

Н. З. Емельянова, Т. Л. Партыка,
И. И. Попов

ОСНОВЫ ПОСТРОЕНИЯ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

*Допущено Министерством образования Российской Федерации
в качестве учебного пособия для студентов учреждений среднего
профессионального образования, обучающихся по специальности
«Программное обеспечение вычислительной техники
и автоматизированных систем»*

Москва
ФОРУМ — ИНФРА-М
2007

УДК 004(075.32)
ББК 32.973я723
Е60

Рецензенты:

д. т. н., профессор каф. «Информатика и программное обеспечение вычислительных систем» Московского государственного института электронной техники (Технического университета) *Л. Г. Гогарина*;
доктор экономических наук, профессор, декан факультета «Информатика» в НОУ «Московский международный институт эконометрики, информатики, финансов и права» (ММИЭИФП) *А. А. Емельянов*;
профессор кафедры информатики Института Информатики РЭА им. Г. В. Плеханова, д. т. н., профессор *Н. В. Максимов*

Емельянова Н. З., Партыка Т. Л., Попов И. И.

Е60 Основы построения автоматизированных информационных систем: Учебное пособие. — М.: ФОРУМ: ИНФРА-М, 2007. — 416 с.: ил. — (Профессиональное образование).

ISBN 978-5-91134-102-2 (ФОРУМ)

ISBN 978-5-16-002967-2 (ИНФРА-М)

В учебном пособии рассматриваются классификация и структура автоматизированных информационных систем (АИС), а также информационных ресурсов и технологий, связанные с ними понятия и определения, роль предметной области. Значительное внимание уделяется стратегиям проектирования АИС.

В качестве примеров рассмотрены базовые классы: документальные АИС, фактографические, экспертные системы и гипертекстовые АИС, а также соответствующие конкретные системы: АИС по законодательству (банк данных ЮРИУС), офисные информационные системы (пакет Lotus Notes), АИС электронной коммерции (АРМ дилера Российской торговой системы), автоматизированные библиотечные системы (продукт TinLib) и др.

Для студентов специальностей «Информационные системы/прикладная информатика (по областям применения)», «Программное обеспечение вычислительной техники и автоматизированных систем», а также для широкого круга специалистов в области информатизации.

УДК 004(075.32)
ББК 32.973я723

ISBN 978-5-91134-102-2 (ФОРУМ)
ISBN 978-5-16-002967-2 (ИНФРА-М)

© Н. З. Емельянова,
Т. Л. Партыка,
И. И. Попов, 2005
© Издательство «ФОРУМ», 2005

Введение

Вторая половина XX в. ознаменовалась крупным технологическим рывком научно-технического прогресса (НТП). Появление в середине XX в. электронно-вычислительных машин/компьютеров открыло новые возможности обработки информации и управления.

Исторически создание вычислительной техники, ЭВМ/компьютеров явилось выдающимся результатом развития электротехники и технических знаний в целом. Постепенно наращивалась память, повышался уровень программного обеспечения, и компьютеры стали приобретать не только вычислительные, но и другие функции [20].

Социальное значение компьютеров первых поколений состояло в возможности автоматизации интеллектуальных функций человека. Это само по себе было чрезвычайным явлением в 40—50-х гг. прошлого столетия, порождало бурные дискуссии типа «может ли машина мыслить», «не вытеснят ли киберы человека» и т. п.

До конца 70-х гг. влияние «техников» в социально-экономическом назначении компьютеров было приоритетным. Это связано с тем, что компьютеры родом были из электротехники и капиталоемкость создания ЭВМ была велика; при этом главное внимание разработчиков сосредоточивалось на процессах роста памяти, быстродействия, комплексирования ЭВМ и создания образцов большой мощности по обработке информации. Результаты технических достижений стали видимы и осязаемы.

Потребитель и его интересы в использовании компьютеров разработчиками, хотя и не отрицались, однако не были приоритетны на этом этапе.

Применение компьютеров для вычислительных задач и особенно для решения управленческих проблем резко изменило ситуацию. «Неожиданно» выяснилась сравнительно невысокая эффективность использования имеющихся в то время ЭВМ. Создание на базе ЭВМ и средств связи автоматизированных систем управления (АСУ):

- технологическими процессами (АСУ ТП);
- предприятиями (АСУП);

- отраслями (ОАСУ);
- общегосударственным уровнем управления (ОГАСУ)

вывело разработчиков и потребителей на новый, более высокий уровень понимания возможностей использования вычислительной техники в социально-экономическом развитии общества.

Создание вычислительных комплексов зеркально отражало жестко централизованную структуру управления в нашей стране и, естественно, в известной мере повторяло трудности и недостатки монолитной структуры управления (по организации потоков информации, их метода свертывания, скоплению, надежности и др.).

Вычислительный комплекс был организован так, что потребитель «обслуживался, приспособлялся, подгонялся» к нему, а не наоборот. Это было неудобством принципиального характера, которое неизбежно сказывалось на эффективности использования комплексов и ЭВМ этих поколений.

Конечно, сравнительно низкая эффективность АСУ объясняется не только слабой ориентацией на потребителя, но и тем, что психология последнего на этом этапе была настроена на неавтоматизированные, традиционные методы обработки информации. Кстати, необоснованной «страшилкой» было мнение, что автоматизация приведет к значительному сокращению рабочих мест и др. Зарубежная и отечественная практика показала, что значительное увеличение возможностей обработки информации при использовании ЭВМ в управлении не уменьшает, а, наоборот, увеличивает число рабочих мест, предъявляя при этом к работнику новые, специфические и более высокие профессиональные требования.

Программное обеспечение как основной инструмент использования ЭВМ за последние полвека превратилось в крупное и обособленное направление высокоинтеллектуальной деятельности, неотъемлемое от компьютера.

Образовалась четко выраженная *компьютерная технология (технико-программное обеспечение)*, которая является исходной базой для всего процесса обработки информации.

Второй технологический рывок — это успехи в микроминиатюризации, нанотехнологиях — позволил создать персональные компьютеры (ПК). Значение создания персонального компьютера в том, *что он стал обслуживать потребителя*, т. е. вектор социального назначения ПК стал таким, каким он и должен быть.

Идеология программного обеспечения также претерпела революционное преобразование. Прежние акценты на ВЦ коллективного использования (ВЦКП) сменились ориентацией на автономного

пользователя, которому доступны собственные вычислительные мощности, а также и ВЦКП через соответствующие сети.

Третий технологический рывок НТП в прошлом столетии — создание новейших средств телекоммуникации, сетей и сетевых методов передачи информации. Появление сетей передачи информации различной сложности и уровня, включая создание «Мировой паутины» — Интернета, создало принципиально новую среду — глобальное информационное пространство без географических границ.

Информационные системы связаны с семантической (содержательной) обработкой информации, которые основаны на использовании закономерностей языкознания, лингвистики, теории классификации, терминологии и др.

Именно от тщательности и глубокого решения семантических проблем *предметной* области зависит непосредственно эффективность конкретных информационных систем, а следовательно, и процесса информатизации как такового.

В настоящем учебном пособии рассматриваются некоторые вопросы классификации, структуры, методы построения и проектирования *информационных систем* с учетом данных аспектов применительно также к *информационным технологиям и информационным ресурсам*.

В *главе 1* представлены общие вопросы терминологии, понятий и классов объектов и процессов, связанных с проблематикой АИС; рассмотрены факторы информатизации и уровни информационных процессов; приводится классификация информации, информационных технологий, автоматизированных информационных систем, мировых информационных ресурсов, их структура; описаны базы данных на машиночитаемых носителях (электронные издания), их типы, а также примеры служб-генераторов баз данных и онлайн-хост-служб. Также здесь рассмотрены обеспечивающие подсистемы (компоненты) АИС, в том числе техническое, программное, информационное (включая лексическое и лингвистическое) обеспечения; классификация типов, форматов, структур данных, файлов, баз данных и коммуникативные (обменные форматы) АИС.

В *главе 2* представлены основные (базовые) типы АИС — фактографические, документальные, интеллектуальные и гипертекстовые; описаны модели и структуры баз данных (БД), табличные БД. В качестве примера сред реализации реляционных и постреляционных БД рассмотрены ADABAS — Natural и Foxpro. Выбор указанных систем объясняется тем, что такие *общеупотребительные* средства, как MS SQL Server, Sybase, Oracle и др. достаточно подробно описаны в имеющейся литературе, в том числе и авторов настоящей

го пособия [6]. Выбранные же для иллюстрации системы, на взгляд авторов, относятся к *элитарным*.

Далее в главе рассмотрены документальные АИС: их происхождение, классические схемы функционирования документальных АИПС. Логическая и физическая структура документальной БД, как и поисковые возможности (а также описание БД и загрузка), приведены на примере классической системы АИПС STAIRS. Рассмотрена также современная система реализации документальных БД — АИПС локального и удаленного доступа Irbis — интерфейс Irbis, структура БД, поиск в АИС Irbis, операторы поискового условия и т. п. Дано также краткое описание информационно-поисковых систем Интернет (более подробно вопрос рассмотрен в пособии авторов [23]).

Рассмотрены экспертные системы (подсистемы, развитие, применение). Закончена глава описанием гипертекстовых технологий и БД (история развития, гипертекстовые БД, информационный поиск в гипертекстовых массивах, классификация гипертекстовых систем).

Глава 3 посвящена описанию основных принципов проектирования АИС, включая традиционные и более современные технологии. Рассмотрены некоторые классы методов проектирования систем (по степеням автоматизации, типизации, адаптивности).

Введено понятие жизненного цикла АИС. Рассмотрены модели жизненного цикла (каскадная, итерационная, спиральная); стадии и этапы канонического проектирования ИС — предпроектная, техно-рабочего проектирования, стадии внедрения, эксплуатации и сопровождения проекта.

Альтернативным подходом к каноническому проектированию в настоящее время является реинжиниринг бизнес-процессов (BPR — business process reengineering), который базируется на сходстве информационных процессов (ИТ-архитектур) и общих бизнес-архитектур и заключается в интеграции информационных технологий в бизнес-процессы. Автоматизация проектирования АИС (CASE-технологии) является одним из подходов к данной интеграции. Рассмотрено ER-моделирование структуры предметной области на примере среды ERWIN (моделирование в ERWIN, идентификация сущностей и связей в ERWIN, графическое редактирование модели, обратное проектирование — reverse engineering).

Здесь же рассмотрены стандарты описания и задания качества прикладного программного обеспечения и АИС. Это ГОСТ Р ИСО/МЭК 9126—93 и зарубежные стандарты (IEC TC, ISO, Baldridge Award, стандарты оценки качества программного обеспечения, методология CMM, модель SCOPE*PROGEPT и др.).

Глава 4 содержит обзор некоторых типичных областей применения АИС, в том числе автоматизированных ИС по законодательству (в качестве примера рассмотрен прототип отечественных АИСЗ — ЮРИУС), офисные информационные системы (пример — офисный комплекс Lotus Notes); автоматизированные информационные системы электронной коммерции (примеры: интернет-магазин, платежная система «Элит», Российская торговая система — РТС); автоматизированные библиотечные информационные системы (примеры АБИС — ISIS, TinLib V280), корпоративные АИС, основанные на принципах MRP, MRP-II, ERP.

Кратко описаны также другие типы систем — географические информационные системы (ГИС), информационные системы вузов (АИС ВУЗ), Информационные системы медицинских учреждений (Hospital Information System — HIS), а также прочие профессионально-ориентированные АИС (рынок прикладного программного обеспечения России — бухгалтерские, складские, кадровые системы, комплексные системы предприятия, «конструкторы» автоматизации, персональные финансовые программы, средства подготовки документов, составления налоговой декларации, бизнес-плана и пр.

В приложениях приведены: глоссарий терминов, методические указания по выполнению лабораторной работы «Исследование физической структуры и навигации в БД на примере АИСЗ ЮРИУС», пример отчета о подобной лабораторной работе.

Следует отметить, что авторы вынуждены рекомендовать читателю в процессе чтения настоящего учебного пособия обращаться к литературе, ранее вышедшей в данном издательстве, а именно:

Голицына О. Л., Максимов Н. В., Попов И. И. Базы данных: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2003. 352 с. (номер [5] в библиографическом списке);

Голицына О. Л., Попов И. И. Основы алгоритмизации и программирования: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2002. 432 с. [6];

Максимов Н. В., Попов И. И. Компьютерные сети: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2003. 336 с. [24];

Партыка Т. Л., Попов И. И. Информационная безопасность: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2002. 368 с. [28];

Партыка Т. Л., Попов И. И. Операционные системы, среды и оболочки. М.: ФОРУМ: ИНФРА-М, 2003. 400 с. [29].

В указанных книгах более подробно освещены вопросы организации и реализации баз данных и файловых систем, правового обеспечения АИС, а также проблемы информационного поиска в Интернете и организации распределенных БД.

Учебное пособие базируется на материалах, накопленных авторами в процессе практической и исследовательской деятельности, а также преподавания в МИФИ, МИСИ, МЭСИ, РГГУ.

Авторы выражают благодарность коллегам, принявшим участие в обсуждении материала: Максиму Н. В. (РГГУ), Курбакову К. И. (РЭА им. Г. В. Плеханова), а также Голицыной О. Л. (РГГУ), Белобровцевой О. В. (РТС), Яхимович Ю. В. (АО Консультант плюс), за предоставленные иллюстративные материалы.

Глава 1

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ. ОСНОВНЫЕ ПОНЯТИЯ И КЛАССИФИКАЦИЯ

Прежде чем перейти к рассмотрению предмета «информационные системы», следует разобраться как в истории вопроса, так и найти его место в общем контексте процессов и задач информатизации.

Мы предполагаем вкратце рассмотреть:

- факторы информатизации и этапы развития информационных процессов и систем;
- уровни информационных процессов;
- классификацию информации, информационных технологий, систем и ресурсов;
- функциональную структуру и обеспечение (подсистемы) информационных систем.

1.1. Информационные объекты и процессы.

Классификация

Факторы информатизации и последовательность развития информационных процессов и систем

В качестве таковых могут быть определены:

- а) технические средства информатизации (аппаратурный фактор);
- б) программные средства и системы (программный фактор);
- в) информационный фактор — собственно информация, т. е. сигналы, сообщения, массивы данных, файлы и базы данных (БД);
- г) интеллектуальные усилия и человеческий труд (человеческий, гуманитарный фактор). Всегда присутствует человек-пользователь,

решающий задачи какой-либо предметной области с использованием инструментария информатики.

Перечисленные компоненты информатизации не случайно названы *факторами*, поскольку подобно классическим экономическим *факторам производства* (труд, капитал, земля) они:

- взаимозаменяемы (одна и та же производительность может быть достигнута при различных сочетаниях факторов — математически это описывается *кривой безразличия*);
- эффективность производства при увеличении одного из факторов, но при фиксированном вкладе остальных, увеличивается, но все медленнее и медленнее (*закон убывающей производительности*), что требует гармоничного развития всех составляющих, и не последняя роль здесь отводится человеческому фактору. В частности, пользователь должен соответствовать уровню информационных технологий.

Все названные компоненты (факторы) в значительной степени взаимозаменяемы при решении задач. Это означает, что в широких пределах некоторый результат может быть получен, а задача — решена как в рамках электронных схем, так и посредством программ или информационных ресурсов (а также интеллектуальными усилиями человека).

Например, функция

$$y = e^x$$

может быть вычислена:

- 1) *аппаратурно* (усилитель с нелинейными обратными связями);
- 2) *программно* (вычисление суммы ряда для e^x);
- 3) на основе *базы данных* (в качестве таковой здесь может фигурировать сборник математических таблиц, содержащих соответствующие значения y и x).

Аналогично могут быть приведены такие примеры, как перемножение двух переменных, построение случайной последовательности чисел и т. п. Ниже будет рассмотрена конкретная аппаратная реализация поисковых процессов в АИС (машины баз данных).

Перечисленные факторы соответствуют также *исторической последовательности развития информатики* (автоматизированных информационных технологий — АИТ и систем — АИС). Можно выделить следующие этапы развития, на каждом из которых доминирует какой-либо из упомянутых факторов:

- *технический период*, в течение которого сложились основные представления о структуре универсальных электронных вы-

- числительных машин (ЭВМ), определилась архитектура и типы устройств — с 1946 по 1964 г. (приблизительно);
- *программный период* — выработалась современная классификация программных средств, их структур и взаимосвязей, сложились языки программирования, разработаны компиляторы и принципы процедурной обработки; с 1954 по 1970 г.;
 - *информационный период* — в центре внимания исследователей и разработчиков оказываются структуры данных, языки описания (ЯОД) и манипулирования (ЯМД) данными, непроектурные подходы к построению систем обработки информации — с 1970 г. по настоящее время;
 - *гуманитарный период*, связанный с резким возрастанием круга пользователей АИТ и повышением роли интерфейсных и навигационных возможностей соответствующих систем (с начала 90-х гг. прошлого века). Кроме этого, основные черты новых информационных технологий связаны с усилением персонального характера компьютера и расширением возможностей пользователя. Традиционные АИТ были подчинены производителю информации и доводили одинаковое содержание до всех адресатов. Новые АИТ направлены на индивидуального пользователя, предоставляя возможность получения информации, нужной именно ему.

Заметим, что чисто аппаратурное решение задач было положено в основу *аналоговых вычислительных машин (АВМ)*, в настоящее время практически забытых, хотя в 1950—1960-е гг. они широко применялись для имитационного моделирования сложных динамических систем, описываемых системами дифференциальных уравнений.

Здесь же надо отметить, что техническое, программное и информационное обеспечение образуют различные *слои обработки информации*, взаимодействие между которыми должно быть сбалансированным (в смысле отсутствия чрезмерно «толстых» или «тонких» слоев).

