы, вручную же (то есть заново) набирали их в Word for Windows. Иными словами, вплотную столкнулись с новой рутинной (переписывание распечаток SPSS в тексты отчетов, досчитывание и/или пересчет – вручную или с помощью встроенного калькулятора – некоторых показателей и т. п.)

Впрочем, рутины хватало и на ранних стадиях исследования. Когда, например, анкета была уже создана, возникала необходимость организовать структуру массива данных, а значит, ту же самую анкету надо было опять-таки набирать заново, но уже не в текстовом редакторе, а в редакторе одного из статистических пакетов, например, в SPSS (Модуль Data Entry). Теоретически SPSS 4.0 (for DOS) предоставляет возможность вводить полный текст вопроса и тексты вариантов ответов к вопросам на русском языке с тем, чтобы использовать их затем при распечатке результатов счета: одномерных и двумерных распределений, условных средних и т.д. Однако приведение в божеский (пригодный для чтения) вид таблиц из SPSS в формате Word for Windows также оказалось занудным и весьма трудоемким делом.

Таким образом, избавляя от определенных трудоемких операций на стадии разработки анкет, компьютер порождает дополнительные заботы на стадиях формирования массива эмпирических данных и, главное — формирования отчета.

Божков Олег Борисович – старший научный сотрудник Санкт-Петербургского филиала Института социологии Российской академии наук.
Адрес: 198147 Санкт-Петербург, Измайловский проспект, д. 14. **Телефон:** (812) 316-21-62. **Факс:** (812) 316-29-29. **Электронная почта:** oleg@solaris.spb.su

Мы не были совсем уж "чайниками". Как-никак за плечами – более чем десятилетний опыт работы на БЭСМ-6 в Институте социально-экономических проблем АН СССР, с одной стороны, а с другой — имелся определенный вкус к методологической рефлексии относительно процедур социологического исследования.

Опыт контент-анализа телевизионных программ убедил нас в некоторых неоспоримых преимуществах реляционной СУБД (например, Paradox) перед плоскими файлами статистических пакетов. Прежде всего, речь идет о возможности работать со сложными (многотабличными) моделями данных. Однако более глубокое осмысление специфики каждого из названных программных продуктов было еще впереди. Тем не менее, постепенно формировалось близкое к оптимальному сочетание различных программных продуктов, "покрывающее" наши основные потребности в рамках той предметной области, которой мы занимались.

Первым шагом стала попытка избавиться от двойного набора анкеты: сначала в текстовом редакторе для тиражирования (анкета как рабочий документ), а затем при формировании структуры базы данных под Paradox 4.0 for DOS (анкета как структурированный "накопитель" информации). Текст, набранный в Word for Windows, конвертировался в формат TXT и импортировался под Paradox, где с помощью специально созданного для этой цели скрипта (программы на языке программирования данной СУБД – PAL) он анализировался и использовался для "автоматического" формирования структуры таблиц Paradox. Чтобы облегчить анализ текстовых файлов, пришлось ввести определенные (и достаточно строгие) правила формализации. Например, те слова, которые составляли имена полей таблицы, требовалось писать в тексте анкеты только прописными буквами. Для написания содержания вариантов ответа было установлено обязательное использование строчных букв и в качестве разделителя между разными вариантами – символа "точка с запятой". И хотя это не решало всех проблем, от двойного набора анкеты удалось избавиться. Один из новых типов рутин мы, кажется, победили.

Когда мы начали осваивать Paradox 5.0 for Windows, то попытались вообще отказаться от набора анкеты в Word for Windows, тем более, что в структуре Paradox имеется такой объект, как отчет, который в принципе позволяет о печатать текстовый документ в хорошем виде, вполне пригодном для последующего тиражирования анкеты к массовому опросу.

В процессе решения этой задачи мы впервые начали отчетливо осознавать необходимость одновременно осмыслить и соотнести, с одной стороны, основные принципы организации реляционной СУБД, а с другой — природу и структуру самого анкетного метода. Задачи компьютерно-технологической поддержки социологических исследований заставили нас не только задуматься над такими вещами, которые казались совершенно очевидными, но и пересмотреть свое отношение к ним. Поэтому, в частности, возникла потребность более строго описать сам процесс социологического исследования в целом и каждого из его этапов в отдельности.

Процесс эмпирического исследования с точки зрения технологии

Детальное описание канонической (в отечественной социологии) программы социологического исследования приводят Т.Заславская и Р.Рывкина [1, с. 144-207]. Взяв его за основу, посмотрим, в какой мере компьютерные технологии проникли в современную социологическую практику. При характеристике традиционных технологий мы обращались к социологическим учебникам и литературе, издававшейся, в основном, в советский период и имевшей целью дать достаточно полное представление о современной социологии; что же касается компьютерных технологий, здесь мы опирались на опыт профессионального общения с коллегами, а также использовали названия программных продуктов, упоминаемых в новых учебниках по социологии [2, с. 218.].

Таблица 1

“Технологические операции”, составляющие процесс социологического эмпирического исследования

Наименование этапа исследования	Операции	Традиционные технологии	Компьютерные технологии
Формулировка темы (проблемы) исследования	Описание проблемной социальной ситуации и формулировка научной проблемы	Изучение литературы, прессы, личные наблюдения, рассуждения, дебаты с коллегами	1. Системы телекоммуникации 2. Тематический и словарный поиск в глобальных сетях типа Internet 3. Средства локального индивидуального накопления знаний (текстовые процессоры*, базы данных и т.п.), на отдельной РС или в локальной сети
Формулировка целей и задач исследования	Формулировка целей и задач исследования	Рассуждения, дебаты с коллегами	То же
Системное описание объекта исследования	Выделение (определение и уточнение) основных понятий	Анализ научной литературы, изучение прессы, личный опыт, рассуждения, дебаты с коллегами	То же

* Среди текстовых процессоров, чаще всего используемых социологами: Lexicon, MS Word, Word Perfect, WinWord, MultiEdit.

