



СОЦИОЛОГИЧЕСКАЯ  
ИНФОРМАЦИЯ В

**ЭВМ**

АКАДЕМИЯ НАУК СССР  
ИНСТИТУТ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОБЛЕМ

# **СОЦИОЛОГИЧЕСКАЯ ИНФОРМАЦИЯ В ЭВМ**

(СБОР, ХРАНЕНИЕ, КОНТРОЛЬ,  
РЕДАКТИРОВАНИЕ)

ЛЕНИНГРАД  
«НАУКА»  
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ  
1983

Социологическая информация в ЭВМ (сбор, хранение, контроль, редактирование). Иголкина Д. В., Каневский Е. А., Клокачев И. В. и др. — Л.: Наука, 1983. — 236 с.

В монографии рассматриваются вопросы, связанные с организацией хранения социологической информации, структурой информации, контролем и редактированием данных, а также общением пользователей с архивом в диалоговом режиме. Большое внимание уделяется техническим средствам коллективного пользования вычислительной системой. Подробно описываются терминальные средства сбора информации, режимы диалога с пользователем и связи с ЭВМ. Особое внимание уделяется автономным устройствам, обеспечивающим интерактивный режим сбора структурированной информации. Изложенный в книге материал базируется на опыте разработки и эксплуатации системы комплексной обработки социально-экономических данных в ИСЭП АН СССР и на обобщении известных практических и теоретических результатов. Лит. — 183 назв., ил. — 37, табл. — 35.

Авторы: Д. В. ИГОЛКИНА, Е. А. КАНЕВСКИЙ, И. В. КЛОКАЧЕВ,  
В. Б. МАРАХОВСКИЙ, Г. И. МЕНДЕРСКАЯ, Ж. И. ПАНЧУК

Рецензенты: В. А. Ядов, С. И. Баранов

Ответственный редактор  
доктор технических наук В. И. Варшавский

## ПРЕДИСЛОВИЕ

---

Роль социологических исследований как источника информации, необходимой для решения широкого круга практических вопросов, непрерывно возрастает. И вместе с тем возрастает значение автоматизации информационной деятельности на базе современных ЭВМ как одного из важнейших средств повышения эффективности социологических исследований. Все это крайне важно для современных методов научного управления и отвечает требованиям углубления хозяйственной реформы в совершенствовании социального механизма в свете постановлений XXV и XXVI съездов КПСС.

Применение ЭВМ в общественных науках началось с решения экономических задач, имеющих преимущественно вычислительный характер [81]. По мере развития вычислительной техники и накопления опыта уже в 60-х годах ЭВМ стали широко использовать демографы, социологи, историки и др. При этом наряду с расширением области применения ЭВМ изменился и характер решаемых задач: вычисления стали заменяться анализом текстовой информации, изучением структуры языка и т. п. Стали проводиться исследования по унификации и стандартизации не только исходной информации, но и методов ее описания и обработки.

В социологических исследованиях используются материалы массовых опросов, которые отличаются разнообразием внутренней структуры и большим числом наблюдений [80]. Здесь исследователь сталкивается со сложными информационными процессами, анализ и изучение которых сопряжены со значительными затратами труда и времени. ЭВМ позволяет учесть большое число факторов и связей, добиться лучшей координации, а также снизить стоимость социологического исследования и повысить его эффективность. Особенно большой выигрыш от применения ЭВМ может быть получен на стадии обработки социологических данных, которая занимает сейчас до 80% времени, затрачиваемого на весь цикл социологического исследования.

Одним из самых трудоемких этапов информационной деятельности является получение, накопление и преобразование данных [80]. Это связано с тем, что до сих пор большинство операций здесь выполняется вручную. Ведь даже физики, которые далеко продвинулись в области автоматизации эксперимента, отмечают, что во многих случаях задержка в получении окончательных результатов исчисляется годами. И они считают, что одна из задач, стоящих перед системой автоматизации, — сокращение сроков

от момента, когда экспериментальные данные уже получены, до момента получения окончательных результатов. Что же в этом случае могут сказать социологи, не имеющие автоматизированных систем сбора данных? Потребность в технических средствах сбора информации для них особенно велика. Основной задачей является придание социальной информации формы, пригодной для непосредственного ввода в ЭВМ, а также специальная организация и преобразование этих данных к виду, удобному для последующей обработки.