Уровни информационных процессов

Данное понятие характеризует степень связи информационных процессов с предметной областью: информационные технологии; информационные системы; информационные ресурсы.

а. *Автоматизированную информационную технологию* определим как целенаправленное и согласованное использование: технических средств информатизации (аппаратурный фактор); программных средств и систем (программный фактор); информационных масси-

вов и баз данных (информационный фактор); интеллектуальных усилий и человеческого труда (человеческий, гуманитарный фактор), для решения задачи (задач) предметной области.

б. *Информационные системы* определяются как комплексы информационных технологий, ориентированных на процедуры сбора, обработки, хранения, поиска, передачи и отображения информации предметной области.

в. *Информационные ресурсы* — комплексы соответствующих информационных систем, рассматриваемые прежде всего на социально-экономических уровнях описания и применения.

В принципе, можно утверждать, что информационные технологии являются менее зависимыми от структуры и специфики предметной области, чем информационные системы и/или ресурсы, однако эта связь всегда существует.

Классификация информации

Понятие *информации* является таким же фундаментальным, как *материи*, *энергии* и других философских категорий. Это атрибут, свойство сложных систем, связанное с их развитием и самоорганизацией. Известно большое количество различных определений *информации*, *данных*, *знаний* и пр. Мы ограничимся только рассмотрением некоторых практически важных понятий и определений.

Основаниями (признаками) классификации информации могут быть:

- а) структура данных и форма представления информации (табл. 1.1);
- б) содержание, предметная область применения (получения) (табл. 1.2).

Более подробно классы информации (данных) рассмотрены ниже.

Отметим, что разделение информации на *табличную (числовую)*, *текстовую* и *графическую* отражает последовательность, в которой эти виды «осваивались» компьютерами. Первоначальные языки программирования (ЯП) были рассчитаны прежде всего на обработку числовой (Fortran, Algol), нежели символьной информации. Раньше появляются и табличные базы данных, также преимущественно рассчитанные на обработку числовых таблиц (файлов). Затем осваиваются текстовые файлы (текстовые редакторы) и текстовые БД (автоматизированные информационно-поисковые системы — библиографические и полнотекстовые). Наконец, с существенным повышением быстродействия и емкости памяти компьютеров на

Таблица 1.1. Основные классы информации по структуре и форме

Основание для классификации	Классы информации			
	По уровням сложности	Сигнал	Сообщение, документ	Информационный массив
По типу сигнала	Аналоговая (непрерывная)	Цифровая (дискретная)	—	—
По уровням доступа и организации	Данные в регистровой памяти	Данные в оперативной памяти	Файлы данных на внешних устройствах	Базы данных
По способам кодирования и представления (данные, файлы и БД)	Цифровая (вычислительные данные, двоичные)	Символьная (алфавитно-цифровая, строчная)	Графическая	—
По организации данных (файлы и БД)	Табличная	Текстовая	Графическая (мультимедиа)	—

Таблица 1.2. Классификация информации по содержанию (предметная область)

Тип информации	Содержание	Поставщик содержания
Биржевая и финансовая	Индексы рынка, котировки, цены, обзоры	Биржи, банки, службы финансовой информации
Экономическая: демографическая статистика	Первичная и вторичная; национальная, региональная статистика	Переписи: опросы, аналитические исследования
Коммерческая	Данные о предприятиях, товарах, услугах	Аналитические службы
Деловые новости	Состояние рынка, события в области экономики	Службы фильтрации, агентства новостей
Научно-техническая	Фундаментальные, прикладные науки	Центры НИИ, издательства, библиотеки
Правовая	Нормативно-правовые акты	Законодательные органы, Минюст
Медицинская	Медучреждения, болезни, лекарства, яды	Информационные центры, библиотеки, госпитали
Потребительская и развлекательная	Образование, музыка, музеи, библиотеки, кино	Справочные службы, учреждения
Бытовая	Погода, туризм, справочники	Информационные службы

сцену выходят графические и другие мультимедийные файлы (графические, аудио, видеоредакторы). Говорить о графических (мультимедиа) базах данных и АИС пока все же преждевременно.

Интересно, что эта последовательность прямо противоположна той, в которой данные виды информации осваивает человек. Действительно, сначала он знакомится с графическими образами (птички, цветочки и бабочки на шкафчиках для одежды в детском саду), затем учится читать и писать, а только потом осваивает таблицу умножения.

В то время как классификация по структуре и форме является более характерной и важной с точки зрения *информационных систем и технологий*, классификация по содержанию (см. табл. 1.2) более соответствует уровню *информационных ресурсов*.

1.2. Основные классы информационных технологий

Рассмотрим вначале некоторые из базовых (элементарных) информационных технологий, связанных в основном с обработкой документов (табл. 1.3).

Таблица 1.3. Основные информационные технологии

№	Тип технологии	Входная информация	Выходная информация	Пример программных средств
«Чистые технологии» — обработка одного типа данных				
1	Обработка текстов (документов)	Текст .txt, .doc, .rtf., .pdf	—	Lexicon, Word, Adobe Acrobat
2	Обработка таблиц	Таблица .xsl, .dbf	—	Access, Excell, Foxpro
3	Обработка статических изображений	Картинка .pix, .pcx, .bmp, .gif	—	Paint, PhotoShop
4	Обработка звукового сигнала	Оцифрованный звук .wav, .mp3	—	Magix 4 in One
5	Обработка видео	Видеосигнал .avi, .mpg, .dat, vob	—	Adobe Premier, Pinnacle Studio
«Смешанные» технологии — преобразование типов данных				
6	Распознавание образов	Картинка .pix, .pcx, .bmp, .gif	Текст .txt, .doc, .rtf., .pdf	FineReader
7	Распознавание голоса	Оцифрованный звук .wav, .mp3	Текст .txt, .doc, .rtf., .pdf	«Змей-Горыныч» (Dragon Dictate)
8	Чтение текста	Текст .txt, .doc, .rtf., .pdf	Оцифрованный звук .wav, .mp3	Govorilka, Speak & Mail

Работа с текстовой информацией

Известно, что существуют различные типы текстовых файлов (плоские, размеченные, ASCII и пр. — см. ниже). Соответственно для ввода, обработки, представления информации в таких файлах требуются различные программные возможности. Для работы с текстами на компьютере используются программные средства, называемые текстовыми редакторами или процессорами. Существует большое количество разнообразных текстовых редакторов, различающихся по своим возможностям, — от очень простых учебных до мощных, многофункциональных программных средств, называемых издательскими системами, которые используются для подготовки к печати книг, журналов и газет. Эти программы обрабатывают различные типы и форматы текстовых файлов, преобразуя их друг в друга по необходимости.

Например, Word позволяет работать с текстовыми файлами в коммуникативном (тип файла .rtf), внутреннем (.doc), и текстовом (.txt) форматах. Только в текстовом формате (плоский текст — .txt) работают редактор Notepad, встроенные редакторы оболочек Norton Commander и Far Manager.

Основное назначение текстовых редакторов — создавать текстовые файлы, редактировать тексты, просматривать их на экране, изменять формат документа, распечатывать его на принтере.

Необходимо отметить, что наиболее развитые редакторы позволяют обрабатывать не просто *тексты*, а *документы* (тексты, содержащие встроенные или внедренные объекты или файлы других типов — табличные, графические, мультимедиа и пр.).

Работа с графической информацией

Общеизвестно, существует значительное разнообразие форматов представления графики в файлах растровой информации.

Для построения, коррекции, сохранения и печати рисунков и других изображений используется *графический редактор*.

Рисовать изображения можно как в режиме ручной прорисовки, так и с помощью базовых инструментов для создания простых фигур (отрезок, прямоугольник, круг, эллипс и т. д.), или *графических примитивов*. Чтобы воспользоваться инструментом, необходимо выбрать соответствующий графический примитив и установить курсор в ту точку экрана, где необходимо изобразить фигуру.

Естественно, создавая изображения на экране компьютера, можно не только рисовать их самостоятельно, но и использовать го-

товые, например, фотографии, рисунки из книг и т. д., полученные с помощью сканера или цифрового фотоаппарата.

Функции всех графических редакторов приблизительно одинаковы (один из простейших графических редакторов для IBM-совместимых компьютеров — PaintBrush).

Они позволяют:

- создавать рисунки из графических примитивов;
- применять для рисования различные цвета и «кисти» (т. е. использовать линии различной ширины и конфигурации);
- «вырезать» рисунки или их части, временно хранить их в буфере («кармане») или запоминать на внешних носителях;
- перемещать фрагмент рисунка по экрану;
- «склеивать» один рисунок с другим;
- увеличивать фрагмент рисунка для того, чтобы прорисовать мелкие детали;
- масштабировать (изменять размер) изображение, выполнять его перемещение и поворот;
- добавлять к рисункам текст.

Кроме простейших, в настоящее время широко распространены достаточно мощные программные продукты обработки изображений и подготовки оригинал-макетов изданий, включающих множество графических образов. Примером может служить Adobe PhotoShop, позволяющий дополнительно осуществлять следующие операции:

- анализировать цветовой спектр изображения или его фрагментов;
- управлять палитрой изображения, обогащая его теми или иными чистыми цветами или их композициями;
- создавать многослойные изображения с управлением прозрачностью каждого слоя;
- наносить спецэффекты, ретушировать в цвете изображения в различных масштабах, вплоть до увеличения в 100—1000 раз.

Табличная информация и ее обработка

Следует различать два уровня табличной информации:

- *электронные таблицы* — размещение в оперативной памяти, сложные алгоритмы обработки, относительно небольшие объемы;
- *табличные базы данных* — размещение на НМД, большие объемы, несложные алгоритмы обработки, базирующиеся на поиске

информации в таблицах и логическом связывании последних (кратко рассматриваются в гл. 2, подробнее, например, в [6]).

Электронные таблицы (или табличные процессоры) — это прикладные программы, предназначенные для проведения табличных расчетов. Они позволяют создавать *динамические таблицы*, которые содержат *вычисляемые поля*, значения которых автоматически пересчитываются по заданным формулам при изменении значений данных, содержащихся в других полях.

Многие фирмы-разработчики программного обеспечения для ПК создали свои версии табличных процессоров — прикладных программ, предназначенных для работы с электронными таблицами. Из них наибольшую известность приобрели Lotus 1-2-3 (фирма Lotus Development), Supercalc (разработка Computer Associates), Multiplan и Excel (Microsoft).

Электронные таблицы просты в обращении, быстро осваиваются непрофессиональными пользователями и во много раз упрощают и ускоряют работу бухгалтеров, экономистов, ученых, конструкторов и людей целого ряда других профессий, деятельность которых связана с постоянными расчетами.

Технологии распознавания изображений

В то время как переход от символьной информации к графической (растровой) достаточно элементарен и без труда осуществим, например, при выводе текста на экран или печать, обратный переход (от печатного текста к текстовому файлу в машинном коде) весьма затруднителен. Именно в связи с этим для ввода информации в ЭВМ первоначально использовались *перфоленты*, *перфокарты* и др. промежуточные носители, а не исходные «бумажные» документы, что было бы гораздо удобнее. «В защиту» перфокарт скажем, что наиболее «продвинутые» устройства перфорации делали *надпечатку* на карте для проверки ее содержания.

В середине 70-х гг. прошлого века была предложена технология OCR — Optical Character Recognition (оптическое распознавание символов) для ввода информации в ЭВМ, заключающаяся в следующем:

- сходный документ печатается на бланке с помощью пишущей машинки, оборудованной *стилизованным* шрифтом (каждый символ комбинируется из ограниченного числа вертикальных, горизонтальных, наклонных черточек, подобно тому, как это делаем мы и сейчас, нанося на почтовый конверт цифры индекса);

- полученный «машинный документ» считывается оптоэлектрическим устройством (собственно OCR), которое кодирует каждый символ и определяет его позицию на листе;
- информация переносится в память ЭВМ, образуя электронный образ документа или документ во внутреннем представлении.

Очевидно, что по сравнению с перфокартами (перфокартами) OCR-документ удобнее хотя бы тем, что он без особого труда может быть прочитан и проверен человеком, и вообще представляет собой «твердую копию» соответствующего введенного документа. Было разработано несколько модификаций подобных шрифтов разной степени «удобочитаемости» (OCR-A, OCR-B и пр.).

Очевидно также, что считывающее устройство представляет собой сканер, хотя и *специализированный* (считывание стилизованных символов), но *интеллектуальный* (распознавание их).

OCR-технология в данном виде просуществовала недолго и в настоящее время приобрела следующую структуру:

- считывание исходного документа осуществляется универсальным сканером, создающим растрового образ и записывающий его в оперативную память и/или в файл;
- распознавание полностью возлагается на программные продукты, которые, естественно, получили название OCR-software.

Типичным представителем данного семейства программ является ABBYY FineReader, технологический процесс которого включает следующие шаги:

- сканирование исходного документа (страницы);
- разметка областей (ручная или автоматическая), требующих различные виды обработки (страницы разворота книги, таблицы, рисунки, колонки текста и пр.);
- распознавание — создание и вывод на экран текстового файла (с вставленными рисунками и таблицами, если это необходимо);
- контроль правильности (ручной, автоматический, полуавтоматический);
- вывод информации в выходной файл в заданном формате (.doc или .rtf для Word, .xls для Excell и пр.).

Данные, полученные на каждом этапе (изображение, текстовый файл), сохраняются под «общей вывеской» *страницы с номером*, что позволяет в любой момент вернуться и повторить разметку, распознавание и пр.

Предметные технологии

В перечень информационных технологий, кроме перечисленных выше простейших (или базовых, т. е. «беспредметных») технологий, принято включать также предметные технологии (т. е. привязанные к предметной области), или комплексы информационных процессов, по существу представляющие собой узкоспециализированные АИС, или *позадачные системы*.

Такие системы ориентированы на обработку специализированных входных и выпуск стандартных выходных форм или отчетов.

На рис. 1.1 приведены примеры структур информационных потоков подобных систем: соответственно физическое (натуральное) представление производственных процессов (а), финансовый аспект процессов управления (б).

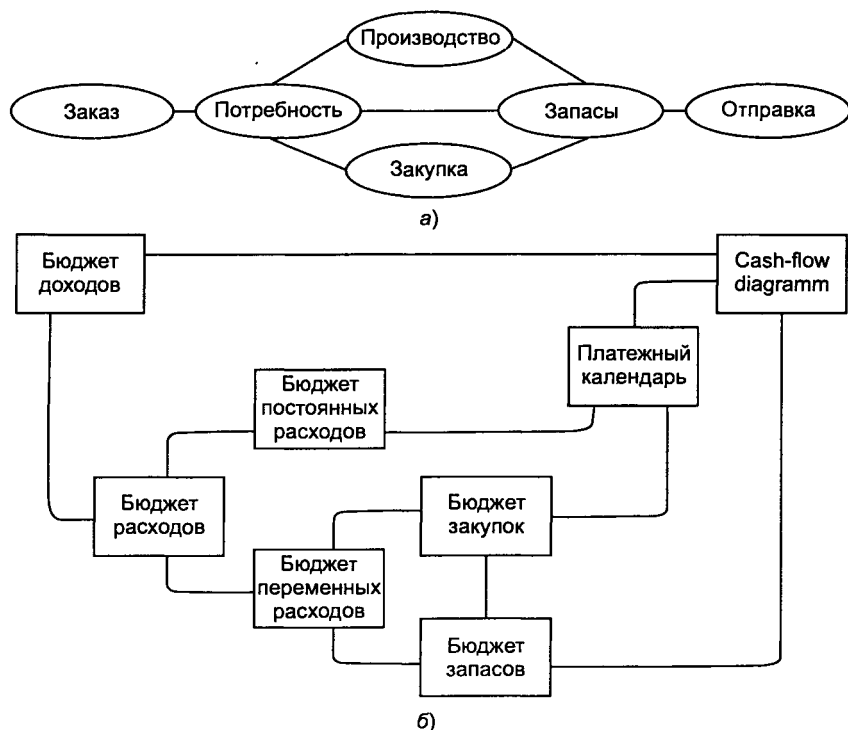


Рис. 1.1. Примеры структур предметных технологий (процессы обработки информации, связанные с управлением производством):

а — физическое представление производственных процессов; б — финансовый

1.3. Автоматизированные информационные системы (АИС), структура и классификация

АИС может быть определена как комплекс автоматизированных информационных технологий, предназначенный для *информационного обслуживания* — организованного непрерывного технологического процесса подготовки и выдачи потребителям научной, управленческой и др. информации, используемой для принятия решений, в соответствии с нуждами для поддержания эффективной деятельности [30].

Компоненты и структуры АИС

Рис. 1.2 отображает структуру типичного совокупного технологического процесса АИС, или представление АИС как совокупности функциональных подсистем — *сбора, ввода, обработки, хранения, поиска, распространения* информации.

Очевидно, это обобщение схем, приведенных на рис. 1.1.