Продолжение таблицы 1

	Теоретическая и эмпирическая интерпретация понятий	То же	То же
	Операциональные определения	Логический и семантический анализ	Отсутствуют
	Операционализация понятий	Логический, семантический и лексический анализ	Отсутствуют
	Выдвижение и обоснование гипотез	Логический и семантический анализ	Отсутствуют
Разработка стратегического плана исследования		Логический и семантический анализ	ORGANIZER
Построение схемы выборки		Логический и семантический анализ	Отсутствуют
Методический инструментарий исследования	Разработка рабочих документов	Логический, семантический и лексический анализ	Текстовые редакторы, DA-система, CATI, CAPI, ORGANIZER
Рабочий план исследования		Логический и семантический анализ	
Сбор эмпирических данных	Наблюдение, опрос	Наблюдение, опрос, сбор (отбор) документов	CATI, CAPI, РИСЭП
Подготовка данных к обработке		Кодирование, ввод данных в ЭВМ	CATI, CAPI, SAS, SPSS, EXCEL, Paradox, DA-система
Обработка и анализ данных		Построение распределений, вычисление корреляционных матриц	CATI, CAPI, SAS, SPSS, EXCEL, Paradox, DA-система
Обобщение результатов	Построение и обоснование выводов	Логический и семантический анализ, написание текстов	Текстовые и графические редакторы
Оформление результатов исследования		Перепечатка текстов, ручное построение графиков, диаграмм	Текстовые и графические редакторы, DA-система**

** Разработка фирмы "Контекст" (Москва). Разработчик математических основ метода детерминационного анализа (ДА) С.В.Чесноков.

Как видно из табл. 1, перечень программных продуктов, используемых социологами, весьма невелик. Кроме того, для некоторых этапов исследования трудно подобрать более или менее адекватное программное обеспечение. Эта весьма серьезная проблема, на наш взгляд, еще недостаточно осознана. Речь идет о поиске программных продуктов, наиболее адекватных предметной области социологии.

Работа социолога — исследователя начинается с **формулировки темы (проблемы), целей и задач исследования, определения и уточнения основных понятий** — теоретических концептов, установления связей между ними и определения содержания этих связей (логических, семантических, функциональных и т.п.). Эта интеллектуальная, творческая работа, требующая достаточно широкой эрудиции, хорошего знания теоретических оснований социологии, технологически поддерживается очень слабо — в лучшем случае автоматизированными системами библиографического поиска. Мощным средством технологической поддержки на данном этапе становится в последнее время глобальная компьютерная сеть Internet и ей подобные с хорошо разработанными средствами поиска. Однако, вероятно, существуют и другие средства компьютерной поддержки этой стадии исследовательской работы, например, базы данных, имеющие дело с иерархическими и древовидными структурами, а также с данными графического (или иконографического) характера.

На этапе **системного описания объекта исследования**, а также теоретической и эмпирической интерпретации понятий картина аналогична. Разница в том, что здесь исследователь конкретизирует проблему и вынужден более глубоко заниматься самостоятельным анализом и собственно творчеством. Здесь рамки компьютерной поддержки как таковой резко сужаются, и из всех программных продуктов наиболее адекватным оказывается текстовый процессор.

Впрочем, прямая компьютерная поддержка эмпирической интерпретации и операционального определения понятий, выдвижения и формулировки гипотез вряд ли возможна. Здесь используются преимущественно догадки, интуиция исследователя. Естественно, все эти операции и процедуры трудно формализуемы, подвержены сильному влиянию индивидуальных (интеллектуальных, психологических и т.п.) особенностей исследователей. Однако и на данной стадии исследования, думаем, необходимы такие средства, которые могли бы четко фиксировать некоторые важные результаты, накапливать их для последующего анализа.

Разработка стратегического плана исследования в принципе может быть поддержана такими программными средствами, как, например, "Organizer", однако, в социологической практике такого рода средства используются редко.

В отечественной социологической литературе описан опыт использования преимущественно вычислительных алгоритмов (факторный, кластерный анализ) при **построении и расчете выборки** [2]. Однако вряд ли это можно считать обычной, распространенной практикой. Как правило, здесь рассматриваются фундаментальные, стратегические проблемы, связанные с построением территориальной выборки.

На Западе в течение последних 15-20 лет параллельно с разработкой универсальных статистических программ идет развитие специализированных систем, направленных на автоматизацию всех этапов социологического исследования. Имеются в виду программы, включающие функции текстового редактора, ориентированного на автоматизацию подготовки анкет, возможности CATI и CAPI¹, автоматического считывания данных с анкет, напечатанных традиционным способом, подготовки таблиц с высококачественным дизайном, статистического анализа данных [3, с. 452].

Для отечественных социологов CATI, CAPI и Bellview и другие программные средства – скорее экзотика, нежели привычные, стандартные инструменты. В начале 80-х годов в журнале “Социологические исследования” промелькнула информация о чудо-машине “РИСЭП” (Регистратор информации по социально-экономическим проблемам), которая, по утверждению разработчиков, существенно облегчала полевую работу и обеспечивала ввод данных в ЭВМ непосредственно в ходе опроса. Скорее всего, эта машина не получила признания.