Одной из форм такой организации является архив форматных анкет (АРФА), создание которого в 1977 г. потребовало нового подхода к классификации и кодированию информации, к разработке более усовершенствованных средств формализованного описания данных. Здесь решающая роль принадлежала исследованию информации в семантическом аспекте и разработке на этой основе содержательных информационных языков, а также переходу к автоматизированному кодированию данных и организации диалога человека с ЭВМ.

Следует отметить, что эффективность применения ЭВМ в настоящее время все более определяется средствами диалога между человеком и вычислительной машиной. По оценке специалистов трудовые затраты только на преодоление информационного барьера между пользователем и ЭВМ в лучшем случае составляют около 50% от общих затрат на вычислительную технику. А это означает, что развитие и широкое внедрение средств диалога вдвое повысит эффективность использования ЭВМ.

В связи с возрастающим объемом данных и ограниченностью числа специалистов, владеющих программированием, встает вопрос о создании средств программирования для обработки информации, которые ориентированы на пользователей без опыта работы на ЭВМ — так называемых конечных пользователей. И диалоговый язык должен быть рассчитан именно на них, а не на системных аналитиков — людей, постоянно работающих с ЭВМ.

Для того чтобы необученный пользователь мог взаимодействовать с системой, необходимо, чтобы язык запросов не требовал каких-либо специальных знаний, легко усваивался и запоминался, т. е. был близок к естественному. Это значительно облегчает широкое использование режима диалога, имеющего по крайней мере два преимущества: 1) получение данных непосредственно на рабочем месте и в виде, удобном для анализа и принятия решения; 2) возможность создания систем, не использующих бумажные документы, что качественно изменит технологию обработки данных.

Сбор и подготовка данных могут осуществляться децентрализованным способом, для чего в местах возникновения первичной информации оборудуются абонентские пункты, позволяющие работать с ЭВМ в режимах прямой (on line) или косвенной (off line) дистанционной обработки.

В режиме прямой дистанционной обработки ЭВМ непосредственно связана с терминалом по каналу связи, что обеспечивает

немедленный ввод информации в ЭВМ и передачу результатов абоненту. В настоящее время все более важную роль начинают играть активные или интеллектуальные терминалы, имеющие целый ряд преимуществ перед пассивными (дисплеями). Первым отечественным активным терминалом явился регистратор информации РИ-7801, разработанный в 1976 г., в котором в качестве внешнего запоминающего устройства использован накопитель на мини-кассетах.

В режиме косвенной дистанционной обработки информация передается в ЭВМ с помощью промежуточных носителей, роль которых могут выполнять перфоленты, перфокарты, магнитные ленты, магнитные диски и т. д. К подобному типу относятся различного рода устройства подготовки данных и первый отечественный электронный интервьюер — регистратор информации социально-экономического профиля (РИСЭП). В качестве промежуточного носителя в последнем используются мини-кассеты емкостью около 70 тыс. символов.

Применение подобных средств позволяет использовать ЭВМ на этапах сбора, классификации и предварительной обработки социологической информации [90], в частности на этапах анкетирования и интервьюирования. Применение электронного интервьюирования резко уменьшает влияние интервьюера на респондента, а также обратное влияние, объясняемое их личным контактом, улучшает условия анонимности; создает возможность увеличения точности и непротиворечивости, а также исключает случайные пропуски ответов; создает возможность для более целенаправленного ведения опроса за счет исключения нарушений порядка ответов на вопросы; обеспечивает автоматическую фиксацию ответов респондента в памяти ЭВМ или на машинных носителях.

Затрудняющими факторами применения ЭВМ для сбора социологической информации являются территориальная рассредоточенность и подвижность респондентов.