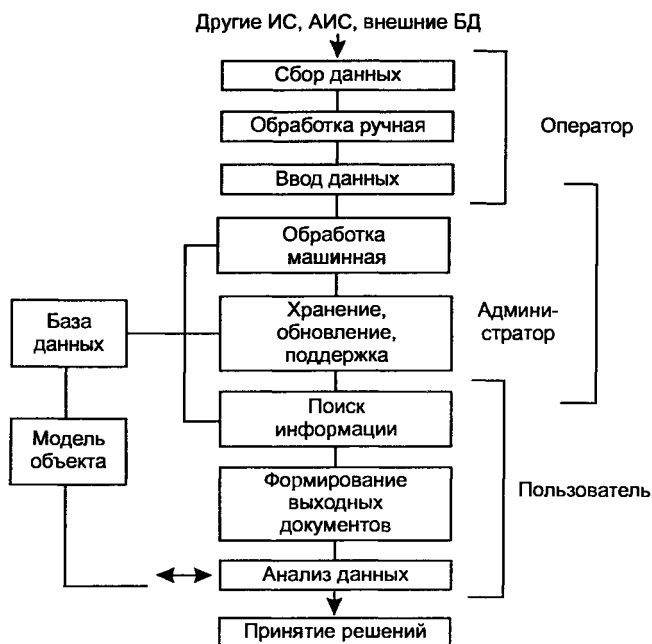


Рис. 1.2. Основные технологические процессы АИС

Некоторые элементы рис. 1.2 являются альтернативными (необязательными):

- *модель объекта* может отсутствовать либо отождествляться с базой данных, которая часто интерпретируется как *информационная модель предметной области*, структурная (для случая *табличных, фактографических* БД) или содержательная (для случая *документальных* БД). В экспертных системах в качестве модели объекта (предметной области) фигурирует *база знаний* (БЗ), представляющая собой *процедурное развитие* понятия БД (БД, по своей сущности, непроцедурный объект);
- *модель объекта и БД* могут отсутствовать (а соответственно и процессы хранения и поиска данных), если система осуществляет динамическое преобразование информации и формирование выходных документов без сохранения исходной, промежуточной, результирующей информации. Если *преобразование данных также отсутствует*, то подобный объект информационной системой не является (он не выполняет *информационной деятельности*), и должен быть отнесен к другим классам систем (например, канал передачи информации и т. п.);
- *процессы ввода и сбора данных* являются необязательными, поскольку *вся необходимая и достаточная* для функционирования АИС информация может уже находиться в БД и составе модели и т. д.

Классификация АИС

В связи с тем, что АИС относятся к сложным системам, целесообразно рассмотреть различные основания для классификации АИС (табл. 1.4).

1. *Отрасли применения* связаны с понятием сектора информационного рынка, соответствуют в основном типам деятельности пользователей, на информационное обеспечение которых ориентирована та или иная АИС. Если использовать укрупненные категории, то можно сказать, что современный информационный рынок представляют три отрасли [34, 35, 36]:

1) информация (базы данных и системы предоставления информационных ресурсов пользователям), в том числе:

- деловая;
- научно-техническая и информация для специалистов;
- потребительская и развлекательная;

Таблица 1.4. Классы автоматизированных ИС

Классы АИС	1. Отрасли применения						
	Деловая информация		Профессиональная информация		Потребительская информация	Электронная коммерция	
	2. Вид и структура информации						
	Библиографическая		Полнотекстовая		Справочная	Табличная	Графическая
	3. Взаимодействие с пользователями						
	Системы разделения времени			Поддержка групповых решений			
	4. Типы принимаемых решений						
	Степень неопределенности			Уровень руководства			
	5. Масштаб АИС						
	АИС организации		Локальные АИС		Глобальные АИС		
	6. Тип организации						
	Производство	Учреждение	Библиотека	Вуз	Госпитали	Прочие организации	
	7. Программное обеспечение						
	СУБД		АИПС		Универсальные оболочки	Программируемые среды	
8. Тип интерфейса с пользователем							
Командный язык		Шаблоны		Меню	Оконный		
9. Наличие и тип моделей, встраиваемых в АИС							
СОПР, КИС (EPR/BRP)		САПР		ЭС	ГИС		

2) электронные сделки (системы электронной торговли, банковские, биржевые и финансовые операции, продажа билетов и резервирование мест и пр.);

3) электронные коммуникации (электронная почта и передача данных).

2. Вид информации отражает структуру данных, поддерживаемую в БД АИС (см., например, ниже рис. 1.4—1.7):

- библиографические данные;
- полнотекстовые документы;
- справочные БД (указатели);
- численные БД;
- графические БД — основную долю составляет растровая или векторная графическая информация.

3. Методы взаимодействия с пользователями определяют две группы информационных систем [30]:

- системы с разделением времени (СРВ), в которых каждый участник как бы пользуется собственной ЭВМ и основной задачей администраторов и разработчиков является защита данных от несанкционированного доступа и взаимная изоляция участников;
- системы обеспечения групповых решений (СОГР) или Computer Supported Cooperative Work, groupware, которые ориентированы на прямо противоположную задачу — обеспечить взаимодействие пользователей в процессе принятия решений. СОГР сочетают коммуникационную, вычислительную технологии и технологию принятия решений для облегчения формулирования и решения неструктурированных проблем группой лиц. Системы, рассматриваемые в настоящем пособии — Lotus Notes и АИС электронной коммерции, — в принципе относятся к данному разряду.

4. Типы принимаемых решений характерны для информационных систем, используемых в экономике и управлении. Под информационными системами управления (ИСУ) принято понимать основанные на компьютерной технологии системы, предназначенные для обеспечения руководителей всей необходимой информацией. Основные подходы к классификации ИСУ базируются на двух различных аспектах классификации управленческих решений: степени неопределенности и уровнях руководства.

По степени неопределенности могут быть выделены четыре способа решений в зависимости от неопределенности целей и структуры взаимосвязи элементов организации:

- формально-логический вывод (вычисление или использование ЭС);
- коллективное обсуждение;
- использование рыночного механизма;
- интуитивное решение.

По уровням руководства выделяют: стратегические, административные, оперативные решения. Отдельный класс составляют ИСУ, предназначенные для контроля за исполнением решений.

Среди систем обеспечения управления (Management Support Systems — MSS) различают: системы информационного обеспечения (СИО), СОПР (см. ниже) и системы обеспечения руководства (СОР). СИО на основе СУБД предоставляют доступ к данным независимо от типа принимаемых решений и предметной области; СОПР обеспечивают принятие решений по специфическому классу

проблем; СОР предназначены для руководителя или группы управляющих.

5. *Масштаб АИС* определяется уровнем организации и функционирования системы, спектром информационного обслуживания, объемом информационных массивов и потоков. Различают следующие классы АИС:

- организации или ее подразделения;
- локальные (региональные или отраслевые);
- глобальные (межотраслевые и, как правило, межрегиональные). Основным типом глобальных АИС являются онлайн-вые службы (хост-службы), предоставляющие доступ удаленным пользователям по телекоммуникационным сетям к некоторому множеству БД.

6. *Тип организации*, использующей АИС, также является основанием для типизации систем; соответственно могут быть выделены следующие АИС:

- различных видов производств;
- административно-управленческих организаций;
- библиотек и информационных центров (АИБС);
- вузов (АИС ВУЗ);
- медицинских учреждений и пр.

7. *Классификация по типу используемого программного обеспечения.*

Во-первых, программные реализации БД различаются по типам БД и структурам данных, их образующих:

- табличные, текстовые, графические БД, что соответствует основным видам данных в ЭВМ;
- документальные и фактографические БД, что соответствует в принципе табличным и текстовым БД;
- реферативные и полнотекстовые, как разновидности документальных БД.

Во-вторых, для реализации документальных БД могут быть использованы по крайней мере два альтернативных средства [30, 31]:

- универсальные оболочки (ISIS, Irbis), относительно закрытые для расширения силами пользователей;
- специальные разработки в среде реляционных или постреляционных СУБД (системы программирования Foxpro, ORACLE, ADABAS), открытые для развития.

Ниже рассматриваются как табличные БД, так и текстовые, реализованные в рамках документальной программной оболочки — реферативные (TinLib), либо в среде реляционной базы данных — реферативные и полнотекстовые (Jurius/Foxpro).

8. *Классы интерфейсов конечного пользователя.* Пользовательский интерфейс должен обеспечивать выполнение следующих функций:

- получение справки о базах данных, доступных пользователю данной АИС;
- получение информации о структуре БД (структура документа/записи, типы полей — доступ к словарию БД);
- информирование пользователя о спектре значений данных в БД (доступ к частотному словарю БД);
- формулирование запроса (поискового выражения, предписания, критерия) и выполнение поиска;
- просмотр результатов поиска в различных представлениях (подсхемах) — фиксированных или произвольных;
- использование полученных результатов для их статистической или содержательной обработки и/или их встраивания в документы, над которыми работает пользователь.

Эти возможности интерфейсов могут реализовываться в рамках следующих основных классов АИС (см. табл. 1.4):

- с *языковым интерфейсом* — предусматривает взаимодействие с пользователем в рамках некоторой системы команд (является ранней формой интерфейсов и в настоящее время обычно используется в профессионально ориентированных системах, рассчитанных на высококвалифицированных пользователей);
- с интерфейсом *форматированного экрана* (иногда именуется Query By Example — QBE — поиск по шаблону);
- с интерфейсом системы *меню*, где фрагменты словаря данных и частотных словарей образуют рубрики меню, отмечая которые курсором или указывающим устройством, пользователь комбинирует поисковое выражение;
- с комбинированными *оконными* интерфейсами, которые включают форматированный экран, меню, фрагменты командного языка, а также элементы графического пользовательского интерфейса или *виджеты* (кнопки, флажки, списки, полосы прокрутки, радиокнопки и др., см. [29]).

9. *Применяемые в контуре АИС модели* являются необязательной компонентой и могут использоваться для поддержки принятия решения пользователем. В зависимости от вида моделей выделяются специфические (интеллектуальные или интеллектуализированные) виды АИС:

- *системы обеспечения принятия решений* (СОПР) — системы, использующие модели объекта управления, опирающиеся на вычислительные или имитационные модели и расчетные методы (линейное программирование, теория массового обслуживания

ния, сетевые модели и пр.). К данному классу относятся и так называемые *интегральные корпоративные информационные системы (КИС)*, базирующихся на принципах MRP (Material Requirement Planning, или методология планирования потребности в материалах), CRP (Capacity Requirements Planning, или планирование производственных мощностей). В процессе развития появились системы MRP с замкнутым циклом, которые впоследствии получили наименование MRP-II (Manufactory Resource Planning) ввиду идентичности аббревиатур. Эти системы были созданы для эффективного планирования всех ресурсов производственного предприятия, в том числе финансовых и кадровых. В последующем системы планирования класса MRP-II в интеграции с модулем финансового планирования FRP (Finance Requirements Planning) получили название систем бизнес-планирования ERP (Enterprise Requirements Planning), которые позволяют наиболее эффективно планировать всю коммерческую деятельность современного предприятия, в том числе финансовые затраты на проекты обновления оборудования и инвестиции в производство новой линейки изделий;

- *системы автоматизированного проектирования (САПР)* — в отличие от СОПР содержат в качестве процедурной или описательной модели предметной области данные и связи, характерные для проектируемого класса объектов (машины и механизмы, электронные схемы, архитектурные сооружения и пр.);
- *экспертные системы* — базируются на логической модели предметной области, реализованной в форме базы знаний и механизма логического вывода. В последнее время все чаще становятся средством обеспечения СОПР и САПР. Кроме того, иногда к интеллектуальным АИС относят некоторые СОПР, которые могут включать в контур обработки информации модели коллективного поведения (базирующиеся на теории игр или экономического равновесия);
- *географические информационные системы (ГИС)* — применяют цифровые модели местности (ЦММ) в различных разновидностях.

В последующих главах относительно подробно рассмотрены примеры некоторых АИС, охватывающих и иллюстрирующих большинство из перечисленных классов:

- *АИС по законодательству* (профессионально ориентированные, документальные, справочные или полнотекстовые, реализованные в универсальных оболочках или реляционных БД,

использующие интерфейсы форматированного экрана или меню, функции СОПР);

- офисные АИС (универсально функциональные, интегрированные, документальные или табличные БД, поддерживают функции СОГР, интерфейсы форматированного экрана);
- системы электронной коммерции (табличные БД, универсальные интерфейсы, функции СОГР, специализированные оболочки);
- библиотечные АИС (табличные и текстовые БД, универсальные и специализированные оболочки, поддерживают тип организации, интерфейсы типа меню).

1.4. Мировые информационные ресурсы, их структура и классификация

Информационные службы и системы, действующие на государственном, международном, глобальном уровне, принято относить к компонентам *мировых информационных ресурсов* (табл. 1.5).

При рассмотрении всего спектра информационных систем нельзя обойти системы данного класса.

На рис. 1.3 приводится обобщенная схема движения мировых информационных ресурсов (структура глобального информационного обмена). Здесь присутствуют: потребители-поставщики инфор-

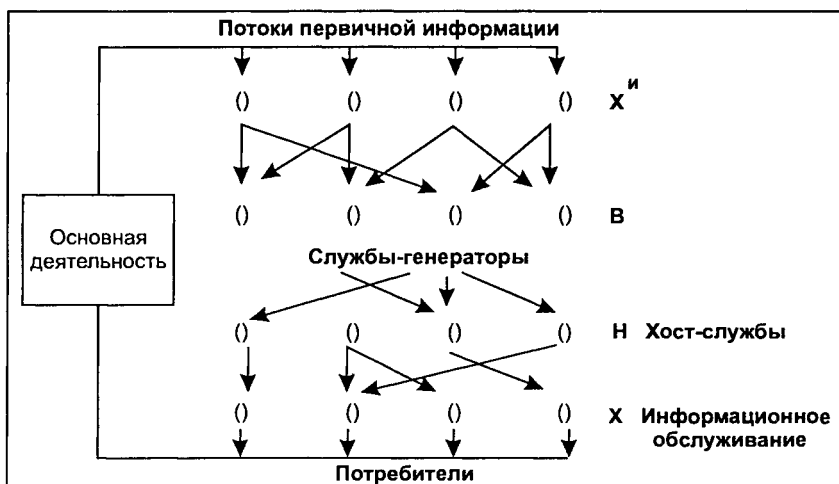


Рис. 1.3. Обобщенная траектория «круговорота» мировых информационных ресурсов

Таблица 1.5. Краткие сведения о типах информационных ресурсов по отраслям

Тип информации	Содержание	Поставщик содержания	Потребители (задачи)	Примеры БД	Хост-служба	Web-серверы
Биржевая и финансовая	Индексы рынка, котировки, цены, обзоры	Биржи, банки, службы финансовой информации	Бюджетирование; управление портфелем; анализ бумаг	BOND BYER BUSINESS DATELINE; FINANCIAL TIMES ONLINE	Dialog, Nexis, GBI	FIC — http://www.fid-inv.com
Экономическая, демографическая статистика	Первичная и вторичная; национальная; региональная статистика	Переписки; опросы, аналитические исследования	Маркетинг	ABI INFORM; LEGAL RESOURCES INDEX; MANAGEMENT CONTENTS	Dialog, GBI	Economic Bulletin Board — http://www.stat-usa.gov/bef/ebhome.htm
Коммерческая	Данные о предприятиях, товарах, услугах	Аналитические службы	Исследования рынка, оценка партнеров	D&B INTERNATIONAL MARKET, MOODIES CORPORATE NEWS	Dialog, GBI	Business Yellow Pages /www.chba.uh.edu/yellowpages.htm
Деловые новости	Состояние рынка, события в области экономики	Службы фильтрации, агентства новостей	Общий бизнес	BUSINESS WIRE, TNT TODAY	Dialog, NEXIS	USA Today — /www/ustoday.com
Научно-техническая	Фундаментальные, прикладные науки	Центры НИИ, издательства, библиотеки	Исследования, разработки	NTIS; INSPEC; AGRICOLA; CHEMNAME; COMPENDEX	Dialog, STN, ORBIT	МЦНТИ — http://icsti.su ; ВИНИТИ — /viniti.msk.su
Правовая	Нормативно-правовые акты	Законодательные органы, Минюст	Применение, правоохрана	CIS, LABORLAW; CONGRESS RECORDS	Dialog, Lexis	Гарант-сервис /www.garant.msu.ru
Медицинская	Медучреждения, болезни, лекарства, яды	Информационные центры, библиотеки, госпитали	Диагностика, лечение	MEDLINE, CANCERLIT, TOXLINE	Dialog, STN	MedNet — /www.mednet.com
Потребительская и развлекательная	Образование, музыка, музеи, библиотеки, кино	Справочные службы, учреждения	Досуг	MAGAZINE INDEX, SPORT	Dialog	Enjoi — http://www.enjoir.ru/
Бытовая	Погода, туризм, справочники	Информационные службы	Досуг	NATIONAL NEWSPAPER INDEX	Dialog, Nexis	Погода России /meteo.infospace.ru

мации (слева) и система управления мировыми информационными ресурсами (справа), в первом приближении включающая в себя два звена:

- службы-генераторы баз данных;
- службы-операторы баз данных (онлайновые или хост-службы).

Базы данных на машиночитаемых носителях (электронные издания)

Эти БД являются основными элементами инфраструктуры *мировых информационных ресурсов*. Первоначально переход к электронной форме издания БД был совершен службами, выпускавшими реферативные журналы, экспресс-информацию, обзоры и прочие виды вторичных документов, поэтому ранее всего появились библиографические (реферативные, справочные) БД. По мере накопления опыта и появления потребностей в электронных изданиях стали выходить полнотекстовые БД, а также статистические сборники, указатели, словари, справочники и прочие виды информации, образующие сегодняшний рынок разнообразных информационных ресурсов.

Типы баз данных

Выделяются следующие типы баз данных.

1. Библиографические данные (записи) — выходные данные первоисточников (включают авторов, заголовки, классификационный индекс, место публикации и пр.), иногда реферат (рис. 1.4).

2. Полнотекстовые документы (записи) — полный (или почти) исходный текст журнальной статьи или другого документа (рис. 1.5).

3. Справочники (указатели) — списки (перечни, словари, описания продуктов, веществ, проектов, организаций, грантов, персон и другая систематизированная информация об объектах) — рис. 1.6.