Особо следует отметить ДА-систему — один из первых отечественных программных продуктов, поддерживающих процесс исследования от создания словаря переменных и шкал до “генерации отчетов”. Однако и этот пакет ориентирован прежде всего на этапы обработки и анализа специфических данных, свойственных гуманитарным и общественным наукам [4, 5].

Этап **обработки и анализа эмпирических данных** наиболее обеспечен компьютерной поддержкой. Кроме SPSS (Statistical Package for Social Sciences), SAS (Statistical Analysis System) и Statgraphic, на данном этапе используется большое количество разнообразных программных продуктов как импортного, так и отечественного производства, как профессиональных статистических пакетов, так и “доморощенных”, полупрофессиональных и просто любительских программ. Наиболее квалифицированные социологи предпочитают использовать SPSS, SAS или SYSTAT.

Обобщение результатов, построение содержательных, логических выводов вынуждает социолога снова использовать ЭВМ в качестве усовершенствованной пишущей машинки, ибо пока нет программных средств, способных взять на себя эти творческие функции.

Итак, пространство социологического исследования весьма неравномерно обеспечивается современными компьютерными технологиями. Очевидно также, что разнообразие действий, осуществляемых социологом в процессе исследования, диктует применение достаточно широкого ассортимента программных средств. Проблемы вроде бы нет, поскольку рынок программных средств сегодня чрезвычайно богат и, в принципе, каждый может установить на свой компьютер любой набор software.

Однако проблема есть. Во-первых, разные программные продукты не всегда “ладят” между собой, во-вторых, они не обязательно обеспечены

¹ CATI – Computer assisted telephone interviewing, CAPI – Computer assisted personality interviewing.

удобным для пользователя интерфейсом, в третьих, довольно часто (особенно в области гуманитарных наук) программы не адекватны предметной области пользователя. Конечно, этим не исчерпывается список проблем. Что мы имеем в виду, когда используем понятие “технологическая поддержка” социологического исследования? Речь идет именно о последовательной **технологии**, включающей в себя средства фиксации и оформления разнообразных объектов с учетом их “природы”. Это – обеспечение необходимых преобразований объектов, вычислительных и аналитических процедур, а также фиксации и оформление результатов вычислений в формате, удобном для пользователя и возможных потребителей (читателей, заказчика и т.п.) и, главное, — обеспечение плавных, без дополнительных (и излишних) затрат, переходов от одних операций над объектами к другим, от одной стадии исследования к другой, а при необходимости и от одного исследования к другому.

Важными шагами, предваряющими разработку такой технологии, являются, во-первых, анализ самого процесса исследования, во-вторых, осмысление метафор, лежащих в основе каждого программного продукта, и установление степени их адекватности метафорам предметной области [6].

От компьютерной технологии к анализу ее предметной области

В свое время автор и его коллеги предпринимали попытку формализованного описания процесса социологического исследования и поиска подходов (тогда сугубо теоретического) к созданию индивидуальной базы социологических знаний, которая по замыслу должна была включать: словарь основных теоретических понятий, результаты их операционализации, и, наконец, методические и процедурные решения, то есть анкетные вопросы и их шкалы, а также алгоритмы фиксации перехода от теоретических понятий к их эмпирическим индикаторам. Речь шла об интегральной информационной технологии, в рамках которой профессиональная деятельность социолога могла бы поддерживаться как непрерывный, технологически единый процесс накопления знаний и опыта в интересующей его предметной области [7, с. 58]. Теперь же, опираясь на предыдущий опыт, мы решали задачу эффективной компьютерной поддержки всего процесса эмпирического исследования.

Что должно храниться (накапливаться) в тезаурусе? Казалось бы ответ очевиден: “Вопрос анкеты”. Но то, что мы называем вопросом анкеты, не обязательно и не всегда выглядит как вопросительное предложение. Иногда это утверждение, иногда задание (предложение в повелительном наклонении). Это может быть, наконец, комбинация утверждения и задания или утверждения и собственно вопроса. Например, *“Многие считают, что взять что-то для личных нужд на производстве, скажем, кусок мыла или какой-либо инструмент или пачку писчей бумаги – вовсе не воровство (утверждение). Согласны Вы с этим или нет? (собственно вопрос)”* На месте собственно вопросительного предложения может быть использовано задание, на-

пример: "Отметьте на линейке, нарисованной ниже, свою позицию по этому поводу". И в первом, и во втором случае мы работаем с предложениями, описывающими определенную жизненную ситуацию.

Феномен, именуемый вопросом анкеты, как правило, не исчерпывается каким-либо одним предложением. Часто он включает в себя целый ряд предложений, например, совокупность вариантов ответа или единицу измерения, в которой может быть выражен ответ, или "линейку", на которой респондент может зафиксировать свою позицию и т.п. Иногда анкетные вопросы сопровождаются комментариями или методическими указаниями. В любом случае необходимо точно определить, что такое "вопрос анкеты".

Из чего состоит вопрос анкеты?

Вопрос анкеты включает в себя три обязательных элемента: (1) краткое имя вопроса (или имя поля базы данных), которое мы назвали "мнемоформулой" вопроса; (2) полный текст или формулировку вопроса (элемент текстового документа) и, наконец, (3) обозначения типа (атрибут) поля в базе данных.