В предлагаемой читателю монографии сделана попытка обобщить отечественный и зарубежный опыт в области сбора, хранения, контроля и редактирования социологической информации. В книге нашли отражение исследования, выполненные авторами в Ленинградском отделении Центрального экономико-математического института АН СССР и Институте социально-экономических проблем АН СССР при разработке архива форматных анкет и создании терминальных устройств РИ-7801 и РИСЭП. Следует отметить, что архив нашел особенно широкое применение: за период его существования архивом пользовалась 21 организация, было введено и обработано более 400 пакетов общим числом около 200 тыс. анкет.

В работе над архивом принимали участие И. П. Макаренко, В. П. Макарычев, Л. И. Пархоменко, П. Ю. Суворов и Э. В. Штенникова, в создании терминальных устройств — Н. Н. Вертенова, Л. М. Гайдукова и Е. Н. Клименко.

## ОСНОВНЫЕ ПРИНЦИПЫ ОРГАНИЗАЦИИ АРХИВА ФОРМАТНЫХ АНКЕТ (АРФА)

### 1.1. Архив как составная часть системы обработки данных. Общие требования

Архив, т. е. система хранения данных, проектируется и реализуется как составная часть проблемно-ориентированной системы программирования. Архив, о котором идет речь здесь, имеет проблемную ориентацию на использование в системах анализа и обработки социологической информации — анкет. Это находит отражение в определении структуры данных, организации систем ввода и хранения, в процедурах доступа. При решении различных вопросов, связанных с повышением эффективности, удобства, простоты, а также с количественными ограничениями, в качестве типовой структуры данных рассматривалась именно социологическая анкета. Это означает, как будет видно в последующих главах, что по своей организации, структуре данных и набору программ-операций архив фактически представляет собой достаточно универсальную базу данных. Однако естественность и удобство языков описания данных и манипулирования ими, эффективность программного обеспечения системы АРФА достигаются при работе с социологическими анкетами.

В системе обработки данных архив занимает центральное место в том смысле, что интерфейс между различными компонентами системы определяется (и осуществляется) через архив. Иначе говоря, лишь структура данных является общим требованием и ограничением для программных модулей. Организация системы из формально независимых модулей является естественной и принципиальной при удовлетворении ряда требований. Перечислим важнейшие из них.

1. Возможность развития системы (пополнение новыми компонентами) и модификации (усовершенствование отдельных компонент) в процессе эксплуатации.

2. Использование различных систем (языков) программирования для реализации различных программных модулей.

3. Простота технологии построения системы и возможность разработки модулей системы разными лицами в различные сроки.

4. Эффективность использования современных средств вычислительной техники при развитии вычислительной базы путем пополнения компонент системы.

При разработке архива анкет авторы руководствовались рядом основных принципов, изложенных в следующих разделах настоящей главы. Некоторые из них носят общий характер:

- 1) работоспособность и надежность системы — основные показатели;
- 2) простота и естественность средств общения человека с системой;
- 3) универсальность системы хранения; но не в ущерб эффективности ее и удобству работы с ней;
- 4) максимальный сервис как для пользования, так и для системного обслуживания.

## 1.2. Организация системы хранения данных

Собственно архив как хранилище информации создается на устройствах внешней памяти ЭВМ: магнитных лентах (МЛ) и магнитных дисках (МД). Две основные задачи встают при разработке способов хранения информации во внешней памяти: эффективность использования памяти (плотность записи информации) и скорость поиска и считывания нужных массивов информации.

Мы исходим из того, что устройства внешней памяти имеют зонную структуру: операции ЭВМ (или экстракоды операционной системы — ОС) выполняют позонное обращение к МЛ или МД. Поэтому организация хранения массивов информации (в АРФА — пакетов анкет) естественно также принята позонной. Поскольку МЛ и МД имеют различные размеры, конкретные физические ленты или диски могут иметь различную фактическую длину (испорченные или занятые чем-то зоны), то каждый физический носитель архива, называемый томом, снабжается каталогом собственных зон. В каталоге для каждой зоны отмечается, занята ли зона и чем (имя массива). Тогда оказывается, что в системе управления архивом может вообще отсутствовать список (или каталог) пакетов и томов. Система работает каждый раз с заданным набором томов.