4. Численные данные — записи содержат таблицы статистических, финансовых и других сведений (рис. 1.7).

В настоящее время по разным оценкам функционирует от 5 до 10 тыс. информационных служб, выпускающих БД, каждая из которых производит от 1 до 20 баз данных. В табл. 1.6 приведены данные по некоторым наиболее продуктивным службам как широкого, так и узкоспециализированного профиля, а в табл. 1.7 — примеры баз данных, отличающихся различными параметрами, существенными с точки зрения использования информационного ресурса.

RN: 512012 Wave-front shearing interferometer for cryogenic ;
TITLE: laser-fusion targets. Tarvin J.A.; Sigler R.D.;
AUTH: Busch G.E. (RMS Fusion, Inc., Ann Arbor Michigan **CORP:**
SOURC: 106 Appl.Opt. ISSN 0003-6935. (1 Sep. 1979). v. 18(17)
 p. 2971-2974. **YEAR:** **VVSS:**
ABSTR: The operating principles of a wave-front shearing
 interferometer designed by Murty are described.
 The interferometer has simple independent adjustment
 for both phase and shear and is stable mechanically.
 It has been used to access the quality of cryogenic
 laser-fusion targets immediately prior to their
 irradiation by a high-power laser. It has also
 been used to study the changes in cryogenic targets
 caused by energy emitted by the high-power laser
 prior to the main pulse.
DESCR: TARGETS: quality control

Рис. 1.4. Пример записи БД INIS. БД содержит более чем 5,5 млн записей:
 поля базы данных: RN — номер реферата; TITLE — заглавие документа;
 AUTH — авторы; CORP — корпоративный автор; SOURC — журнал-источник;
 ABSTR — реферат; DESCR — дескрипторы

CN: Rhonda L. Cassity, Petitioner
 vs.
 Commissioner of Internal Revenue, Respondent
CT: United States Tax Court
DK: Docket No 2545-95
TX: Text: Filled April 6, 1997 Harley G. Shneider and James R. Lorander, for petitioner
 Ronald J. Long, for the respondent.
 Memorandum Finding of Fact and Opinion
 GERBER, JUDGE: Respondent determined deficiencies in petitioner's Federal Income
 tax for the taxable years 1985, 1986, 1987 and 1988 in the amount of \$ 166,730.93,
 \$210,648.41, \$225,715.20 and \$ 51,404.14, respectively. The issues for our
 consideration are:
 (1) Whether Federal income tax returns for the years 1985, 1986, 1987 and 1988, filed
 in the names of petitioner and her husband, were joint returns and
 (2) if they are joint returns, whether petitioner is relieved of liability as an «innocent
 spouse» under the provision of section 603(e).

CF: Code Classification — Section 6013 (Joint Returns)
GI: Geographic Identifier: US (United States)
DE: Descriptors: Joint returns

Рис. 1.5. Пример записи БД TNT (Tax Notes Today). БД содержит информацию
 по вопросам налогообложения, включая протоколы рассмотрения спорных дел в
 судах и пр.:
 поля БД: CN — наименование дела (процесса); CT — суд, в котором рассматри-
 валось дело; DK — номер описи судебных дел; TX — полный текст документа

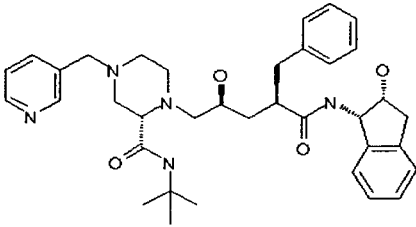
	DIALOG(R)File 453:Prous Science Drugs of the Future
AA=	ENTRY NUMBER: 199183
/NA, NA=	DRUG NAME: L-735524
/SY, SY=	GENERIC NAME: Indinavir sulfate (proposed INNМ; USAN)
/TN, TN=	BRAND NAME: Crixivan (Merck & Co., BR, CA, US)
CN=	CHEM NAME: 2(R)-Benzyl-5-(2(S)-(N-tert-butylcarbamoyl)4-(3-pyridylmethyl) piperazin-1-yl)-4(S)-hydroxy-N-(2(R)-hydroxyindan-1(S)-yl)pentanamide sulfate (1:1)N-tert-Butyl-1-(2(S)-hydroxy-4(R)-(N-(2(R)-hydroxyindan-1(S)-yl)carbamoyl)-5-phenylpentyl)-4-(3-pyridylmethyl)piperazine-2(S)-carboxamide sulfate(1:1)
MF=	FORMULA: C36H47N5O4.H2SO4
RN=	CAS REG. NO.: 150378-17-9 (free base) 157810-81-6
IM=	
ST=	DEVEL. PHASE: Launched (1996)
/LO, LO=	ORIGINATOR: Merck & Co.
/TC, TC=	CLASS: 71521 (Antiviral (AIDS)) 71523 (HIV-1 Protease Inhibitor)
SS=	SYNTHESIS: 361961
CX=	CONTEXT TABLE:71521C (Anti-HIV Agents)
/TX	Synthesis : Indinavir has been obtained by several different ways: 1) The reaction of cis-(1S,2R)-indanediol (I) with acetonitrile and concentrated H2SO4 gives cis-(1S,2R)-1- aminoindan-2-ol (II), which is

Рис. 1.6. Пример записи базы данных по химическим структурам и именам CHEMNAME. БД Chemical Abstracts Chemical Names Dictionary выпускается службой CAS (США) и содержит более 2,5 млн записей: поля: CN — наименование вещества; RN — регистрационный номер CAS; TX — текстовое описание процедуры синтеза; RF — библиографические ссылки на связанные публикации; IM — графический образ

	<p>cyclocondensed with 3- phenylpropionyl chloride (III), isopropenyl methyl ether and triethylamine to yield the acetone amide (IV) (1).</p> <p>The condensation of amide (IV) with (S)-(+)-glycidyl p-toluenesulfonate (V) in the presence of lithium hexamethyldisylazide (LHS) affords the chiral epoxide (VI), which is condensed with 4-(tert-butoxycarbonyl)-N-tert-butylpiperazine-2(S)-carboxamide (VII) in refluxing isopropyl acetate and deprotected with aqueous HCl to give the dihydroxy-diamide (VIII). Finally this compound is condensed with 3-(chloromethyl)pyridine (IX) by means of triethylamine in DMF (2, 6). Scheme 1.</p> <p>2) The amide (IV) can also be alkylated with allyl bromide and butyl lithium to the pentenyl amide (X), which is diastereoselectively converted to the chiral iodohydrin (XI) by means of N-.</p> <p>According to data from an ongoing phase II study presented at the XIth International Conference on AIDS, over half (54%) of HIV-positive patients treated for 48 weeks with Merck & Co.'s Indinavir(TM) at 2.4 g/day (800 mg every 8 h) continued to have virus levels reduced below detection. An increase in CD4 cells was also observed and treatment was generally well tolerated. (119)</p> <p>Merck & Co.'s indinavir sulfate (IndinavirTM) in combination with zidovudine and lamivudine has been shown to suppress the HIV virus below the level of detection for 48 weeks in 6 of 7 patients according to data presented at the XI International Conference on AIDS. (120)</p> <p>PREV. PUB. IN: Drugs of the Future, Vol. 21, No. 6, p. 600, 1996 Drug Data Report, Vol. 15, No. 10, p. 965, 1993 Drug Data Report, Vol. 18, No. 5, p. 454, 1996</p>
/RF, RF=	<p>REFERENCES:</p> <ol style="list-style-type: none"> 1. Mealy, N.; Castauer, J., «Indinavir Sulfate», Drugs of the Future, Vol. 21, No. 6, p. 600, 1996 (Synthesis Scheme available. SEE 361961) 2. Verhoeven, T.R.; Roberts, E.F.; Senanayake, C.H., «Ryan, (Merck & Co., Inc.)», Regiospecific processes to make cis-1-amino-2-alkanol from diol or halohydrin. US 5449830. 3. Askin, D.; Eng, K.K.; Rossen, K.; Purick, R.M.; Wells, K.M.; Volante, R.P.; Reider, P.J., «Highly diastereoselective reaction of a chiral, non-racemic amide enolate with (S)-glycidyl tosylate.», Synthesis of the orally active HIV-1 protease inhibitor L-735,524. Tetrahedron Lett 1994, 35: 673-6. (...) 126. Lin, J.H.; Chiba, M.; Balani, S.K.; Chen, I.W.; Kwei, G.Y.S.; Vastag, K.J.; Nishime, J.A., Species differences the pharmacokinetics and metabolism of indinavir, a potent human immunodeficiency virus protease inhibitor», Drug Metab Dispos 1996, 24(10): 1111

Рис. 1.6 (Продолжение)

CO:	Seward Grain, Inc. Ceral	Processing Div	
AD:	400 CORNSHUSKER HWY		
CY:	Lincoln, St NE	ZIP - 68507-3121	
TE:	Phone:	402-467-8441	
CN:	County:	Lancaster	
DE:	Business:	Whol Grain	
SC:	Primary SIC	- 5153 -	Grain - Whols
	Secondary SIC	- 4221	Farm Product - Storage
YR:	Latest year organized:	1987	
IN:	State of incorporation:	NE	
DI:	Date of Incorporation:	March 10, 1986	
	Latest Year	Trend Year (1996)	Base Year (1993)
SA:	Sales(\$): 240,000,000	214,000,000	146,000,000
ET:	Employees total: 350	340	325
EH:	Employees here 35	30	28
SG:	Sales growth (%)	47	
EG:	Employment growth (%)	5	
SQ:	Square footage:	140,000	
BA:	Bank:	Boatmens First	Natl BK
PO:	Chairman:	NA:	Lincoln George A.
	President:		Carmill Robert S.
	Vice President:		Sigourney, David S.
			Mc Kinney, Gary W.
			Miller, Oren L.
	Secretary:		Arndt, Adrian
			Smith, Dennis
	Treasurer:		Wheeler, Charles

Рис. 1.7. Пример записи БД D&B Dun's Market Identifiers. Производится фирмой Dun & Bradstreet и содержит данные более чем по 8 млн предприятий США, с персоналом от 5 чел. или объемом продаж выше 1 млн, данные создаются персоналом аналитиков численностью более 2000 чел.:

поля БД: CO — наименование компании; DE — описание бизнеса; AD — адрес; CY — город; TE — телефон; CN — графство; ST — штат; SC — Standard Industrial Code — классификатор продуктов и услуг; YR — год основания организации; IN — штат инкорпорации; DI — дата инкорпорации; SA — объемы продаж; ET — всего занятых; EH — число занятых здесь; SG — рост продаж; EG — рост занятости; SQ — площадь офисов; BA — банк; PO — должность руководителя; NA — имя руководителя

Табл. 1.8 содержит данные о доступности некоторых БД через различные онлайн-службы (*рассеяние информации*), характеризующие профили как тех, так и других.

В качестве примеров рассмотрим далее одну из служб-генераторов (провайдеров) БД — АИС INIS и службу-оператор (хост-службу) STN.

Таблица 1.6. Профили ряда информационных служб-генераторов БД

Служба	Выпускаемые БД
Службы широкой специализации	
1. Predicasts (Кливленд, Огайо, США), основные БД	Infomat International Business; PTS Aerospace/Defense Markets & Technology; PTS Annual Reports Abstracts; PTS F&S Index; PTS International Forecasts; PTS Marketing & Advertising Reference Service; PTS New Product Announcements/Plus; PTS Newsletter Database; PTS PROMT; PTS U.S Forecast; PTS U.S. Time Series
2. R.R. Bowker (Нью-Йорк, США)	American Library Directory; American Men & Women of Science; Books in Print; Energyline; Enviroline; Microcomputer Software Guide; Publishers, Distributors and Wholesalers; Supertech; Ulrich's International Periodical Directory
3. Chemical Abstract Service (Колумбус, Огайо, США)	CA Search; Chemical Industry Notes; Chemname; Chemsys; Chemzero
4. University Microfilms International (UMI) – (Анн Арбор, США)	Criminal Justice Periodical Index; Dissertation Abstracts Online; Insurance Abstracts; Newspaper Abstracts
5. Information Access Company (IAC)	Computer ASAP; Computer Database; Industry Data Sources; Legal Resources Index; Magazine ASAP; Magazine Index; Management Contents; National Newspaper Index; Newsearch; Newswire ASAP; Trade and Industry ASAP; Trade and Industry Index.
Службы узкой специализации	
1. Institute of Electrical Engineers (Лондон, Великобритания)	INSPEC (Физика электричества, управление)
2. U.S. National Library of Medicine – Medline; Cancerlit.	
3. INIS (IAEA – МАГАТЭ – Вена, Австрия)	INIS (Атомная наука и техника)
4. Derwent Publication Ltd. (Лондон) – World Patent Index (Патенты)	
5. Chemical Abstract Service (CAS)	CA FILE, CA REGISTR, CASREACT (Химическая информация)

Таблица 1.7. Основные эксплуатационные характеристики выпускаемых БД

Предметный охват	
Бизнес	AGIBUSINESS, BUSINESS DATELINE
Наука и техника	COMPENDEX, INSPEC, CURRENT TECHNOLOGY
Досуг	NEWSPAPER INDEX, SPORT
Видовой охват	
Книги	BOOKS IN PRINT, LC-MARC BOOKS, REMARC
Патенты	CLAIMS/U.S. PATENT ABSTRACTS, WORLD PI

Окончание табл. 1.7

Диссертации	DISSERTATION ABSTRACT ONLINE
Газеты	NEWSPAPER ABSTRACTS
Частота обновления	
Постоянно	BOND BYER FULL TEXT, PTS PROMT, CHIC. TRIBUNE
Еженедельно	ABI/INFORM, CURRENT CONTENTS SEARCH
Ежемесячно	AGRICOLA, ANALYTICAL ABSTRACTS
Ежеквартально	AMERICA: HISTORY AND LIFE, ART/LITER INTERNAT.
Тип данных	
Полные тексты	BUISSNESS DATELINE, BUISSNESSWIRE
Библиографические	ABI/INFORM, AEROSPACE, INIS, INSPEC
Справочные (указатели)	AMERICAN MEN AND WOMEN IN SCIENCE
Числовые	D&B – DUN'S FINANCIAL RECORDS, INVESTEXT
Ретроспективность	
До 10 лет	BUISSNESSWIRE
Свыше 20 лет	ABI/INFORM, AGRICOLA, INSPEC, INIS
Свыше 30 лет	BEILSTEIN, AMERICA: HISTORY AND LIFE
Наличие печатного аналога (обычно – РЖ)	
INSPEC	«Science Abstracts»
CA Search	«Chemical Abstracts»
Compendex	«Engineering Index» и т. п.

Таблица 1.8. Сведения о доступе к некоторым БД через различные онлайн-хосты

Хост-службы	Базы данных			
	ABI/INFORM	INFOMAT	INSPEC	DERWENT
Dialog Inc. (Калифорния, США)	+	+	+	+
STN International (Карлсруэ, ФРГ)	+	–	+	+
DataStar (Берн, Швейцария)	–	–	+	–
Orbit (McLean, VA, USA)	–	–	+	+
GBI (Gesellschaft fuer Betrieb-wirtschaftliche Information mbH) Мюнхен, ФРГ	+	+	–	–

Международная система по атомной науке и технике INIS

Система INIS (*INIS — International Nuclear Information System*) является структурным подразделением Международного агентства по атомной энергии (МАГАТЭ), секретариат которого базируется в Вене (Австрия).

Система основана в 1965 г. и предназначена для организации международного информационного обмена в области мирного использования атомной энергии; ее первые продукты появились в 1970 г. INIS является первой в мире децентрализованной системой, основанной на международной кооперации; ее участники — около 100 стран и международных организаций.

Страны и международные организации — участники INIS — взаимодействуют с центром в Вене следующим образом. Национальные центры готовят входной поток и получают выходную продукцию из Вены. Некоторые страны осуществляют подготовку входного потока через региональные центры. Например, система INSPEC готовит входной поток в машиночитаемой форме от имени ряда скандинавских стран, Нидерландов по общедоступным первоисточникам. В то же время национальные центры этих стран самостоятельно подготавливают к вводу в INIS труднодоступные и непубликуемые материалы (отчеты, патенты, стандарты, диссертации и т. д.).

Таким образом, входная обработка и распространение продуктов системы осуществляются участниками и полностью децентрализованы. Только входная проверка, слияние и совместная обработка массивов проводятся в центре (рис. 1.8). Такая структура имеет следующие преимущества:

- более полный охват информации по атомной науке и технике;
- наиболее эффективную обработку документов на различных языках;
- равномерное распределение затрат на сбор и обработку информации между крупными и мелкими поставщиками (потребителями) информации;
- развитие национальных информационных инфраструктур.

Входная информация представляется национальными центрами на рабочих листах или в машиночитаемой форме. Копии труднодоступных первоисточников пересылаются в Клирингхауз INIS для микрофильмирования и изготовления диамикрокарт (микрофиш), а также оцифровываются и хранятся на CD-носителях.

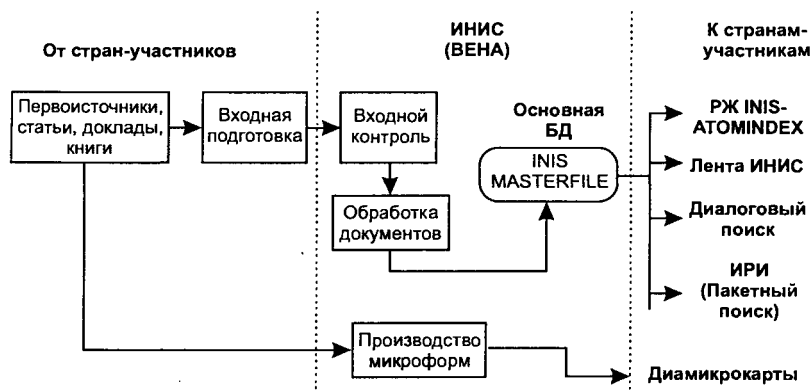


Рис. 1.8. Макросхема технологического процесса в INIS

Программы ввода преобразуют входную информацию во внутренний рабочий формат, причем каждый документ разделяется на две компоненты:

- реферат и библиографические данные;
- дескрипторы поисковых образов документов (ПОД).