К необязательным (сменным) элементам вопроса можно отнести: (1) единицы измерения; (2) шкалы или совокупность вариантов ответов; (3) комментарии или пояснения. При анализе различных социологических анкет мы обнаружили, что комментарии бывают по меньшей мере двух типов. Один помещается, как правило, перед формулировкой вопроса и представляет собой *преамбулу* или введение, а другой, чаще всего помещаемый после формулировки вопроса, имеет характер *методического указания* респонденту или интервьюеру.

С точки зрения содержания, любой вопрос социологической анкеты есть описание того или иного фрагмента социальной реальности с помощью "общепринятых ассоциаций" (или типизаций), соотносимых с определенным теоретическим понятием (концептом). Когда речь идет о «валидности», «достоверности» или «надежности» социологического вопроса, имеется в виду прежде всего мера соответствия между описанием фрагмента реальности на естественном языке и его же описанием на языке теории.

Каждый из элементов социологического вопроса вносит «свой вклад» в это описание и обеспечивает необходимую и достаточную меру понимания вопроса для всех участников взаимодействия: исследователя, интервьюера и респондента. Например, вопрос "Сколько вам лет?" сопровождается указанием единицы измерения и/или правилами использования этой единицы: "Укажите полное число лет, исполнившихся вам на данный момент" или просто: "___ лет". В традиционном вопросе о душевом доходе, кроме подробного описания способа подсчета душевого дохода обязательно приводится единица измерения: либо "рублей", либо "тысяч руб." — хотя бы для того, чтобы не затруднять ни респондента, ни интервьюера написанием "лишних" нулей.

В процессе разработки и опытной эксплуатации нашей системы возникли мелкие, но требующие решения проблемы. В частности, оказалось, что

вопросы не всегда могут быть просто перенесены (скопированы) из одной анкеты в другую. При обсуждении содержания и структуры новой анкеты типична фраза: "Ну, а сюда пойдет вопрос "о политических ориентациях"... из ноябрьской анкеты прошлого года". Однако, когда доходит до дела, выясняется, что вставить сюда данный вопрос нельзя: какой-нибудь его элемент приходится изменять (например, редактировать). Причем, изменения могут касаться и шкалы, и комментария, и типа вопроса (признака, шкалы) или даже самой формулировки.

Причины требуемых изменений различны. Они могут быть вызваны тем, что в новом исследовании ставятся другие задачи, или при аналогичных задачах не требуется высокая точность измерения, а достаточно приближительной, «качественной» оценки, или тем, что в качестве объекта выступает другая категория населения.

Тогда возникает резонное сомнение: в новой анкете мы используем тот же самый вопрос или уже другой? Во всяком случае появляется новое понятие — «модификация» или «вариант» вопроса. И, как следствие этого, новая проблема — что же накапливать в тезаурусе: отдельные вопросы, блоки вопросов, вопросы с их модификациями и комментариями по поводу каждой из них? Не будет ли это дублированием или неоправданным порождением «лишних», избыточных сущностей? С другой стороны, все вариации отдельного вопроса — это практический и подчас довольно ценный опыт, который полезно сохранить. Наконец, надо учитывать, что в исследовательской практике анкетные вопросы изменяются еще и потому, что происходит новое, более глубокое, более точное осмысление жизненных ситуаций или теоретических понятий. Фиксация таких изменений (в частности, сохранение более ранних вариантов) имеет историческую и методологическую ценность.

Очевидно, что мы имеем дело с одним и тем же вопросом, а разные шкалы показывают, что в первом случае данный вопрос выполняет в анкете роль ориентира (фоновой характеристики), во втором — он выступает средством получения детальной, концептуальной информации, где позиции шкалы свидетельствуют о масштабности предпринимательской деятельности. Очевидно также, что вопрос в первой редакции применим в «массовой» анкете, во второй — уместен при обращении к специфическому контингенту — предпринимателям и руководителям предприятий.

Три ипостаси анкетного вопроса

Хотя на интуитивном уровне ясно, что такое анкетный вопрос, тем не менее, само слово "вопрос" в данном контексте не совсем точно отражает суть дела. Во-первых, это — описание определенной жизненной (или социальной) ситуации, если угодно, ее вербальная модель. Это описание, как правило, и есть вариант эмпирической интерпретации понятия, посредством которого данная модель соотносится с однозначно не эксплицируемым теоретическим понятием. Данное качество анкетного вопроса можно назвать описательным или концептуальным. В определенном смысле вопрос — это

метафора репрезентируемого в вопросе понятия или совокупность описывающих понятие "общепринятых ассоциаций". С этой точки зрения первую ипостась можно квалифицировать и как метафорическую.

Во-вторых, вопрос является ключевым элементом текстового документа — обращенного вовне (к интервьюеру или респонденту), выражаемого средствами естественного языка. Это диалоговая или дискурсивная его ипостась.

И, в-третьих, наряду с функциями "содержательной модели" социальной ситуации и дискурса, вопрос выполняет также функцию формирования признакового пространства исследования — техническая или инструментальная его ипостась. С этой точки зрения целесообразно различать вопросы, возмущающие (1) один единственный признак (например, пол, образование, возраст, тип жилья и т.п.); (2) совокупность однородных, однотипных признаков (например, различные стороны удовлетворенности жильем или работой, частота обращения к различным видам деятельности и т.п.); (3) совокупность (или набор) разнородных, разнотипных признаков.