При заполнении архива новым массивом информации очередность зон, занимаемых под данный массив, не фиксируется, т. е. считается, что порядок безразличен. В этом выражается одна из специфик архива анкет, хранимое и обрабатываемое множество которых рассматривается неупорядоченным. Но зато упрощается и ускоряется процесс поиска и считывания зон (достаточна шкала, а не список зон, см. гл. 4). Поскольку упорядоченность зон в архиве не имеет значения, то в процессе работы с томами дополнительных переписей зон при записи новой зоны не требуется. Заметим также, что такая организация обеспечивает единство системы хранения на устройствах последовательного — МЛ — и произвольного — МД — доступа.

Эффективность использования памяти существенным образом обеспечивается уплотненной формой записи информации в архиве. При разработке форм представления информации в архиве



(см. гл. 2) учтена специфика массивов социологических анкет. Описание формата анкеты «обслуживает» множество анкет. Из всех типов элементов, составляющих анкеты, самая компактная и удобная для быстрой выборки и распаковки форма представления выбрана для наиболее распространенных типов: целочисленных значений и номеров подсказок. В итоге, в среднем упакованная информация занимает в 8—10 раз меньше места во внешней памяти по сравнению со стандартной формой представления значений.

На первый взгляд может показаться, что требование экономии внешней памяти путем уплотненной записи информации в архиве и требование быстрого считывания массивов информации в процессе обработки противоречивы. Ведь в случае упакованной формы при считывании порции информации еще нужно затратить время на ее распаковку. Однако «накладные расходы» при обращении к внешней памяти оказываются настолько большими, что для архива, реализованного в ОС ДИСПАК на БЭСМ-6, процессорное время, которое затрачивается на считывание одной зоны, в несколько раз больше времени, требуемого для распаковки зоны. Специальные эксперименты показали, что для пакетов из 500—800 анкет типичного формата (50—100 признаков) процессорное время считывания распакованных анкет в 5—6 раз больше, чем время считывания упакованных анкет с последующей распаковкой по формату.

### 1.3. Структура данных и форма их представления

При создании базы данных должны быть определены объекты, подлежащие хранению и обработке. При задании объекта в системе прежде всего нужно определить его структуру. От структуры объекта зависит в значительной мере «лицо» всей системы: ее назначение, возможности развития, степень универсальности, простота и удобство пользования, эффективность реализации. Хотя понятие типа объекта определяется не только его структурой, но и допустимыми операциями над ним [121], однако в этом разделе мы рассматриваем лишь структуру.

Имея в виду главное назначение описываемой базы данных как архива социологических анкет, основной объект определяется как анкета. Структура анкеты была выработана в процессе анализа нескольких десятков практических анкет, применяемых различными группами исследователей [21, 22, 27, 39, 40, 42, 75, 79, 81, 86, 113].

Анкета, т. е. объект разрабатываемой системы программирования, определяется как структура в языках программирования [4, 115]. Она представляет собой упорядоченную последовательность элементов (полей структуры) разных типов. В качестве допустимых вводятся простые типы (целые, вещественные, текстовые) и сложные, или составные (массивы, последовательности, анкеты). Массивы и последовательности образуются из элементов одного типа. Описание структуры анкеты, называемое описанием фор-

мата, содержит перечисление типов составляющих элементов, в случае составных — типов элементов, их составляющих.

В обработке участвует пакет однотишных анкет, т. е. анкет с одним форматом. Поэтому при организации архива предусмотрено задание описания формата анкет для пакета, которое вводится в архив при образовании пакета. Представление анкет в памяти ЭВМ жестко связано с заданным форматом: по формату анкета записывается и читается.

Среди элементов, составляющих анкету, могут быть такие, размеры значений которых не фиксируются форматом, т. е. не ограничены. Из элементов простых типов это — текстовые, составных — последовательности. Поэтому в представлении анкеты в памяти целесообразно выделить форматную и неформатную части. В случае значений элементов, выходящих за границы фиксированного поля в форматной части, заводится ссылка, указывающая на начало соответствующего участка неформатной части. Конец участка определяется длиной, задаваемой для каждого участка в анкете (а не в формате). Подробно это описано в следующей главе.