Каждая компонента затем обрабатывается соответствующими наборами контрольных программ. Выявленные ошибки устраняются, и массивы передаются на обработку завершающими программами технологического процесса, формирующими главный массив (*Master File*) INIS, являющийся интегральной информационной базой для подготовки выходных продуктов.

Центр в Вене ведет и контролирует тезаурус (INIS Thesaurus), на основе которого осуществляется индексирование документов национальными центрами. Тезаурус непрерывно обновляется, и в него включаются новые термины и связи в процессе обработки входного потока INIS, он содержит примерно 16 тыс. дескрипторов и 5 тыс. запрещенных терминов.

Годовой информационный поток INIS составляет более 100 тыс. документов, включая 20 тыс. труднодоступных и непубликуемых. За период с 1970 г. накоплен фонд копий первоисточников и БД по 5,5 млн документов, из них — 1,5 млн труднодоступных. Предметный охват системы: физика — 40 %, химия, науки о земле, материалы — 23 %; биологические науки — 13 %; изотопы — 2 %; техника, технологические процессы — 18 %; прочее — 4 %.

Основные выходные продукты INIS следующие:

1. БД на CD (CD/МЛ INIS), выходящие 2 раза в месяц и содержащие 5000 документов, представленных на вход системы. Форматы

представления данных выбираются подписчиками и могут быть непосредственно рассчитаны на загрузку БД в АИПС, использующую ту или иную программную оболочку поиска информации. Выпускается с 1975 г.

2. Реферативный журнал по атомной науке и технике (INIS Atomindex), доступный подписчикам печатный журнал, который выходит 2 раза в месяц, подготавливаемый на основе главного массива или МЛ INIS с использованием автоматизированного типографского набора.

3. Труднодоступные документы на микроформах и CD представляются Клирингхаузом INIS по запросам потребителей (имеются в виду те документы, которые освещены в РЖ INIS Atomindex или МЛ INIS — около 20 % всего ввода в систему).

4. Обслуживание запросов пользователей в режиме диалога INIS осуществляет доступ к массивам центра в Вене по сетям связи.

5. Обслуживание разовых запросов в пакетном режиме осуществляется в ограниченном масштабе по запросам европейских международных организаций.

В процессе функционирования центр службы INIS выпускает справочную серию (INIS Reference Series), распространяемую среди стран-участниц МАГАТЭ и регламентирующую подготовку входных документов INIS. Справочные руководства INIS включают в себя:

- нормативный список названий периодических изданий (INIS Authority List for Journal Titles);
- нормативный список написаний названий корпоративных авторов (INIS Authority List for Corporate Entry);
- список префиксов номеров отчетов (эквивалентен ISSN, но для непубликуемых источников) — INIS Authority List for Report Number Prefixes;
- тезаурус (INIS Thesaurus, обсуждался выше);
- многоязычный словарь (переводы дескрипторов на некоторые европейские языки — INIS Multilingual Dictionary).

Онлайновая хост-служба STN International (Scientific and Technical Network)

STN International — научно-техническая информационная сеть, предоставляющая пользователю доступ к большинству мировых БД по науке и технике. Она содержит свыше 170 БД суммарным объемом более 170 млн записей, ретроспектива 20—30 лет (табл. 1.9).

Таблица 1.9. Эксплуатационные характеристики доступа к некоторым БД STN International

Наименование БД	Год	Объем (млн записей)	Периодичность обновления	Провайдер	Час работы (долл.)	Просмотр документа (долл.)	ИРИ ¹ (долл.)	Тип БД	STN Easy ² (долл. в час)
ABI-INFORM	1971	0,66	Еженедельно	University Microfilms Inc. (UMI), USA	90,00	2,10	15,00	Библиография, полный текст	—
BELLSTEIN	1830	4,90	Ежеквартально	Bellstein Institute for Organic Chemistry, Germany	0,00	9,00	0,00	Справочная	—
BIOSIS Previews /RN	1969	8,5	Еженедельно	BIOSIS, U.S.A.	60,00	1,35	10,50	Библиография	2,50
CA	1967	10,0	То же	CAS, U.S.A.	32,00	1,91	6,80	То же	2,50
COMPENDEX	1970	3	» »	Engineering Information, Inc., U.S.A.	60,00	1,85	1,89	» »	3,00
INPADOC	1968	21,0	» »	European Patent Office, Austria	55,00	0,44	1,75	» »	1,00—10,00
INSPEC	1969	5,38	» »	Institution of Electrical Engineers, UK	60,00	1,85	1,89	» »	3,00
METADEX	1966	0,92	Ежемесячно	Materials Information, a joint service of the Institute of Materials, UK and ASM International, U.S.A.	40,00	1,69	1,89	» »	—
MEDLINE	1972	6,30	Еженедельно	U.S. National Library of Medicine, U.S.A.	30,00	0,16	6,00	» »	0,25
NTIS	1964	1,65	Раз в 2 недели	National Technical Information Service, U.S.A.	40,00	1,69	1,89	» »	—

¹ Стоимость обслуживания одного профиля избирательного распределения информации (см. ниже, гл. 2, режимы работы АИПС) за месяц.

² Обслуживание пользователей через Web-серверы STN.

Предметный охват STN:

- медицина и биологические науки — 23 базы данных;
- химия — 45;
- техника — 40;
- энергетика и проблемы окружающей среды — 31;
- материалы и конструкции — 20;
- математика и физика — 16;
- общественные науки — 6;
- патентная информация — 19.

По структуре базы данных STN образуют следующие группы [2]:

- библиографические/bibliographic — 107 БД;
- полнотекстовые/full-text — 17;
- справочные/directory — 29;
- числовые/numeric 24;
- по химическим структурам — 14.

Генераторами БД для STN являются 37 организаций, в том числе:

- Chemical Abstracts Service — 10 БД;
- Deutsche Bibliothek — 2 БД;
- Institution of Electrical Engineers — 1 БД;
- International Patent Documentation Center — 2 БД;
- Engineering Information — 2 БД.

1.5. Обеспечивающие компоненты (подсистемы) АИС

Важным аспектом рассмотрения АИС является описание структуры *обеспечивающих подсистем* (табл. 1.10), соответствующая вышеупомянутым компонентам — *техническим, программным, информационным* средствам, интегрированным *человеческим фактором*.

Техническое обеспечение АИС

Техническое обеспечение (ТО) АИС может быть разделено на две группы — *универсальное и специализированное*.

Универсальное ТО включает в себя средства ввода, обработки, хранения и передачи информации.

Три первых типа средств в настоящее время являются стандартными компонентами универсальных ЭВМ (вычислительных комплексов или установок) независимо от их класса и масштаба. Суще-

Таблица 1.10. Виды обеспечения АИС

Обеспечение АИС	Техническое	Универсальное	ЭВМ и средства коммуникации (передача, ввод, хранение, обработка данных)		
		Специальное	Машины баз данных		
	Программное	Операционные системы	(MS-DOS, OS/2, UNIX, Windows)		
		Системы программирования	Общего назначения (Паскаль, Си, Бэйсик, Фортран)		
			Разработки АИС (FoxPro, ADABAS)		
		Приложения (оболочки АИС)	СУБД (ADABAS, Oracle)		
	АИПС (STAIRS, Irbis, ISIS)				
	Лингвистическое	Форматы (форматная база)	Базы данных онлайн-хостов		
			Данных		
			Файлов		
			Документов	Библиографические (карточный, МЕКОФ)	
		Полнотекстовые (ODA, SGML)			
		Лексические средства	Классификаторы		
			Кодификаторы (ОКП, GS, МКИ, УДК, INIS)		
			Тезаурусы (INIS, INSPEC, ВИНТИ)		
		Информационные языки	Языки манипулирования данными	ЯЗ	
				ИПЯ	
	Языки описания данных		Языковые средства (SQL)		
	Информационное	Файлы	Текстовые		
			Табличные		
			Графические и др.		
Базы данных		Фактографические			
	Документальные				
Организационное	Документы	Общеправовые			
		Профессиональные			
	Процессы	Техническое обслуживание			
		Эксплуатация задач			
		Подготовка данных			
		Обслуживание пользователей			
	Пользователи	Администраторы системы			
		Операторы			
Интерактивные пользователи					
		Конечные пользователи			

ствование тех или иных устройств отображения и ввода информации в электронную форму независимо и вне ЭВМ постепенно сходят на нет.

Средства передачи информации включают в себя сетевое и телекоммуникационное оборудование ЭВМ, а также системы и средства связи общего назначения, обеспечивающие разнообразные виды коммуникации, в том числе между техническими средствами первых трех групп, независимо от того, входят они в состав каких-либо вычислительных комплексов или нет.

Кроме традиционных технических средств (универсальных ЭВМ и систем), длительное время существует альтернативное направление, ориентированное на реализацию АИС.

Здесь нет необходимости в подробном рассмотрении современных технических средств вычислительной техники (ВТ), поскольку читатель и так должен иметь о них достаточно представления, а, кроме того, некоторые из этих аспектов подробно освещены в упоминавшихся выше книгах [6, 24, 28].

Специализированное ТО (Машины баз данных). Основы современных информационных систем составляют БД и системы управления базами данных (СУБД), роль которых как единого средства хранения, обработки и доступа к большим объемам информации постоянно возрастает. При этом существенным является постоянное повышение объемов информации, хранимой в БД, что влечет за собой требование увеличения производительности таких систем. Резко возрастает также спрос на интеллектуальный доступ к информации. Это особенно проявляется при организации ее логической обработки в системах баз знаний, на основе которых создаются современные экспертные системы.

Быстрое развитие потребностей применений БД выдвигает новые требования к СУБД, такие, как:

- поддержка широкого спектра типов представляемых данных и операций над ними (включая фактографические, документальные, мультимедийные данные);
- естественные и эффективные представления в БД разнообразных отношений между объектами предметных областей (например, пространственно-временных с обеспечением визуализации данных);
- поддержка непротиворечивости данных и реализация дедуктивных БД;
- обеспечение целостности БД в широком диапазоне разнообразных предметных областей и операционных обстановок;

- управление распределенными БД, интеграция неоднородных баз данных;
- существенное повышение надежности функционирования БД. Вместе с тем традиционная программная реализация многочисленных функций современных СУБД на ЭВМ общего назначения приводит к громоздким и непроизводительным системам с недостаточно высокой надежностью. Тем более затруднительным оказывается наращивание программных средств, обеспечивающих перечисленные выше требования. Это обусловлено рядом причин:
- фон-Неймановская архитектура ЭВМ неадекватна требованиям СУБД, в частности реализации поиска, обновления, защиты данных, обработки транзакций только программным способом неэффективны как по производительности, так и по стоимости;
- многоуровневое и сложное программное обеспечение СУБД снижает эффективность и надежность функционирования БД;
- универсальная ЭВМ оказывается перегруженной функциями управления базами данных, что снижает эффективность работы собственно прикладных систем;
- централизация и интеграция данных в сетях персональных и профессиональных ЭВМ нереализуема с приемлемой стоимостью без включения в состав сетей специализированных ЭВМ для поддержки функции СУБД.

Эти соображения приводят к мысли о необходимости создания *специализированных автономных информационных систем*, ориентированных исключительно на реализацию функций СУБД. Однако системы, реализованные на обычной универсальной мини- или микро-ЭВМ, не способны полностью решить указанные проблемы. Необходим поиск новых архитектурных и аппаратных решений. Исследования в этом направлении привели к появлению проектов и действующих прототипов машин баз данных — МБД (Database machine — DBM).

В общем случае *МБД принято называть аппаратно-программный мультимикропроцессорный комплекс, предназначенный для выполнения всех или некоторых функций СУБД.*

Такие свойства *реляционной модели данных* [6], как возможность расчленения отношений на непересекающиеся группы, возможность массовой и параллельной обработки, простота и независимость данных, а также наличие развитой теории и аппарата обусловили разработку МБД, ориентированных в основном на поддержку реляционных баз данных. В настоящее время очевидна правильность такого выбора в связи с установлением возможности опери-

ровать объектами баз знаний на реляционном концептуальном уровне посредством операций реляционной алгебры.

Первые публикации по МБД появились в 1974 г., в последующие годы было реализовано более 50 проектов, некоторые в виде промышленных прототипов и коммерческих изделий. Основными критериями для оценки того или иного проекта являются полнота выполняемых функций СУБД и ожидаемое повышение производительности при их выполнении.

Можно выделить два обобщенных направления, в которых ведутся исследования по структурным методам повышения производительности МБД: *многопроцессорные неоднородные (МН) и сетевые МБД*.

К многопроцессорным неоднородным можно отнести большинство проектов МБД, основными особенностями систем являются следующие.

1. Наличие нескольких уровней обработки данных, в частности, таких, как:

- селекция и первичная фильтрация данных непосредственно в контроллере дисковой памяти;
- вторичная обработка, заключающаяся в реализации операций реляционной алгебры над вспомогательными отношениями, полученными на первом этапе;
- структурная обработка или обработка метаданных, заключающаяся в поддержке вспомогательных структур (индексация, кластеризация).

2. Наличие системной буферной памяти (СБП) между первыми двумя уровнями обработки, где размещаются отношения или вспомогательные структуры данных, полученные на первом уровне обработки. Такая архитектура предполагает наличие опережающей выборки и подкачки данных из уровня первичной обработки (стадирование).

3. Наличие функционального параллелизма, при котором различные функции первичной и вторичной обработки реализуются на физически распределенной аппаратуре: универсальных микропроцессорах.

Наиболее типичными примерами таких МБД являются Delta и Grace. Действующий в настоящее время прототип состоит из следующих подсистем.

Основным функциональным узлом МБД Delta является реляционный процессор (РП) баз данных, назначение которого — выполнение операций реляционной алгебры над отношениями произвольного объема с высокой производительностью.

Каждый из четырех РП системы Delta может выполнять отдельную операцию реляционной алгебры независимо от других или параллельно (например, сортировку отношений).

Кроме этого РП в своем составе имеет центральный процессор (ЦП) с памятью для реализации операций с обширной логикой (например, агрегатных функций типа *min*, *max* и т. п.). Для облегчения обработки входного и выходного потоков данных РП содержит два адаптера иерархической памяти, а также входной модуль для подготовки кортежей отношений (например, перестановки значений атрибутов). Собственно операция реляционной алгебры реализуется в РП. Процессор слияния (ПСЛ) сортированных сегментов отношений предназначен для слияния сортированных сегментов отношений, а также соединения двух отношений или селекции. Двенадцать процессоров сортировки (ПСО) предназначены для реализации конвейерной однопроходовой сортировки сегмента отношения. ПСО и ПСЛ реализованы полностью аппаратно.

Иерархическая память в Delta является наиболее сложной подсистемой, в функции которой входят:

- стадирование данных (в виде сегментов отношений) в соответствии с заявками РП;
- селекция и вертикальная фильтрация отношений с привлечением специального (атрибутного) метода хранения отношений;
- поддержка индексных структур, кластеризация отношений и организация с их помощью быстрого поиска.

Как показали многочисленные исследования, СУБД не может быть эффективной, если большая часть ее работает под управлением операционной системы общего назначения, поэтому повышение эффективности МБД связано с полной изоляцией СУБД в рамках МБД, т. е. реализацией функционально полных МБД, выполняющих все функции управления транзакциями. Учитывая сложность соответствующей операционной системы МБД, реализовать функционально полную и высокопараллельную МН МБД сложно.

Вторая основная проблема в создании высокопараллельных МН МБД, названная «дисковым парадоксом», заключается в том, что скорость ввода-вывода современных УМП (одноканальные и многоканальные НМД с перемещающимися головками) является узким местом и ограничивает достижение высокого параллелизма в обработке. В МН МБД для решения этой проблемы в качестве кэш-диска применяется большая полупроводниковая буферная память.

Для решения этой проблемы предлагаются *сетевые МБД*, в которых распределенное хранение больших БД осуществляется на большем количестве НМД.

Сетевые МБД воплощают принципы однородности структуры, сегментации данных в устройствах массовой памяти (УМП) и распределения процессоров обработки. Таким образом, основная идея сетевых МБД — приближение дешевой обрабатывающей логики (в виде универсальных микропроцессоров) к УМП и связывание таких «обрабатывающих хранилищ» в сеть. Учитывая быстро снижающуюся стоимость процессоров и жестких НМД и успехи в технике коммуникации процессоров, в составе такой сети могут быть сотни УМП, с каждым из которых соединен свой обрабатывающий процессор.

Особый интерес приобретает создание сетевых МБД в связи с появлением серийных однокристалльных транспьютеров, содержащих наряду с процессором и памятью каналы (порты ввода-вывода). Например, промышленный транспьютер фирмы INMOS IMS T414 имеет следующие характеристики: в одном кристалле реализован 32-разрядный процессор; статическое ОЗУ; четыре канала связи; 32-разрядный интерфейс памяти и контроллер динамического ОЗУ.

Конструктивно транспьютерная матрица, являющаяся основным элементом транспьютерных МБД, может быть реализована посредством серийных транспьютерных плат этой же фирмы.