Примером последнего типа служит вопрос об удовлетворенности различными сторонами жилища, которые не могут быть описаны единообразными шкалами. Правда, ситуации, описываемые разными шкалами, обычно оформляются в анкетах в виде отдельных вопросов. Другим примером может служить вопрос о способах проведения свободного времени. В данном случае, кроме явно представленных в анкете признаков — способах проведения свободного времени, могут быть построены серии производных типологических признаков — например, количество отмеченных респондентами публичных и домашних форм досуга и т.п. При этом подразумевается, что для исследователя актуальны понятия "публичные" и "домашние" формы досуга и он уверенно умеет их различать вне зависимости от того, умеют ли их так же хорошо различать респонденты. Более того, исследователь считает, что сами эти свойства имманентны занятиям, заполняющим свободное время.

Здесь нам понадобятся понятия эксплицитных и латентных признаков. *Эксплицитными* мы называем признаки (шкалы), явно представленные в исследовательском документе (например, в анкете), тогда как *латентными* — такие, которые, не будучи представленными респонденту, могут быть вычислены на основе эксплицитных или приписаны респондентам по определенным правилам.

Каждой ипостаси вопроса присущ свой собственный язык. Традиционный язык методики и техники обычно используется, когда мы имеем в виду "дискурсивную" (*открытый — закрытый*) и "концептуальную" (*программный — вспомогательный — фильтрующий*) его ипостаси.

Понятия признака и шкалы отражают иные ипостаси вопроса. В повседневной социологической практике эти понятия сплошь и рядом используются как синонимы. Однако тождество понятий "вопрос", "признак" и "шкала" является частным случаем, общее же правило — их относительно автономное существование.

Цепочка "вопрос — признак — шкала" неполная, в ней, очевидно, пропущен важный элемент. Вопрос — всегда вопрос о чем-то. Признак — это

свойство, качество, состояние, показатель (или указатель) этого “чего-то”. Это “что-то” имеет место в реальности (статус этой реальности — самостоятельный вопрос, требующий отдельного обстоятельного рассмотрения) и обозначается с помощью какого-либо понятия. Чаще всего содержание понятия неустойчиво и трудноуловимо. Иногда (крайне редко) понятие просто и достаточно очевидно связано с известными признаками (например, пол, возраст, образование), как правило же, оно непосредственно не выражается каким-либо одним признаком.

Кроме того, признак всегда встроен в теоретическую конструкцию и репрезентирует определенное теоретическое понятие. Однако признак — тоже понятие, но более низкого уровня, чем то, которое репрезентируется с его помощью. С другой стороны, именно признак выступает в качестве элемента базы данных, где ему однозначно соответствует одно-единственное из всех возможных состояний, фиксируемое в поле (единичной) записи в базе данных.

Шкала же есть совокупность всех конкретных значений (список, перечень состояний), обозначающих диапазон (или границы) эмпирической интерпретации признака. Ситуацию запутывает то обстоятельство, что и шкала, и признак “привязаны” к полю записи в базе данных. То, что мы называем признаком, есть имя поля записи (или понятие “низкого уровня”), а то, что называем шкалой, составляет содержимое (или значение) этого поля.

Когда речь идет о “технической” ипостаси анкетного вопроса, в традиционном языке методики и техники обсуждаемым понятиям соответствуют иные классификационные основания: во-первых, *номинальный — ранговый (порядковый) — интервальный — метрический*; во-вторых, *альтернативный — многоальтернативный*. Оба основания акцентируют внимание на “устройстве” совокупности значений признака, на том, как эти значения соотносятся друг с другом и взаимодействуют.

В первой версии нашей системы, исходя из способа и характера порождения шкал, мы использовали их рабочую классификацию, которая представлена в табл. 2.

Таблица 2

Рабочий вариант классификации социологических шкал

Класс шкал	Содержание	Комментарий
Предопределенные	Шкалы, которые основываются на общепринятых единицах измерения или на ряде натуральных чисел	Возраст в годах, доход в тыс. руб., метраж жилья, число членов семьи и т.п.
“Универсальные”	Шкалы, относительно индифферентные к конкретному содержанию вопроса	Шкалы типа “Да — Нет”, а также стандартные шкалы удовлетворенности, согласия, оценок, отношения и т.п.
Конструируемые или определяемые	“Уникальные” шкалы, разрабатываемые для определенных, конкретных вопросов	Любые типологические шкалы

Объекты формализации

Сам по себе вопрос не является объектом нашей модели. На эмпирическом (в частности, технологическом) уровне социолог почти не работает с вопросом, который, формируя признаковое пространство, остается как бы во “внешнем”, не формализуемом, контуре исследования. Это, скорее, некоторая исследовательская метафора, лишь многообразная форма для содержания, которое сосредоточивается и эксплицируется в используемых исследователем понятиях, признаках, шкалах, правилах и комментариях. Следовательно, объекты, элементы модели надо искать именно здесь.

Бесспорно, ключевым элементом модели является *признак, характеристика, свойство* того или иного изучаемого объекта. Именно над признаком исследователь производит наибольшее число операций, начиная с определения, операционализации и кончая обработкой (вычислениями) и обобщениями. Концептуализации подлежат признаки, явно представленные респондентам, так как распределения ответов по пунктам шкалы (вариантам ответов) свидетельствуют о “поведении”, состоянии и качественной определенности отображаемого ими признака (свойства, качества изучаемого явления). Поэтому шкала есть способ представления и конкретизации признака (если угодно, первый слой его интерпретации).