Таким образом, структура анкеты оказывается определенной достаточно просто, но в то же время и достаточно универсально. В принципе структура анкеты допускает задание любых иерархических структур объектов, состоящих из элементов разного типа. Но если для социологических анкет задание формата оказывается естественным, то для других структур оно может оказаться неестественным или громоздким.

Мы определили (пока неформально) структуру данных в отрыве от физического представления. В информационных системах, характеризующихся сложностью и большими объемами данных, с одной стороны, и разнообразными видами работ с этими данными, с другой стороны, отделение понятия типа объекта от конкретных форм представления объекта существенно. Такой подход позволяет обеспечить эффективность системы на различных этапах манипулирования с данными, делает возможным использование одной и той же информации различными программами обработки, составленными в разных системах программирования на разных алгоритмических языках. В частности, допустимо подключение к системе программ ввода и первичной обработки данных, использующих различные физические устройства ввода и терминальные устройства, предполагающие различные формы кодирования информации на внешних носителях и различные формы представления информации в ЭВМ.

Специальные программы обеспечивают интерфейс (связь, стыковку) различных модулей системы, настроенных на разные конкретные формы представления данных. Эти программы (будем называть их программами интерфейса) осуществляют соответствующие преобразования значений объектов из одной формы представления в другую. По способу использования программ интерфейса можно различать статический и динамический интерфейс. При статическом интерфейсе преобразование информации производится

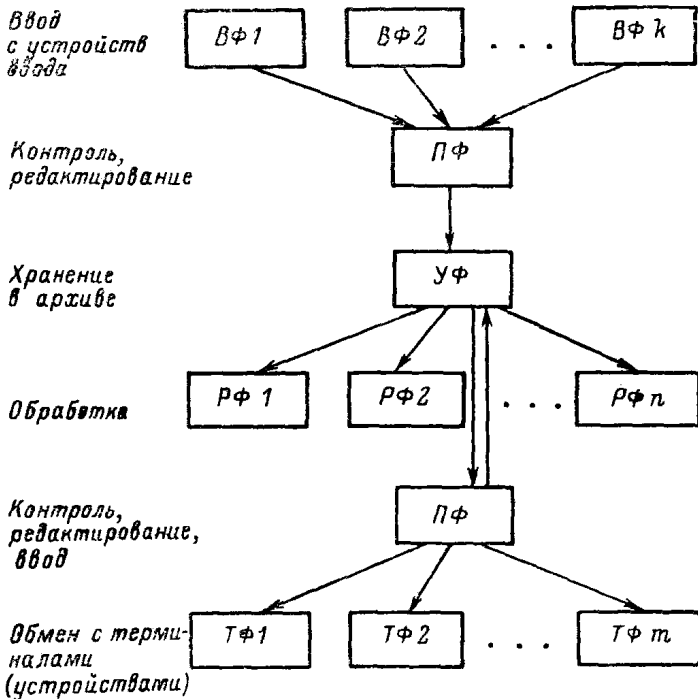


Рис. 1.1. Формы представления данных в системе АРФА.

$ВФ1, ВФ2, \dots, ВФk$  — различные формы представления при вводе;  $ПФ$  — промежуточная форма;  $УФ$  — форма представления в упакованном виде;  $РФ1, РФ2, \dots, РФn$  — формы представления в распакованном виде для разных систем программирования;  $ТФ1, ТФ2, \dots, ТФm$  — формы представления данных для обмена с терминалами (устройствами) разного вида.

после завершения одного этапа обработки информации до начала следующего этапа. В этом случае преобразование информации является как бы дополнительным этапом в цепи последовательных действий. Динамический интерфейс характеризуется тем, что информация преобразуется из одной формы представления данных в другую в процессе обработки. Программы динамического интерфейса оказываются встроенными в программы основных операций и участвуют в одном непрерывном (конвейерном) процессе обработки информации.