К классу сетевых относится коммерческая МБД фирмы Teradata DBC 1012, которая интенсивно распространяется и находит широкое применение в различных информационных системах. Конфигурация DBC 1012 содержит восемь обрабатывающих процессоров ПМД, каждый из которых имеет НМД и подключается к коммуникационной сети типа двоичного дерева (Y-сеть). В узлы этой сети встроены сетевые высокоскоростные процессоры и программируемые логические матрицы, реализующие функции управления сетью. Y-сеть позволяет осуществлять дуплексный обмен между обрабатывающими процессорами. К ней подключаются коммуникационные процессоры для осуществления интерфейса с главной ЭВМ.

Два свойства DBC 1012 характерны для всех сетевых МБД:

- 1) обеспечение возможности увеличения мощности наращиванием числа обрабатывающих процессоров, так что производительность при этом растет линейно (показатель линейности роста производительности DBC 1012 от числа процессоров составляет 97 %);
- 2) обеспечение надежности функционирования за счет дублирования данных в локальных УМП (т. е. обеспечение работы без краха системы при выходе из строя отдельных процессоров или УМП).

Третье направление исследований в области МБД заключается в создании недорогих коммерческих устройств на серийных процессорных элементах с шинным интерфейсом.

Примером является МБД фирмы Britton Lee IDM 500. Хотя эти изделия не ориентированы на высокопараллельную обработку и содержат ограниченное число функциональных процессоров, они удовлетворяют сформулированным ранее принципам МН МБД и полностью реализуют все основные функции МБД.

Роль СБП выполняет полупроводниковая память, к которой через общую шину подключаются периферийные контроллеры НМД со встроенными микропроцессорами, процессор обработки (процессор БД) и до восьми канальных процессоров для подключения к главной ЭВМ или к локальной сети (Ethernet). Кроме того, к общей шине может подключаться особый функциональный процессор (акселератор БД) для выполнения тех операций, которые являются узким местом (например, сортировка отношений).

Развитием этого направления в разработках фирмы Britton Lee является *реляционный файловый сервер (data file-server) RS310* — автономное устройство, подключаемое к локальной сети Ethernet или непосредственно к главной ЭВМ по интерфейсу RS232. Он включает:

- собственно процессор базы данных;
- соединенную с этим процессором оперативную память;
- два жестких диска типа винчестер с соответствующим контроллером;
- до четырех интерфейсных плат двух типов (интерфейс RS232 с восемью выходами или интерфейс локальной сети Ethernet).

Перспективы развития МБД. Создание высокопроизводительных МБД связывается с решением следующих проблем, по которым ведутся интенсивные исследования.

1. Создание специализированных архитектур МБД, сочетающих достоинства горизонтального параллелизма при осуществлении одной операции с функциональным параллелизмом при выполнении последовательности операций и транзакций. Особую роль здесь играет реализация конвейерной потоковой обработки (data flow) применительно к операциям реляционной алгебры.

2. Создание ассоциативной памяти большой емкости для системного буфера МБД. Как известно, ее использование в качестве СБП позволяет существенно повысить эффективность поисковых и некоторых других операций в МБД на уровне вторичной обработки. Интерес представляет гибридная ассоциативная память, которая имеет один порт — обычный для подключения памяти к общей шине, а другой — ассоциативный для подключения к соответствующему контроллеру.

3. Использование процессора потоковой сортировки отношений для их слияния. Например, в проекте Delta он предназначен для по-

токового слияния двух сортированных страниц отношения со скоростью поступления кортежей второй страницы. На базе такого двухвходового процессора слияния может быть реализовано устройство соединения двух отношений (Delta).

4. Аппаратная реализация потоковой фильтрации данных непосредственно в каналах УМП. Такая фильтрация позволяет снизить объемы данных, передаваемых из массовой памяти на обработку, что является существенным источником повышения производительности МБД в целом.

5. Устранение препятствия в увеличении производительности многопроцессорной МБД с двумя уровнями обработки и системной буферной памятью за счет улучшения системы коммутации и связи процессоров обработки с СБП и между собой. Это определяется интенсивностью обмена данными между процессорами и СБП и их объемом, которые, как правило, являются частями файлов (отношений БД). Кроме этого, организация параллельной работы процессоров требует интенсивного обмена сообщениями между ними.

Программное обеспечение

Программное обеспечение (ПО, software) является следующей компонентой обеспечения АИС и разделяется на три группы [6, 29, 30]:

- 1) операционные системы — operating system — (ОС);
- 2) системы программирования — application development system, software development system — (СП);
- 3) приложения (application), или пакеты прикладных программ (ППП) — software package.

В исторической последовательности развития программных средств первыми возникли узко ориентированные приложения («программа, предназначенная для вычисления числа π с точностью до 200-го знака», «программа, предназначенная для расчета и печати платежной ведомости» и пр.), затем — системы программирования (первые версии назывались системами автоматизации программирования), наконец появились операционные системы.

Операционная система предназначена для выполнения следующих основных (тесно взаимосвязанных) функций [29]:

- управление данными;
- управление задачами (заданиями, процессами);
- связь с человеком-оператором.

В различных ОС они реализуются в различных масштабах и с помощью разных технических, программных, информационных методов.

Ранние версии OS/360 были ориентированы на пакетную (batch processing) обработку информации — входной поток заданий (на МЛ, МД или перфокартах) подготавливался заранее и поступал в непрерывном режиме. В дальнейшем возникли расширения OS/360/375, допускающие диалоговую обработку данных с терминалов пользователя, последняя из версий (OS SVM) фактически предоставляла в распоряжение пользователя «виртуальную персональную ЭВМ» с полной мощностью вычислительной установки IBM/360/375.

ОС других семейств (поколений), например RSX (для PDP/11 DEC) или Unix, с самого начала ориентировались на интерактивное взаимодействие с пользователями. Относительно Unix следует отметить, что она в настоящее время является самой популярной ОС (если судить *не по количеству, а по качеству ЭВМ*, а также учесть, что MS DOS функционально является подмножеством Unix).

Широко распространены также системы, разработанные под влиянием концепций UNIX — MS DOS, Windows 95/NT, OS/2.

Система программирования представляет собой инструментальную среду программиста, которая обычно позволяет ему разрабатывать *прикладные программы (программировать приложения, разрабатывать приложения)*, а в более общем случае — и *системные компоненты ПО (ОС и СП)* для соответствующих ЭВМ и операционных систем. Это совокупность средств разработки программ (языки программирования — ЯП, текстовые редакторы, трансляторы, редакторы связей, библиотеки подпрограмм, утилиты и обслуживающие программы), обеспечивающих автоматизацию составления и отладки программ пользователя. Системы программирования классифицируются, как правило, на [6]:

- *одноязыковые/многоязыковые*. Отличительной особенностью последних является то, что отдельные части (секции, модули или сегменты) программы могут быть подготовлены на различных ЯП и объединены во время или перед выполнением в единый модуль;
- *замкнутые/открытые*. В последнем случае в систему можно ввести новый входной язык с транслятором, не требуя изменений в системе;
- *интерпретирующие/компилирующие*. В первом случае осуществляется покомандная расшифровка и выполнение инструкций входного языка (в среде данной системы программирования); во втором — подготовка результирующего модуля, который

может выполняться на ЭВМ практически независимо от программно-аппаратурной среды.

В системе OS/360 (а также RSX, UNIX) СП имеют *многоязыковую открытый характер*, являются системными компонентами, входящими в состав ОС (программы, написанные на языках Фортран, Си, Паскаль и пр., после компиляции и сборки без ограничений могут комбинироваться в сложные исполнительные модули и образовывать библиотеки процедур).

В менее развитых системах (типа DOS) СП являются внешним по отношению ОС слоем программных средств и потому разрабатываются и поставляются отдельно. Это приводит к появлению не вполне совместимых разработок, выполненных для одного и того же исходного ЯП (Turbo-C и Quick-C, например), каждая из которых является *замкнутой одноязыковой СП*.

Отдельно следует отметить *системы программирования АИС* (подобные ADABAS, Foxpro, Clipper и пр.), которые занимают промежуточное положение между СП и приложениями и будут рассмотрены ниже.

Приложения включают в себя программные продукты и оболочки систем. *Программным продуктом* является полностью или частично замкнутый программный комплекс, настроенный на решение одной или нескольких задач — комплексов информационных технологий, ориентированных на обработку (обычно — периодическую) стандартных входных форм и выпуск стандартных выходных документов.

Оболочки информационных систем (системы программирования ИС) представляют собой гибкие программные комплексы, настраиваемые на задачи пользователя. Наиболее распространенными классами данных программных средств являются *системы управления базами данных* и оболочки *автоматизированных информационно-поисковых систем* (АИПС).

АИПС (IRS — Information Retrieval System). В узком смысле под АИПС принято понимать открытый (обычно) или замкнутый (реже) программный продукт, предназначенный для реализации практически большинства функций (см. рис. 1.2) процессов — *ввода, обработки, хранения, поиска, представления данных* (организованных в записи или документы, находящиеся в БД). В этом смысле часто отождествляют АИПС с АИС, и это трудно оспаривать.

Среди АИПС в узком смысле принято выделять (см. гл. 2):

- *фактографические системы* (отличающиеся фиксированной структурой данных или записей), для разработки которых, как правило, используются СУБД, поддерживающие *табличные (реляционные) БД*;

- *документальные системы* (отличающиеся неопределенной или переменной структурой данных или документов), для разработки которых часто (но не обязательно) применяют оболочки АИПС.

В более широком смысле под АИПС подразумеваются также *программные оболочки, ориентированные на разработку продуктов типа АИПС* (в узком смысле). Это связано с тем фактом, что первые системы типа СУБД и оболочек АИПС были предложены в 60—70-е гг. фирмой IBM (и сотрудничавшими с ней организациями) и включали в себя:

- IMS/360 (Information Management System) — по-видимому, первую реальную СУБД, поддерживавшую *иерархическую модель данных* (понятие появилось позже, в связи с необходимостью систематизации СУБД), нашедшую достаточно широкое применение (в частности, для информационного обеспечения проекта Apollo, завершившегося, как известно, высадкой граждан США на Луну в 1969 г.);
- DPS/360 (Document Processing System) — первый промышленный ППП, предназначенный для реализации документальных АИПС. В дальнейшем путем развития принципов DPS фирмой в 1972 г. был выпущен пакет STAIRS (STorage And Information Retrieval System), предназначенный для диалогового обслуживания множества (удаленных) пользователей;
- IRMS (Information Retrieval and Management System), TEXT-PAC и другие аналогичные пакеты.

Как следует из наименований продуктов, разработчики понимали под АИПС именно ППП-оболочки.

Системы управления базами данных и программирования АИС. Среди различных программных средств данного класса следует различать [5, 6, 30]:

- СУБД в «чистом виде» (IMS, SETOP и пр.);
- СУБД с элементами систем программирования АИС (ADABAS/Natural, реже ORACLE);
- системы программирования АИС с элементами СУБД (Foxpro, Clipper).

Первый тип фактически относится и к начальному этапу развития систем второго (реже — третьего) типов.

В этом случае СУБД состоит только из *системы интерпретации вызовов (обращений)* из пользовательской программы (call-interface) на выборку (корректировку, занесение) информации из (в) БД, причем программа написана на одном из универсальных ЯП (Кобол, Фортран, Паскаль и пр.), получивших название *включающих языков*

СУБД. Данная система в последующих СУБД (второй тип) получила наименование *ядра*.

Соглашения о форматах и структурах такого взаимодействия обычно пытаются оформить в виде некоторого формального *языка (языка ядра)*. В частности, вдохновленная успехами в разработке и распространении универсального ЯП PL/I (Programming Language #1), фирма IBM разработала описание форматов интерфейса пользовательских программ с БД IMS в форме языка DL/I (Data Language #1), который однако значительного успеха не имел.

Второй тип представляет собой расширение первого в направлении создания универсальной системы разработчика АИС, включающей также специализированные языковые средства. В этом случае СУБД представляет собой совокупность:

- специализированных программных средств;
- вспомогательных файлов;
- управляющих таблиц (иногда находящихся в составе БД, реже — это файлы ОС).

Система обеспечивает доступ пользователей к БД при соблюдении следующих существенных критериев:

- целостность и непротиворечивость данных, описывающих различные аспекты объектов реального мира;
- защита информации от несанкционированного доступа на чтение/обновление содержимого БД;
- установление и поддержание связей между зависимыми данными, удобство их использования.

Третий тип представляют собой (разработанные обычно для ПК) системы, содержащие как элементы непроцедурного типа (язык запросов), так и процедурного (язык программирования) во входном языке, предназначенном для управления данными и обработки информации.

Элементы СУБД здесь также заключаются в наличии:

- простейшего словаря данных;
- возможностей создания модели предметной области в форме совокупности таблиц, связанных между собой простейшим образом;
- средств генерации отчетов и управления доступом пользователей.

В структурном составе СУБД в двух последних случаях могут быть выделены *ядро* и *среда* (рис. 1.9) [5].

Ядро СУБД — программный комплекс (модуль или модули), обеспечивающий непосредственное выполнение физических опера-

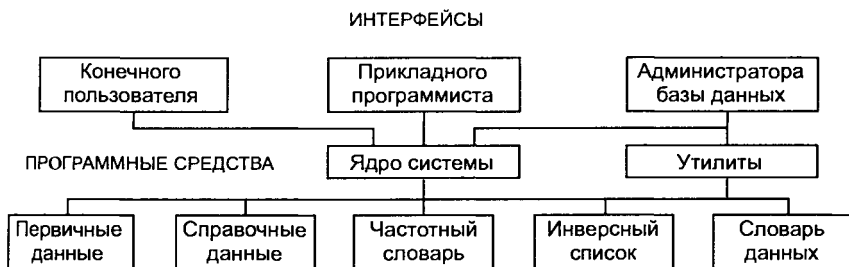


Рис. 1.9. Структура систем управления базами данных

ций над БД (в ранних системах функции Ядра выполняли *программы методов доступа ОС ЭВМ*).

Среда — совокупность интерфейсных модулей, обеспечивающих связь пользователей с Ядром и через него с БД. Среда включает в себя *пользовательские интерфейсы* и *утилиты администратора БД* (АБД).

Утилиты АБД образуют библиотеку программ обслуживания БД в привилегированном режиме (работа пользовательских средств параллельно утилитам не разрешена) и выполняют основные функции, к которым относятся:

- физическая подготовка дисковой памяти к размещению БД;
- подготовка справок о составе БД, структуре файлов, количестве данных и занимаемом объеме;
- загрузка файла БД из последовательного набора данных ОС;
- дозагрузка (расширение существующего файла);
- модификация БД: расширение или перемещение физических наборов данных, реорганизация;
- модификация файла: добавление новых полей в структуру записи; инвертирование полей или освобождение (превращение инвертированных полей в сканируемые);
- выгрузка образа БД для сохранения в архивном наборе данных;
- создание и ведение словаря данных и др.

Средства пользователя. Стандартными средствами этого типа, предоставляемыми фирмой-разработчиком, являются следующие:

- диалоговые интерфейсы;
- генераторы отчетов;
- система конструирования и поддержки интерактивных технологий в информационных системах (ЯП АИС).

1.6. Состав и структура информационного обеспечения АИС

Информационное обеспечение АИС включает в себя две компоненты:

1. *Лингвистическое обеспечение (ЛО)*, к которому относятся:

- *форматная база (типы, форматы, структуры информации* — данных, записей, документов);
- *лексическая база (классификаторы, кодификаторы, словари и тезаурусы* или иные лексико-лингвистические таблицы, используемые при вводе, обработке, поиске, представлении информации (данных, записей, документов);
- *языковые средства* описания (ЯОД, словари данных) и манипулирования (ЯМД) данными.

2. *Собственно информационное обеспечение составляют:*

- *файлы* операционной системы (именованные совокупности данных, находящиеся под управлением ОС в процессе реализации ею функции *управления данными* (см. выше));
- *база (базы) данных* — именованная взаимосвязанная совокупность физических файлов ОС ЭВМ, поддерживающая информационную модель предметной области.

Рассмотрим вначале элементы *форматной базы*.

Типы, структуры, форматы данных и документов в информационных системах

Типы данных — это совокупность соглашений о программно-аппаратурной форме представления и обработки, а также ввода, контроля и вывода элементарных данных.

Структуры данных — способы композиции простых данных в агрегаты и операции над ними.

Форматы файлов — представление информации на уровне взаимодействия операционной системы с прикладными программами.

Форматы данных — соглашения о представлении агрегатов информации при передаче (поэтому часто говорят о коммуникативных или обменных форматах), в том числе:

- библиотечных и справочных БД;
- полнотекстовых документов.

Типы данных (табл. 1.11). Ранние ЯП, а точнее, системы программирования (СП) — Фортран, Алгол, будучи ориентированными исключительно на вычисления, не поддерживали развитых систем типов и структур данных.