Тем не менее, на наш взгляд, шкала обладает статусом самостоятельного объекта именно в силу своей относительной автономности от признака. В частности, одна и та же шкала может “обслуживать” разные признаки и в то же время один признак может быть конкретизирован с помощью разных шкал. Конкретизация признака осуществляется также с помощью комментариев, которые, однако, не являются самостоятельными объектами в системе технологической поддержки исследования, хотя над ними и производится ряд операций: ввод, форматирование, верстка, привязка к определенным признаку или шкале. Но комментарий, по сути, есть атрибут (причем не обязательный) либо признака, либо шкалы, либо анкетного вопроса в целом как метафорического отображения концепта.

В процессе осмысления первой версии системы мы с удивлением обнаружили, что формулировки вопросов анкеты имеют по меньшей мере две типичные “прописки” или два “воплощения”: во-первых, в качестве “полного имени вопроса”, а во-вторых, в качестве комментария “до вопроса” в ситуации, когда вопрос реализуется с помощью совокупности (блока, набора, множества) признаков. И в том, и в другом случаях он выглядел как текст достаточно большой длины (порядка 80 — 100 символов). Иными словами, сам вопрос оказывается *de facto* не самостоятельным объектом.

Итак, реально мы имеем всего два необходимых, значимых и тесно связанных друг с другом объекта: признак и шкалу. Именно вокруг этих объектов разворачивается вся “содержательная” деятельность исследователя.

Признак может быть охарактеризован с помощью следующих обязательных, или постоянных, атрибутов (свойств): *краткое имя* или “мнемоформула”, *полное имя*, описывающее его содержание (собственно “вопрос”),

тип признака (альтернативный, то есть допускающий выбор одного и только одного варианта ответа, или многоальтернативный: цифровой (код) или текстовый). В первой версии компьютерной технологии “мнемоформула” была определена как уникальное имя признака. Именно с помощью мнемоформулы осуществлялась идентификация признака при его поиске на всех стадиях работы.

Описание шкалы более развернуто. Во-первых, так как шкала представляет собой самостоятельный объект, она имеет собственное уникальное имя. Во-вторых, шкала непременно содержит список (перечень) пронумерованных (закодированных) вариантов ответов (значений или состояний). В-третьих, существуют шкалы нескольких типов: номинальная, ранговая и метрическая. К числу необязательных (или переменных) атрибутов шкалы следует отнести: *единицы измерения*, если шкала метрическая, то есть в нашей терминологии (табл. 2) — “предопределенная”.

В процессе эксплуатации первой версии системы возникла потребность выделить еще один объект, а именно — распределить все признаки по *рубрикам*. Эту необходимость мы ощутили, когда в базе накопилось порядка пяти-шести исследований, более 200 признаков и около 100-120 шкал и появилась проблема поиска нужных объектов. (Например, типична такая ситуация: “Точно знаю, что такой признак (или вопрос) в базе есть, но чтобы я помнил, как он называется”.)

Введение нового объекта потребовало серьезного реструктурирования основной базы — БД всех признаков. Рубрики были организованы как многоуровневая структура таким образом, чтобы исследователь имел возможность завести новую рубрику (любого уровня) в любой момент, и, кроме того, переопределить принадлежность признака к рубрике. Правда, один признак должен принадлежать одной и только одной рубрике. Это не сняло всех проблем, однако, с одной стороны, существенно облегчило поиск, а с другой, — еще раз заставило обратить внимание на многозначность смысла признаков и на принципиальную множественность их интерпретаций.

В системе изначально предполагалась поддержка процесса создания вторичных (или производных) признаков. С этой целью был разработан специализированный тезаурус (БД новых признаков) для накопления объектов данного класса. *Производный признак* отличается от исходного тем, что обладает некоторыми дополнительными специфическими атрибутами: *именами исходных* (первичных) признаков, и *алгоритмом построения* (вычисления), который хранится в тезаурусе и может быть воспроизведен при необходимости для любого исследования. В процессе реализации алгоритма производится реструктурирование (добавление новых полей) по меньшей мере двух таблиц: анкеты данного исследования и массива данных, а также собственно вычисление значений нового поля для каждой записи.

О взаимосвязи вопросов с теоретическими понятиями (концептами)

Каждый вопрос, с одной стороны, есть отражение определенной жизненной ситуации, фрагмента реальности, с другой — референт определенно-

го теоретического понятия (концепта). Достаточно сложное понятие вряд ли может быть описано с помощью какого-либо одного признака или даже вопроса. В программе исследования само явление может быть представлено (развернуто, описано) более детально с указанием связей и отношений между различными его аспектами (эмпирическими референтами). Понятие можно описать с помощью условной номинальной или ранговой шкалы (как правило, латентной в употребляемом выше смысле). Более того, программа может предусматривать и алгоритм обобщения (агрегирования) отдельных эмпирических референтов в единый показатель (индекс). Производные, латентные признаки, шкалы, обобщенные понятия могут быть “запланированы” и подготовлены к реализации уже на подготовительных стадиях исследования.

С одной стороны, такого рода действия следует квалифицировать как конструирование реальности, навязывание последней не свойственных ей моделей. С другой стороны, исследователь имеет право на теоретическое (модельное) представление действительности. И сама эта деятельность должна быть предъявлена научному сообществу, как для критики, так и для воспроизведения в сходных или иных условиях. И, наконец, такую деятельность необходимо технологически поддерживать и фиксировать (документировать) должным образом.

Общая композиция системы компьютерной поддержки исследования

Есть еще один объект, который до сих пор подразумевался “по умолчанию”, – это собственно *исследование*. В рамках конкретного исследования (исследовательского цикла) прописаны и анкета, и массив эмпирических данных, и различные процедуры обработки и анализа этих данных, и такой документ как отчет. Система компьютерной поддержки предназначается для обеспечения *всех* этапов исследовательского цикла, однако, пока мы говорили только о ключевых его объектах, ибо именно они, с одной стороны, представлялись самоочевидными, с другой, скрывали в себе наибольшие “напряжения”.