В описываемой системе АРФА (рис. 1.1) предусмотрена возможность нескольких форм представления данных на этапе ввода с наращиванием набора этих форм. После непосредственного ввода информация преобразуется во внутреннюю промежуточную форму, с которой работают программы синтаксического контроля, редактирования и записи в архив. В архив информация записывается в уплотненной форме, предусмотренной логической организацией архива. При выборке из архива для обработки информация преобразуется в форму представления данных, соответствующую системе программирования, в которой составлены программы обра-

ботки. Эти преобразования (см. главу 4) осуществляются специальными процедурами в соответствующей системе программирования. Для каждой системы программирования формой представления является некоторая адекватная модель структуры анкеты в этой системе программирования.

#### 1.4. Описание данных, ввод, контроль, редактирование

Разработке системы архива социологических анкет предшествовал период изучения этапов работы исследователя, начиная со сбора информации. Это изучение показало, что этапы сбора, кодирования, перфорации, ввода в ЭВМ, предшествующие проведению исследований, весьма трудоемки, выполняются людьми разной квалификации. Перед разработчиками системы архива встал вопрос о создании такой автоматизированной системы, которая в возможно большей мере облегчила бы процесс подготовки, ввода, контроля и редактирования анкет.

С этой целью прежде всего был разработан язык ввода анкет (иногда в тексте этот язык называется стандартным в тех случаях, когда рассматриваются некоторые отклонения от него). Язык ввода разработан с учетом не только структуры анкеты, но и ряда реальных обстоятельств, специфичных для анкет или для больших массивов информации:

- допустимость отсутствия значений для элементов анкет или компонент, составляющих элементы;
- возможность повторения значений для нескольких подряд идущих элементов;

- необходимость исправления замеченной ошибки без возвращения назад (это замечание существенно при работе непосредственно на терминале, при перфорации на перфоленду и т. д.).

Разработанный язык соответствует структуре и содержанию анкет, благодаря чему могут быть опущены такие традиционные трудоемкие операции, как кодирование и перфорация, ввод может осуществляться непосредственно с анкет на терминалах ЭВМ. Однако в ряде случаев перфорация может оказаться более целесообразным и дешевым способом подготовки данных. Язык ввода допускает модификации, соответствующие конкретным устройствам подготовки данных или частным случаям анкет.

Язык описания данных, т. е. описание формата, синтаксически совпадает с языком ввода, лишь вместо значений элементов вводятся коды типов этих элементов.

На всех предварительных этапах работы с анкетой, начиная от заполнения ее респондентом, возможны самые разнообразные ошибки и некорректности. Поэтому необходима разработка и создание эффективного аппарата контроля и коррекции анкет. В системе АРФА разработаны специальные программы контроля анкет по задаваемым критериям корректности. Проверке подвергается пакет анкет или заданная часть его. Анкеты, в которых

обнаруживаются ошибки, распечатываются с указанием ошибочных (или подозреваемых) элементов.

Основной режим редактирования в системе АРФА — диалоговый с использованием терминалов с экранами. Язык редактирования практически совпадает с языком ввода. Собственно говоря, это одна система. Разница заключается лишь в том, что в режиме редактирования предварительно вызывается нужная анкета. Кроме того, несколько дополнительных операций имеют смысл именно при редактировании: сдвиг части анкеты, замена элементов и др. (см. главу 3). Синтаксическое совмещение языков ввода и редактирования сделано умышленно не только для удобства реализации. Тем самым для пользователя не специалиста в области вычислительной техники и программирования оказывается определенным только один, достаточно простой, экономичный и универсальный язык. Даже человек, впервые имеющий дело с вычислительной техникой, овладевает всеми «премудростями» языка в течение одного часа.

### 1.5. Доступ к архиву, преобразование массивов информации

Программы, осуществляющие анализ и обработку информации, хранящейся в архиве АРФА, составлены и составляются в рамках систем программирования, имеющихся на ЭВМ. Для использования архива в программах обработки служит набор процедур (подпрограмм) в соответствующей системе программирования, реализующих необходимые операции доступа к архиву. Одно из основных назначений этих процедур — обеспечение эффективного интерфейса между представлением данных в архиве и в системе программирования.