Таблица 1.11. Типы и структуры данных в некоторых системах программирования и управления данными

Данные		Система – язык программирования, СУБД, ИПС						
		Algol	Cobol	PL/1	FoxBase/ Clipper	Adabas/ Natural	Oracle/ SQL	STAIRS, IRBIS, ISIS
Тип данных	Целое короткое (2 байта)	–	–	–	–	–	Small-int	–
	Целое норм. (4 байта)	Integer	Compu- tational	Int	N(x)	N(x)	Int	–
	Целое длинн. (8 байт)	–	–	Double	–	–	–	–
	Действительное норм. (4 байта)	Real	Compu- tational	Float	N(x.y)	N(x.y)	Float Real	–
	Действ. двойное (8 байт)	–	–	–	–	–	Float Double	–
	Двоичное	–	–	Binary	–	B(x)	–	–
	Десятичное упакованное (2 цифры на байт)	–	PIC(9)	Decimal	–	P(x)	–	–
	Десятичное распакованное (1 цифра на байт)	–	PIC(X)	–	N(x)	U(x)	–	–
	Логическое	Boolean	–	+	Logical	–	–	–
	Символьное	–	PIC(A)	Char	C(x)	A(x)	Char	+
	Длинный текстовый или бинарный объект (BLOB)	–	–	–	Memo	–	VarGrafic VarChar	–
	Дата	–	–	–	Date	–	Date	–
Время	–	–	–	–	–	Time	–	
Структуры данных	Массивы	Array	–	Dim	Dimention	VAR(n)	–	–
	Записи (структуры)	–	+	+	+	+	+	–
	Множественные (векторные) поля записи	–	–	–	–	MU	–	+
	Групповые поля записи	–	+	+	–	GR	–	+
	Повторяющиеся группы в записи	–	–	–	–	PE	–	–
	Текстовые поля (параграфы, предложения, слова)	–	–	–	–	–	–	+

В Алголе символьные величины и переменные вообще не предусматривались, в некоторых реализациях строки (символы в апострофах) могли встречаться только в операторах печати данных.

Типы числовых данных Алгола: INTEGER (целое число), REAL (действительное) — различаются диапазонами изменения, внутренними представлениями и применяемыми командами процессора ЭВМ (соответственно арифметика с фиксированной и плавающей точкой). Нечисловые данные представлены типом BOOLEAN — логические, имеющие диапазон значений {TRUE, FALSE}.

Позже появившиеся ЯП (СП) Кобол, ПЛ/1, Паскаль предусматривают новые типы данных:

- символьные (цифры, буквы, знаки препинания и пр.);
- числовые символьные для вывода;
- числовые двоичные для вычислений;
- числовые десятичные (цифры 0—9) для вывода и вычислений.

Разновидности числовых данных здесь соответствуют внутреннему представлению и машинным (или эмулируемым) командам обработки. Кроме того, вводятся числа двойного формата (два машинных слова), для обработки которых также необходимо наличие в процессоре (или эмуляция) команд обработки чисел двойной длины (точности).

Уместно привести пример представления числовой информации в различных перечисленных формах. Пусть задано число $135_{10} = 207_8 = 87_{16} = 100000111_2$:

- внутренняя стандартная форма — тип BINARY — представления для обработки в двоичной арифметике сохраняется (100000111_2). Объем — 1 байт, 8 двоичных разрядов;
- внутренняя форма двоично-десятичного — тип DECIMAL — представления — каждый разряд десятичного числа представляется двоично-десятичной (4 бита) комбинацией. Представление 135 есть 001011101_2 . Объем — 2,5 байта, 12 двоичных разрядов;
- символьное представление (для вывода) — тип ALPHABETIC — каждый разряд представляется байтом в соответствии с кодом ASCII.

Представление 135 есть — $00110001\ 00110011\ 00110101_2$. Объем — 3 байта.

Некоторые системы программирования (Фортран IV, например), поддерживают операции над *комплексными числами* вида $Z = A + Bi$ (где A, B — действительные коэффициенты, а i — мнимая единица). Очевидно, для размещения таких чисел необходим как минимум двойной расход оперативной памяти (по одному слову для

размещения действительной и мнимой частей при обычной и по два слова при двойной точности). Кроме того, очевидно, что процессоры обычных универсальных ЭВМ вряд ли поддерживают операции над такими числами, в связи с чем операции над ними требуют написания соответствующих подпрограмм или эмуляции комплексной арифметики.

Появление СУБД и СП для разработки ИС приводит к появлению новых типов данных:

- *дата и время*;
- *бинарные* (BLOB — Binary Large Object) и *текстовые объекты* без внутренней структуры (интерпретация возлагается на прикладные программы).

Понятие типа данных ассоциируется также с допустимыми значениями переменной и операциями над ними, например, данные типа *время* (ЧЧ:ММ:СС) или *дата* (ГГ/ММ/ДД) предполагают определенные диапазоны значений каждого из разрядов, а также машинные или эмулируемые операции (сложение/вычитание дат и/или моментов времени). Основной причиной «проблемы 2000 г.» являлось не столько двухразрядная запись года в базах данных, сколько встроенные в огромное количество программ (часто не документированных) операции над данными типа DATE — ГГ/ММ/ДД.

Структуры данных. В языке Алгол определены два типа структур: *элементарные данные и массивы* (векторы, матрицы, тензоры, состоящие из арифметических или логических переменных). Основным нововведением, появившимся первоначально в Коболе, (затем ПЛ/1, Паскаль и пр.) являются *агрегаты данных* (структуры, записи), представляющие собой именованные комплексы переменных разного типа, описывающих некоторый объект или образующих некоторый достаточно сложный документ.

Рассмотренные выше экзотические типы данных (комплексные числа) очевидно занимают промежуточное положение между элементарными переменными и массивами (структурами).

Термин *запись* подразумевает наличие множества аналогичных по структуре агрегатов, образующих *файл* (картотеку), содержащих данные по совокупности однородных объектов, элементы данных образуют поля, среди которых выделяются элементарные и групповые (агрегатные).

С появлением СУБД и АИПС возникают новые разновидности структур:

- множественные поля данных;
- периодические групповые поля;

- текстовые объекты (документы), имеющие иерархическую структуру (документ, сегмент, предложение, слово).

В ранее рассмотренных БД — INIS, TNT, CHEMNAME, D&B (см. рис. 1.4—1.7) — легко могут быть идентифицированы следующие виды полей данных:

- элементарные: RN, YEAR, VVSS (INIS); DK, GI (TNT); AA, NA, RN (CHEMNAME); EH, SQ (D&B);
- периодические группы: [PO, NA], [SA, ET] (D&B);
- текстовые поля: ABSTR, DESCR, TITLE (INIS); TX, RF (CHEMNAME), TX (TNT);
- числовые поля: SQ, SA, ET, EN (D&B);
- символьные поля, которые составляют большинство в приведенных примерах БД.

Форматы файлов. Понятие файла появляется впервые в операционной системе OS/360 фирмы IBM, причем в ранних версиях системы «настоящим файлом» считался только перфокарточный массив (file = картотека), данные на МД и МЛ обозначались как DS (Data Set — набор данных). В последующих ОС (RSX, UNIX, MS DOS) файлами становятся именованные организованные наборы данных на любых носителях и устройствах, за сохранность и обновляемость которых (а также передачу в прикладные программы и обратно) несет ответственность ОС ЭВМ.

В зависимости от типа и назначения файлов и возможностей ОС (методов доступа) файл может передаваться в прикладную программу как целое или блоками (физическими записями) и логическими записями (строками, словами, символами).

В системе OS/360 основную роль играли два типа файлов:

- символьные (исходные программы или данные);
- двоичные (программы в машинных кодах).

В современных системах активно используется значительно большее разнообразие файлов, из которых мы перечислим наиболее типичные (табл. 1.12):

- *текстовые файлы* — обобщенное название для простых и размеченных текстов, ASCII-файлов и других наборов данных символьной информации, которые интерпретируются и обрабатываются текстовыми редакторами, процессорами, анализаторами (Lexicon, Word, TEC, анализаторы SGML, HTML);
- *текст без разметки* (планарный) — файл, содержащий только отображаемые (воспроизводимые на всех печатающих устройствах и терминалах) символы кода ASCII, а также простейшие управляющие символы — CR — возврат каретки; LF — пере-

Таблица 1.12. Основные типы файлов, обрабатываемых в ПЭВМ

Тип, расширение имени	Вид информации, содержащейся в файле
exe, com	Программа, готовая к исполнению
bat	Текстовый командный файл
sys	Системный файл
ovi, ovr	Оверлейный файл
pif	Программно-информационный файл Windows
txt, lst	Текстовый файл в формате DOS
doc	Документ (чаще всего в формате WinWord)
rtf	Размеченный текстовый файл (Rich Text Format)
dot	Файл формата документа (Document Type)
pdf	Формат документа Adobe Acrobat
wri	Документ редактора Write для Windows
wps	Документ текстового процессора MS WORKS
bak, old	Старая копия файла, создаваемая перед его изменением
arj, rar, zip, lzh, arc	Архивные файлы
bas	Текст программы на языке Basic
pas	Текст программы на языке Turbo Pascal
c	Текст программы на ЯП Си
bmp, pcx, gif, tif, jpg	Графические файлы
dbf	Файлы базы данных формата DBase, Foxpro, Cliper
wdb	Базы данных формата MS WORKS
wks	Электронная таблица формата MS WORKS
xls	Электронные таблицы EXCEL
lib, dll	Файлы библиотек
dat	Файл данных
ini	Файл инициализации
hlp	Файл справки (подсказки, помощи)
ext	Файл расширений
mnu	Файл меню
wav, mid, mp3, mod	Звуковые файлы
avi, mov, mpg	Файлы видеоклипов

вод строки; TAB — символ табуляции, иногда LF — новая страница;

- *текст с разметкой* — планарный файл, содержащий *бинарную* (см. табл. 1.12, колонки 1, 2) или *символьную* (остальные колонки) разметку, управляющую отображением информации (программно и/или аппаратно);
- *ASCII-файл* — содержит только отображаемые коды левой части кодовой таблицы ASCII (латиница и служебные символы), обычно применяется для хранения документов с символьной разметкой (RTF, SGML, HTML);
- *табличный файл* — содержит форматированные данные (символьные, численные и др), образующие строки и столбцы таблиц, создаваемых и обрабатываемых табличными СУБД (Foxpro, Clipper, MS Access), и/или процессорами (SuperCalc, MS Excell и др.);
- *графический файл* — бинарный файл, содержащий графическую информацию. Форматы: tif (Tagged Image File), bmp (Bit-Mapped Picture), а также ряд других — psx, pic и т. д.;
- *мультимедиа файл* — бинарный, содержащий оцифрованную аудио- (типы wav или MIDI-Sequencer) видео- (формат MPEG) или смешанную информацию.

Проблема обмена информацией и коммуникативные (обменные) форматы

Данные форматы связаны с проблемой передачи информации между системами с различной организацией и структурой данных. Даже при простейшей ситуации (считывании записей из внешнего файла при загрузке информации в БД) возникают две проблемы:

- идентификации данных;
- локализации описания данных.

Проблема идентификации заключается в следующем:

- необходимость правильного распознавания и «привязывания» *данных, размещенных на внешнем носителе* (файл) к тем *областям памяти*, которые выделены для размещения этих данных (им соответствуют какие-то *имена переменных* в обрабатываемой программе);
- обнаружение ошибок при считывании данных (например, несоответствие типа или длины данного ожидаемому и т. п.);
- пропуск ошибочных данных (записей, строк и пр.) или вывод их в специальные *файлы ошибок*.

При считывании информации из файла (эта функция может быть возложена как на операционную систему, так и на пользовательскую программу или библиотечную процедуру) необходимо уметь:

- определять начало и окончание элементарного данного внутри записи;
- определять начало и окончание записи файла.

Здесь необходимо отдельно рассмотреть записи постоянной и переменной (неопределенной) длины. Выделяют следующие методы записи.

1. Ввод, управляемый редактированием (GET EDIT в ЯП ПЛ/1). В этом случае данные на носителе (в файле) должны иметь строго ту длину, которая задана в их описании (в прикладной программе).

Это ограничение, очевидно, имеет смысл только для файлов с записями постоянной длины. При этом символьные данные (поля) должны быть дополнены до стандарта *хвостовыми пробелами (trailing blanks)*, а числовые — *ведущими нулями (leading zeros)*. Для записей фиксированной длины, состоящих из элементов постоянной или ограниченной длины в буфере считывания выделяется область, равная общей длине записи. Всякое нарушение длины и типа приводит к ошибке считывания и выбраковке записи.

2. Ввод, управляемый списком (GET LIST в ЯП ПЛ/1). Этот метод называется также «с разделителями». При этом записи должны быть отделены друг от друга *разделителями (ограничителями) записей (record terminators, delimiters)*, а элементы данных внутри записи — *разделителями данных (data terminators)*. Этот подход действителен как для записей постоянной, так и фиксированной длины.

3. Ввод, управляемый данными (GET DATA в ЯП ПЛ/1). Здесь каждое данное в файле снабжается идентификатором, или меткой, которая совпадает с именем элемента данных в программе. Это способствует «точному приземлению» указанной информации в отведенную память. Также подходит для всех типов записей.

Пусть структура (запись) в некоторой программе имеет следующее описание

TOWN CHAR(20),	*наименование города
PEOPLE NUM(8),	*население
YEAR_F NUM(4);	*год основания

Тогда при первом методе записи информация в файле должна выглядеть так:

МОСКВА _____ 080000001147
(после «Москвы» — 13 «пробелов»);

при втором:

```
#МОСКВА#8000000#1147#$
```

(если разделитель данных — #, а записей — \$);

третий метод допускает следующую форму записи:

```
TOWN='МОСКВА', YEAR_F='1147', PEOPLE='8000000'$ (порядок данных здесь может быть произвольным, а разделитель записей все же нужен — $, а также ограничитель данного — ').
```

Мы предлагаем читателям самостоятельно сформулировать, каковы достоинства и недостатки перечисленных методов записи.

Кроме перечисленных возможны также и другие методы, например, первый байт каждой записи может содержать длину всей записи, а первый байт каждого элементарного данного — его длину.

Записи неопределенной длины возникают тогда, когда ограничителем является физическая метка, распознаваемая устройством.

Проблема локализации описания данных. Рассмотренные выше приемы распознавания программой элементов данных или записей относятся к такому типу взаимодействия, когда описание данных размещено в программе, а файл организован в соответствии с ним. Однако этот способ может привести к нарушению функционирования или разрушению данных, если из-за ошибок программиста или оператора к программе будет подсоединен «неправильный файл».

Для установления независимости программ от данных в некоторых системах и ситуациях описание данных размещают совместно с файлом данных. По такому принципу организован весьма распространенный *формат файла данных* (dbf-формат), происходящий от систем dBase — Clipper — Foxbase — Foxpro, а затем принятый и рядом других систем. В этом случае в начале файла создается заголовок, содержащий описание полей записи (имя, тип, длина данного, код информации и пр.). В этом случае описание данных файла (внешних) в программе не требуется (см. [6]).

Недостатком данного подхода является, например, необходимость использования программистами тех же имен данных, что содержатся в описании файла.

Следующим шагом явилось полное отделение описаний от данных и программ и сосредоточение их в специальных файлах (таблицах) — *словарях данных*, которые относятся к *базам данных и системам управления базами данных*.

В рассматриваемых ниже примерах коммуникативных форматов проблемы *идентификации и локализации* в той или иной степени (по-разному) решены. Предлагаем читателям самостоятельно опи-

сать особенности этих решений в каждом случае (МЕКОФ, карточный формат, SGML).

Типы коммуникативных форматов

- Коммуникативные форматы разделяются на следующие категории:
- форматы обмена *библиографической* информацией (МЕКОФ, карточный формат и пр.);
 - форматы обмена *библиотечно-справочной* информацией (семейство MARC);
 - форматы обмена *полнотекстовой* документальной информацией (языки процедурной и описательной разметки документов, средства моделирования документов) — RTF, ODA, SGML, HTML.

Форматы первой группы удовлетворяют *спецификациям ISO 2709* — рекомендациям Международной организации по стандартизации (ISO) и различаются наполнением и применением — МЕКОФ предназначен для распространения библиографических БД, а MARC — для онлайн-овых библиотечных каталогов (OPAC).

Карточный формат (рис. 1.10) представляет каждый документ (запись) БД как совокупность строк с одинаковым номером (RN), в каждой из которых указана метка (три цифры), идентифицирующая поле или группу полей.

RN:	INIAE8505504	000	INIAE85005504 000339
VI:	INIAE8505504	001	ОТЧЕТ
TI:	INIAE8505504	100	SELF-CONSISTENT, TIME DEPENDENT, HARTREE-FOCK
	INIAE8505504	100	CALCULATION FOR A REACTIVE MOLECULAR COLLISION
AU:	INIAE8505504	200	TISZAUER, D. J. KULANDER, K. C.
CO:	INIAE8505504	320	LAWRENCE LIVERMORE NATIONAL LABORATORY,
	INIAE8505504	320	LIVERMORE (CALIFORNIA)
DA:	INIAE8505504	330	1984 MAY
IS:	INIAE8505504	350	ISSN 0056-2971
AB:	INIAE8505504	400	WE REPORT THE FIRST CALCULATION ON A MOLECULAR
	INIAE8505504	400	SYSTEM USING A FULLY SELF CONSISTENT TIMEDEPENDENT
	INIAE8505504	400	HARTREE-FOCK (TDHF) METHOD. THE NUCLEAR MOTION IS

DE:	INIAE8505504	500	ACTIVATION ENERGY; HARTREE-FOCK METHOD; HYDROGEN;
	INIAE8505504	500	HYDROGEN IONS 1 PLUS; ION-MOLECULEE COLLISION;
	INIAE8505504	500	REACTION KINETICS; TIME DEPENDENCE; TRAJECTORIES;
	INIAE8505504	500	WAVE FUNCTIONS
CL:	INIAE8505504	600	A12
CY:	INIAE8505504	700	US 85 1523 AM

Рис. 1.10. Пример документа БД INIS, представленного в карточном формате: VI — вид документа; TI — заглавие; AU — автор; CO — корпоративный автор; AB — реферат; CL — рубрика INIS; CY — страна публикации

Формат МЕКОФ. Стандарт для коммуникативного формата МЕКОФ устанавливает структуру записи и ее наполнения для обмена библиографическими или другими данными на машиночитаемых носителях (рис. 1.11).