В центре схемы – “База всех исследований” и форма “Создание анкеты”. Эти два структурных элемента моделируют начальные этапы исследования (разработка анкеты, подготовка к проведению опроса, ввод данных и т.п.) и обеспечиваются “инфраструктурой” системы. За обработку эмпирических данных, их агрегирование и обобщение “отвечают” специализированные формы: “Задания на обработку”, включающие формирование и запуск собственно алгоритмов обработки, и “Каталог результатов”, позволяющий накапливать и соответствующим образом готовить результаты обработки данных к включению в отчет.

Элемент системы, имеющий сугубо вспомогательный характер обозначен в главном меню нашей системы кнопкой “Сборка”. Этот элемент обеспечивает преобразование стандартных (или “штатных”) форм представления результатов Paradox в вид, более привычный для социологических отчетов, а

именно: расчет процентов, статистические оценки распределений, оформление таблиц приемлемое для нормального “восприятия” стандартными текстовыми редакторами и иные операции. Кроме этого, здесь же осуществляется и собственно “сборка” результатов, полученных в разных исследованиях, в один тематический “пул” для последующего сравнительного анализа.

Каталог результатов — не просто хранилище. Он оснащен процедурами, обеспечивающими перевод (конвертирование) результатов счета в текстовый формат, а также позволяющими использовать “штатные” возможности СУБД для графического представления результатов счета и обработки данных.

“Задание” в первой версии системы жестко связано с “текущим исследованием”. “Каталог” в этом смысле автономен и позволяет обращаться ко всем результатам, имеющимся в базе данных и (по первоначальному замыслу) — строить агрегированные и “сводные” таблицы с привлечением результатов разных исследований.

В описываемой версии пополнение основных БД (баз данных) могло осуществляться двумя способами. Во-первых, пользователь мог вводить новые объекты непосредственно в соответствующий тезаурус, во-вторых, он имел возможность создавать новые признаки и шкалы “на строительной площадке” при формировании текущей анкеты нового исследования. Одни признаки и шкалы он копировал из БД, а другие сочинял заново. При закрытии формы “Создание новой анкеты” система автоматически пополняла соответствующие базы данных, занося в них вновь созданные объекты: признаки, шкалы, рубрики. Кроме этого, в системе фиксировалось, в каком именно исследовании впервые появились те или иные объекты. Это было достаточно удобно, хотя и порождало определенные трудности и сбои в работе системы. В новой версии скорее всего пользователю будет предоставлен лишь один способ — прямого пополнения соответствующих БД с тем, чтобы создание новых полевых документов (анкет, бланков интервью и т.п.) осуществлялось исключительно на основе уже имеющихся в БД объектов.

И, наконец, следует добавить, что в первой версии процессы формирования задания и собственно обработки данных были построены по образцу SPSS и довольно удачно, на наш взгляд, (с большим быстродействием) имитировали “пакетный режим”.

На рис. 1 отображены общие принципы организации объектов (баз данных или таблиц Paradox) системы информационно-технологической поддержки исследований. На “верхнем” уровне располагаются базы “общего характера”, которые могут быть использованы практически в любом исследовании. Для каждого конкретного опроса из общей базы делается соответствующая выборка или редакция. На этом же уровне находятся “адреса” имеющихся в общей базе исследований. Выбрав в каталоге нужное исследование, пользователь “автоматически” открывает именно его вне зависимости от предстоящих операций.

Второй уровень содержит директории конкретных исследований, каждая из которых построена по единой схеме и включает ряд таблиц: структуру анкеты, массив эмпирических данных, характеристики выборки, документа-

цию об интервьюерах (включая ведомость оплаты труда) и, наконец, результаты статистической (или, шире, машинной) обработки данных, необходимых для написания отчета.