Важным моментом является выбор наиболее рационального представления анкеты в системе программирования. В данном случае под «рациональным» мы имеем в виду такое представление, которое дает возможность составлять эффективные программы обработки и при этом является достаточно естественным и простым для восприятия программистом, составляющим программы обработки, и пользователем этих программ. В системах программирования, используемых на БЭСМ-6, в соответствии со структурой и представлением данных в них, принята форма представления структуры социологической анкеты, описанная в главе 4. Форма представления данных фиксирована в системе, менять ее нельзя, ибо это повлечет за собой необходимость изменения многих программ. Помимо формы представления самой анкеты, должны быть фиксированы также структура и смысл сопутствующей информации, описывающей формат, каталог тома, рабочие поля и разную служебную информацию.

Программы, реализующие операции доступа к архиву, опираются на единую форму представления данных и описания их как в архиве, так и в системе программирования. Набор этих опе-

раций может пополняться, видоизменяться в процессе эксплуатации системы. В процессе разработки и эксплуатации системы АРФА создано два набора операций доступа к архиву.

Первый набор операций ориентирован на анализ и обработку анкет без их преобразования в процессе обработки, поэтому их называют операциями выборки анкет. Они реализованы в виде процедур системы БЭСМ-АЛГОЛ [105] и употребляются в программах на языке АЛГОЛ-60. Для достижения большей скорости работы процедур в них широко применяется автокод, и они созданы в виде процедур, работающих в два этапа. В программе обработки социологической информации используются пакеты достаточно большого числа анкет одного формата. Учитывая это обстоятельство, все общие необходимые действия, связанные с выборкой и подготовкой служебной информации — о физических носителях, формате анкет, рабочем поле, — выделены в отдельную процедуру. Эти действия осуществляет процедура *format*, обращение к которой должно быть выполнено в начале программы. Выборку очередной анкеты осуществляет процедура *ankf*, которая в качестве результата на заданном (в виде параметра) массиве выдает анкету в распакованном виде, т. е. в форме, предусмотренной для системы программирования. Таким образом, операции первого набора обеспечивают только выборку (чтение) из архива анкет только одного формата.

Второй набор операций доступа к архиву реализован в рамках Мониторной системы ДУБНА [66]. Они предоставляют более широкие возможности, допуская работу с пакетами анкет разных форматов и преобразование пакетов в архиве.

Для реализации возможности работы с анкетами разных форматов вводится понятие канала. Канал — это часть архива, состоящая из одного или нескольких пакетов анкет с одинаковым форматом. Каналов может быть объявлено (открыто) в программе несколько, каждому задается свой номер. Операция ОТКР (открыть канал) аналогична по своим действиям операции *format*. Для каждого канала отводятся рабочее поле текущей распакованной анкеты, буфер каталога, поле формата и ряд служебных регистров. Все операции работы с анкетами в качестве одного из параметров должны иметь номер канала.

Основными операциями являются: чтение следующей (очередной) анкеты; запись из рабочего поля на место текущей анкеты; то же — на свободное место (в пакете, соответствующем текущей анкете); исключение текущей анкеты.

Для удобства с каждым каналом связана стрелка, указывающая на текущую анкету. Введены операции установки и сдвига стрелки. Для каждой операции определено положение стрелки после выполнения операции. Для операции чтения предусмотрена вторая стрелка, специальная, указывающая на следующую анкету (подробности см. в главе 4).

Каждая анкета имеет имя, заданное или приписанное системой (см. главу 2). Операции доступа имеют варианты работы с анке-

тами по имени. Предусмотрены операции выдачи распакованных анкет на печать.

Преобразование распакованной анкеты на рабочем поле выходит за рамки «компетенции» системы АРФА. Предполагается, что все необходимые операции осуществляются в программе обработки средствами соответствующего языка системы программирования, поскольку представление анкеты на рабочем поле соответствует представлению данных в этой системе.

Операции доступа к архиву допускают, как нетрудно видеть, преобразование информации, хранящейся в архиве. С их помощью можно произвольным образом корректировать пакет анкет. Пользуясь двумя каналами одновременно, можно запись преобразованных анкет производить по другому каналу, образуя таким образом частные (возможно, временные) подмассивы с сокращенными анкетами, что повышает производительность программ обработки. В главе 4 приведены примеры преобразований анкет с разными форматами, выбираемых по двум каналам. При этом образуемая третья анкета с новым форматом засылается по третьему каналу. Преобразования такого рода позволяют создавать пакеты анкет наиболее оптимальной структуры для последующей эффективной обработки.

#### 1.6. Системное обслуживание архива

Большой объем информации, хранение которой предполагается долговременным (несколько лет), большое число пользователей системы разной квалификации, широкое использование человека в процессе ввода и предварительной обработки анкет — вот основные обстоятельства, определяющие необходимость организации специального системного обслуживания архива. Система АРФА с самого начала проектировалась и разрабатывалась с учетом наличия в ней аппарата системного обслуживания, выполняемого специальной ответственной группой.

В системе предусмотрены программы, реализующие специальные служебные операции, пользование которыми является компетенцией группы системного обслуживания. К их числу относятся операции заведения тома архива с организацией каталога и контролем состояния физического носителя, заведения нового пакета и нового формата, контроля, просмотра и распечатки всей служебной информации, находящейся на томе, операции переписи пакетов анкет с одного тома на другой с переупорядочиванием пакетов. Системное обслуживание опирается на общесистемный аппарат контроля и обслуживания, имеющийся в операционной системе: защита информации от случайных разрушений путем введения паролей для физических носителей, служебные программы проверки, переписи и дублирования информации, хранимой в системе долго, и т. д.

В функции группы системного обслуживания входит, как необходимый элемент, оказание методологической помощи пользо-

вателям системы на разных этапах ее применения, а в некоторых случаях и организация работ на отдельных наиболее трудоемких этапах. Опыт эксплуатации архива в течение нескольких лет показал, что работоспособность, надежность и эффективность использования системы во многом определяются четкой организацией системного обслуживания.

Как уже было сказано, АРФА — открытая система, допускающая свое развитие в процессе эксплуатации. Выявление необходимости модернизации системы, включения новых возможностей, усовершенствования аппарата системного обслуживания, а затем реализации этих потребностей также являются одной из важнейших задач группы системного обслуживания.



---

## ПРЕДСТАВЛЕНИЕ ДАННЫХ В АРХИВЕ И ВВОД ДАННЫХ

### 2.1. Языки описания и ввода анкет

В рассматриваемом ниже типичном языке описания анкеты учтены все основные требования к языкам описания данных подобного типа [21, 39, 75, 79, 86].

Язык должен: определять типы элементов; задавать правила построения сложных элементов; допускать идентификацию каждого элемента.

Основная цель языка — дать постановщику задачи (социологу) возможность описать в естественной и компактной форме всю информацию, содержащуюся в анкете.

Язык описания данных (в нашем случае — анкеты) служит для описания структуры данных, т. е. типов составляющих элементов и их структурной взаимосвязи: упорядоченности, вхождения одного элемента в другой и т. п. В некоторых случаях (в целях повышения эффективности) в язык описания данных включаются средства задания информации о конкретном представлении. Эта дополнительная информация, как правило, необязательная и носит рекомендательный характер.

Язык ввода данных служит для описания (или задания) значений данных, в случае структурированных данных — значений всех составляющих элементов. Мы говорим «всех», имея в виду, разумеется, определение их значений и в случаях не заданных явно значений. Естественно, что язык ввода должен быть полностью согласован с языком описания данных, т. е. структура задаваемых значений должна соответствовать абстрактно заданной структуре соответствующих объектов.

В нашем случае под структурными данными, подлежащими описанию, понимаются анкеты. Здесь, следовательно, анкета нас интересует лишь как основной структурный объект разрабатываемой системы. Но этот структурный объект должен адекватно отражать анкету, являющуюся объектом социологического исследования. Сформулированное выше требование о естественности языка описания данных приводит к необходимости максимально возможного синтаксического согласования этих двух понятий анкеты, включая терминологию.

Во избежание недоразумений уточним, что анкета как объект социологического исследования существует в двух видах, которые