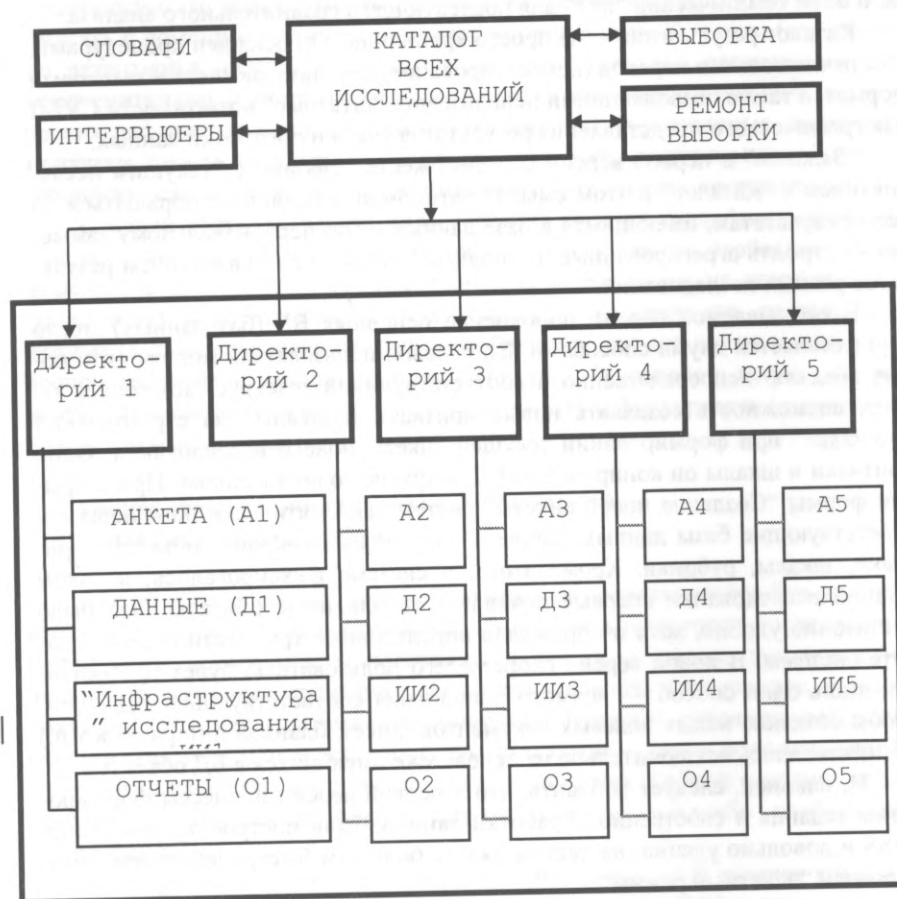


Рис. 1. Общая схема организации объектов (баз данных) информационно-технологической поддержки социологических анкетных исследований

Понятием “инфраструктура исследования” мы обозначаем такие документы, которые непременно сопровождают эмпирическое исследование и отнимают у исследователей немало времени и сил: кодировальные листы, задания интервьюерам по выборке, в случае телефонных опросов – номера АТС, списки интервьюеров, занятых в исследовании, инструкции для них. Кроме перечисленного, сюда мы относим и специфические процедуры, связанные с контролем качества эмпирической информации и ремонтом выборки программными средствами. В первой версии системы реализованы практически все названные элементы инфраструктуры, кроме инструкции для

интервьюеров, последняя создается независимо в текстовом редакторе Word for Window's. Все документы, так или иначе связанные с выборкой, помимо программных средств реализации, требуют весьма солидного информационного обеспечения, в том числе автономных баз данных (например, списков интервьюеров, списков номеров АТС городской телефонной сети с их привязкой к административным районам города и т.п.), которые были созданы нами ранее и затем интегрированы в систему.

Отдельными элементами системы стали средства (формы) организации контроля за вводом исходной исследовательской информации и ремонта выборки. И хотя, как правило, мы тщательно контролируем параметры выборки и в ремонте выборки нужды обычно не возникает, этот блок системы оказывается необходимым, особенно когда возникают задачи прогнозного характера.

Промежуточные итоги

В настоящее время ведется работа по созданию новой версии системы. Начав с попытки хоть немного избавиться от рутинных и трудоемких операций, мы были вынуждены заняться анализом собственной профессиональной деятельности. В процессе этого анализа сформировалось представление о системе и ее существенных принципах. В качестве таковых мы выделили следующие: (1) ориентация на анкету, то есть на полную автоматизацию всех "внутренних", служебных структур — форм ввода, контроля данных и т.п.; (2) полная каталогизация основных элементов исследования: прежде всего, самих исследований, признаков и шкал, заданий на обработку, результатов обработки (одномерных, двумерных распределений, условных средних и др.), новых, то есть производных или вторичных признаков, характеристик интервьюеров, генеральной и выборочных совокупностей; (3) расширение возможностей ассистирования при построении новых признаков (автоматическая подстановка имен исходных признаков из соответствующих БД, формирование заготовки алгоритма их построения с развернутыми комментариями, каталогизация алгоритмов, автоматическое реструктурирование базы исходных данных, вычисление значений новых признаков); (4) полная интегрированность в среду Paradox.

Следует обратить внимание на такой элемент системы, как словари. Это действительно словари, которые позволяют учитывать синонимичность естественного языка при обработке и анализе ответов на открытые вопросы, не занимаясь при этом перекодировкой исходной информации. Последняя полностью сохраняется, тем самым сохраняется возможность повторного обращения к ней и анализа данных под любым иным углом зрения.

Исследования неструктурированного типа, скажем, биографические или основанные на наблюдении или глубинном интервью, также нуждаются в специализированном и глубоком самостоятельном анализе. По всей вероятности, в рамках этих исследований будут выделены другие объекты поддержки, скорее всего иной будет и технология поддержки. Вероятно, когда-нибудь их удастся интегрировать в единое целое, но это задача будущего.

ЛИТЕРАТУРА

1. Заславская Т.И., Рывкина Р.В. О принципах и основных элементах программирования исследования трудовой мобильности // Методологические проблемы социологического исследования мобильности трудовых ресурсов. Новосибирск: Наука, 1974.
2. Территориальная выборка в социологических исследованиях. М.: Наука, 1980. С. 218.
3. Социология: Основы общей теории. М.: Аспект Пресс, 1996.
4. Чесноков С.В. Детерминационный анализ социально-экономических данных. М.: Наука, 1982.
5. Чесноков С.В. Силлогизмы в детерминационном анализе // Известия АН СССР. Сер. "Техническая кибернетика". 1984. № 5.
6. Каныгин Г.В. Предметная область при компьютерном моделировании анкетного опроса // Социологический журнал. 1997. № 1/2. С. 93-119.
7. Божков О.Б., Лезин Г.В. Технологическая поддержка эмпирических социологических исследований // Информационные технологии в гуманитарных и общественных науках. СПб: ЭМИ РАН, 1994. С. 56-78.