Содержание позиций		Позиции символов	Наполнение
Маркер	-	0-4	Длина записи
		5	Статус записи
		6-9	Коды применения
		10	Длина индикатора
		11	Длина идентификатора
		12-16	Базовый адрес данных
		17	Набор кодов
	18, 19	Зарезервировано	
	План справочника	20	Длина «длины поля данных»
		21	Длина «позиции начального символа»
22		Длина «части, определяемой при применении»	
23		Зарезервировано	
Справочник	Статья	-	Метка Длина поля данных Позиция начального символа Часть, определяемая при применении
		-	Статья ... Статья Разделитель поля
Поля данных	-	-	Идентификатор записи Разделитель поля записи
		Резервные поля	Резервное поле Разделитель поля Резервное поле Разделитель поля
	Библиографические поля	-	Данные Разделитель поля
	-	-	Разделитель записи
	-	-	Следующая запись

Рис 1.11. Структура записи формата МЕКОФ

В отличие от карточного формата, для интерпретации которого необходимо описание данных в программе загрузки или в таблицах словаря данных, МЕКОФ содержит максимально полное описание в собственно записи. Оно находится в *маркере записи* и *справочнике*, причем маркер описывает справочник, а справочник — собственно

данные, чем достигается высокая степень независимости данных от программ обработки.

Библиографическая запись — совокупность данных, включающая: маркер данных; справочник; поля данных; разделитель записи.

Маркер — структурный компонент, находящийся в начале каждой записи, содержащий параметры для ее обработки и состоящий из следующих компонентов: *длина записи, статус записи, коды применения, длина идентификатора, длина индикатора* (количество символов индикатора); *базовый адрес данных, набор кодов, план справочника*.

Справочник — указатель местонахождения полей данных в записи. Состоит из переменного числа статей, каждая из которых идентифицирует соответствующее поле данных и содержит следующие компоненты — *метка; длина поля данных, позиция начального символа поля*.

Поле данных (поле) — часть библиографической записи, имеющая переменную длину, следующая за справочником и связанная с одной из его статей; может содержать одно или несколько *подполей* и должно заканчиваться разделителем.

План справочника — набор параметров, определяющих структуру статей справочника.

Идентификатор — элемент, состоящий из одного или нескольких символов, непосредственно предшествующий подполю и идентифицирующий его.

Индикатор — элемент, расположенный в начале поля и несущий дополнительную информацию о содержании поля, взаимосвязи с другими полями или об операциях, требуемых при определенной обработке данных.

Метка — три символа, связанные с полем и идентифицирующие его.

Разделитель — управляющий символ, разделяющий данные.

Форматы серии MARC имеют иерархическую структуру, которая поддерживает отношения между компонентами внутри записи и позволяет комбинировать связанные записи.

Аналогичное назначение и структуру имеет *Единый Коммуникативный Формат (CCF) ЮНЕСКО*, содержащий три компонента:

- правила структуризации среды хранения (включая физические размеры элементов среды — длина блоков и т. п.);
- идентификационные коды элементов данных (поля, подполя, метки, разделители);
- правила представления содержимого элементов данных (словари и алгоритмы индексирования и кодирования информации).

Форматы полнотекстовых документов. Понятие *модель документа* охватывает аспекты создания, преобразования, хранения, поиска, передачи и отображения документов. Принято рассматривать структуру документа в двух аспектах: логическом (содержание) и физическом (макетном) — представление [30].

Логическая структура определяет составные компоненты и их соотношения в понятиях, отвечающих взгляду на документы как смысловые структуры. Например, к основным смысловым компонентам относятся: авторские данные (имя автора, место работы), аннотация, оглавление, главы, разделы, параграфы, рисунки, сноски.

На рис. 1.12 приведен пример структуры документа «Пояснительная записка к дипломному проекту (работе)». Здесь выделены такие базовые понятия структуры, как *обязательность/необязательность*

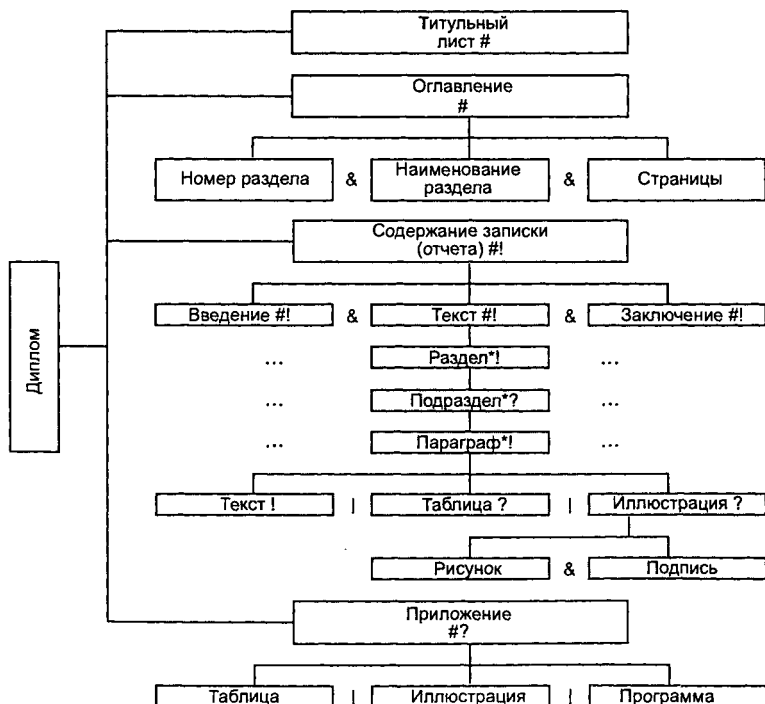


Рис. 1.12. Иерархическая структура документа «Пояснительная записка к дипломной работе»:

— уникальный элемент; * — повторяющийся элемент; ? — необязательный элемент; ! — обязательный элемент; & — вхождение типа И; | — вхождение типа ИЛИ

ность элемента, *уникальность* или повторяемость, вхождение нижестоящих элементов в вышестоящие по принципу «И» (оба типа данных должны или могут входить в элемент) либо «ИЛИ» (только какой-либо *один* из типов данных может или должен входить в элемент).

Макетная структура содержит описание документа в терминах физических единиц — страниц, полос, колонок, колонтитулов, рамок для рисунков, шрифтов, стилей и пр.

Подходы к моделированию документов опираются на два стандарта:

- ISO 8613 (*ODA Office Document Architecture — архитектура управленческой документации*);
- ISO 8879 (*SGML — Standard Generalized Markup Language — стандартный обобщенный язык разметки*).

Документ в ОДА представлен в виде профиля и собственно документа, организованных в форме древовидной структуры.

Профиль содержит информацию о документе в целом и его прохождении, формальные признаки — дата составления, вид, регистрационный номер и т. д.

Собственно документ содержит текст и сведения о его структуре и стиле, а именно:

- структура документа — заглавие, параграфы, оглавление и т. п. (логическая структура), а также абзацы, расположение текста, шрифты (физическая структура);
- архитектура содержания — набор графических элементов, выделение определенных слов, строк и т. п.;
- коммуникативный формат — способы кодирования объектов, признаков и содержания документов.

Язык SGML разработан на базе программного продукта DCF GML фирмы IBM и представляет собой метод создания структурированных документов, а также языков для их разметки.

В системах обработки текстов в документ включается дополнительная информация, называемая *разметкой* и выполняющая следующие функции:

- выделение логических элементов данного документа;
- задание функций обработки выделенных элементов.

В обычных текстовых процессорах существуют встроенные команды включения/выключения шрифтов и другие, аналогичные командам управления размещением информации на экране или при печати (*Escape-последовательности*). Такой подход называется командной или *процедурной разметкой* (табл. 1.13).

Таблица 1.13. Примеры разметки текстовых файлов (управление представлением)

Вид разметки	Принтер EPSON	Редактор Лексикон	Формат rtf (Rich Text Format), в т. ч. Word	Формат электронной почты (стандарт MIME)	html (Hypertext Markup Language)
Полужирный шрифт	ESC G... ESC H	chr(255)2... chr(255)0	{\b...}	<bold>...</bold>	...
Курсив (италик)	ESC 4... ESC 5	chr(255)1... chr(255)0	{\i...}	<italic>...</italic>	<i>...</i>
Подчеркивание	ESC - 1... ESC - 0	_chr(255)... chr(0)	{u\...}	<underline>...</underline>	<u>...</u>
Индекс верхний	ESC S 0... ESC T	chr(255)5... chr(255)0	{\super...}	<superscript>...</superscript>	^{...}
Индекс нижний	ESC S 1... ESC T	chr(255)4... chr(255)0	{\sub...}	<subscript>...</subscript>	_{...}
Перевод строки	chr(12)	.chr(12)	\page	np	
Выравнивание	-	<Shift><F8>	\qc – по центру \ql – влево \qr – вправо	<Center> <FlushLeft> <FlushRight>	<align =center> =left> =right> =justify>
Абзац	Табуляция (TAB, chr(9))	TAB	\par	Paragraph	<p>

Примечание. Аналогично могут быть рассмотрены — управление размером и формой шрифта, вставка математических символов и пр.

Альтернативный способ разметки заключается в выделении части текста без указания способа обработки, который затем назначают другие команды. Такая разметка называется *описательной* (дескриптивной). Она включает метки (*tags, tagu*) начала и окончания элемента текста и указывает, как интерпретировать данный фрагмент. Изменяя набор процедур, сопоставленный описательной разметке, можно изменить внешнее представление одного и того же документа. Развитие идей описательной разметки привело к определению разметки как формального языка, что позволяет проверить правильность разметки и минимизировать разметки за счет подстановки умолчаний.

В языке SGML каждый документ включает:

- *декларации языка SGML*, привязывающие к определенным значениям параметры обработки, а также имена синтаксиса;
- *пролог*, состоящий из деклараций о типе документа. Они определяют типы элементов, их взаимосвязи и атрибуты, а также условные обозначения, которые могут быть задействованы при разметке;

- *данные*, которые состоят из разметки документа и собственно информации.

Основные типы конструкций языка — описания *элементов* `<!ELEMENT...>`, *объектов* `<!ENTITY...>` и *атрибутов* `<!ATTRIBUTE LIST...>`, образующих структуру документа (документов), при этом элемент является основной его компонентой, объект — группа, род элементов, атрибут — характеристика элемента. Все «квадратики», приведенные на рис. 1.13, являются элементами. Запишем одну из возможных конструкций, соответствующую выделенной на рис. 1.13 цепочке элементов («Подраздел» — «параграф» — «текст»...):

<code><!ELEMENT SUBDIV (PAR*)></code>	подраздел состоит из параграфов (повт.)
<code><!ELEMENT PAR (TEXT TABLE? PICT?)></code>	параграф - из текста или таблицы (необяз.) или рисунка (необяз.);
<code><!ELEMENT PICT (IMAGE & CAPT)></code>	рисунок - из изображения и подписи

Декларации и пролог на языке SGML задают структуру документа и, будучи отделенными от размеченного текста, образуют описание типа документа (*DTD — Document Type Definition*). На сегодня известно более 5000 DTD, соответствующих различным национальным и международным стандартам, из которых наиболее важным является HTML.

Лексическое обеспечение

К данной компоненте лингвистического обеспечения в первую очередь относятся *кодификаторы*, *классификаторы*, *тезаурусы*.

Кодификаторы АИС представляют собой словари, не обязательно несущие семантическое соответствие между обозначаемым (класс, понятие, сущность) и обозначающим (код, символ). Например, в персональной БД атрибут SEX (Пол) может обозначаться либо М/Ж, либо 1/0 или 0/1, либо еще как-то в рамках двухсимвольного алфавита (домена). Это соответствие необязательно и может произвольно выбираться разработчиками различных систем. Основанием для использования кодификаторов (впрочем как и других типов нормативных словарей) являются: экономия времени для заполнения форм и ввода документов; минимизация ошибок; экономия памяти.

Классификаторы АИС являются словарями, обязательно несущими определенную смысловую нагрузку. Они подразумевают иерархическое разбиение предметной области на совокупность сужающихся классов и последующее отнесение обозначаемого объекта реального мира (документ, продукт и пр.) к одному из классов (индексирование).

Иерархические классификаторы могут быть разделены на два типа:

- с фиксированным числом уровней;
- с неопределенным числом уровней.

Ниже приведены примеры некоторых классификаторов фиксированной глубины (числа уровней иерархии), используемых как в фактографических, так и в документальных АИС, а именно:

- гармонизированной системы (рис. 1.13);
- классификатор службы INIS (рис. 1.14);
- ОКП (рис. 1.15).
- международная классификация изобретений (рис. 1.16, а).

Раздел	Подраздел	Содержание рубрики
2200.00	Напитки алкогольные и безалкогольные, уксус	
	2201.00	Воды, включая минеральные, натуральные или искусственные, газированные, без добавления сахара или других подслащивающих веществ, неароматизированные лед и снег
	2202.00	Воды, включая минеральные и газированные, с добавлением сахара или других подслащивающих или ароматических веществ, безалкогольные напитки прочие, за исключением фруктовых или овощных соков, классифицируемых в товарной позиции 20.09
	2204.00	Вина виноградные натуральные, включая крепленые сусло виноградное, кроме классифицируемого в товарной позиции 20.09
	2208.00	Спирт этиловый неденатурированный, крепостью менее 80 %. Крепкие спиртные напитки, ликеры и прочие алкогольные напитки, прочие смешанные спиртовые полуфабрикаты, используемые для изготовления напитков
		2208.30 Виски
		2208.40 Ром и тафия
		2208.50 Джин и можжевельная настойка
		2208.90 Прочие
	2209.00	Уксус и его заменители, полученные из уксусной кислоты

Рис 1.13. Фрагмент «Гармонизированной системы для товарной номенклатуры внешнеэкономической деятельности (Harmonized Commodity System)»

Универсальная десятичная классификация (УДК) (рис. 1.16, б), представляет собой пример классификатора с переменной (неопределенной) глубиной.

A00 - PHYSICAL SCIENCES.

A10 - GENERAL PHYSICS.

- A11 - Theoretical Physics.
- A12 - Atomic and Molecular Physics.
- A13 - Solid States and Fluid Physics.
- A14 - Plasma Physics and Thermonuclear Reactions.
- A15 - Astrophysics and Cosmology, Cosmic Radiation.

...

A20 - HIGH ENERGY PHYSICS.

- A21 - Elementary Particles (Theory).
- A22 - Elementary Particles (Experimental).

...

B00 - CHEMISTRY, MATERIALS AND EARTH SCIENCES.

B10 - CHEMISTRY

- B11 - Chemical & Isotopic Analysis.
- B12 - Inorganic, Organic & Physical Chemistry.
- B13 - Radiochemistry and Nuclear Chemistry.

...

B20 - MATERIALS

- B21 - Metals and alloys (Production and Fabrication).
- B22 - Metals and Alloys (Physical Properties and Structure).
- B23 - Ceramics and Cermets.
- B25 - Radiation Effects on Physical Properties of Materials.

...

F00 - OTHER ASPECTS OF NUCLEAR ENERGY.

F10 - ECONOMICS.

- F11 - Nuclear Power Economics.
- F12 - Reactor Fuel Economics.

...

F30 - NUCLEAR DOCUMENTATION.

- F31 - Data Handling.
- F32 - Literature Handling.

F40 - SAFEGUARD AND INSPECTION.

- F41 - Technical Aspects.
- F42 - Non-Technical Aspects.

Рис. 1.14. Фрагменты классификатора INIS
(трехуровневая иерархическая классификационная система)

Раздел	Подраздел	Содержание рубрики
600000	Резисторы постоянные	
	601000	Резисторы постоянные непроволочные
	602000	Резисторы постоянные проволочные
	606000	Резисторы постоянные проволочные и фольговые
610000	Резисторы переменные и прочие	
	611000	Резисторы переменные непроволочные
	615000	Резисторы переменные проволочные и фольговые
	619000	Резисторы прочие
	619100	Резисторы полупроводниковые
	619900	Резисторы разные
620000	Конденсаторы	
	621000	Конденсаторы постоянной емкости керамические
	623000	Конденсаторы постоянной емкости на основе стекла
	628000	Конденсаторы прочие
	628100	Конденсаторы постоянной емкости бумажные
	628200	Конденсаторы постоянной емкости тонкопленочные с неорганическим диэлектриком
	628300	Конденсаторы постоянной емкости с газообразным диэлектриком и вакуумные
	628600	Конденсаторы переменной емкости

Рис. 1.15. Фрагмент «Общероссийского классификатора продукции (ОКП)»

<p>Н-электричество (раздел)</p> <p>Н05-Специальные области электротехники (класс)</p> <p>Н05-плазменная техника (подкласс)</p> <p>Н05Р1-00-получение плазмы (группа)</p> <p>Н05Н1-02-устройства для удержания плазмы (подгруппа)</p> <p>Н05Н1-04- —*— с использованием магнитных полей (подгруппа)</p> <p>Н05Н1-06 - устройства для сжатия канала плазмы (подгруппа)</p> <p>Н05К-печатные схемы (подкласс)</p> <p style="text-align: center;">а)</p>	<p>б-прикладная наука, медицина, техника</p> <p>681-точная механика</p> <p>681.1-</p> <p>...</p> <p>681.5-автоматика, техническая кибернетика</p> <p>681.51-системы автоматизированного управления</p> <p>...</p> <p>681.514 — стохастические САУ</p> <p>681.518 — информационные системы</p> <p style="text-align: center;">б)</p>
--	---

Рис. 1.16. Примеры иерархических классификаторов:
а — МКИ; б — УДК

Тезаурусы АИС. Тезаурус (Thesaurus) представляет собой толковый *дескрипторный словарь*, в котором значение каждой стандартной лексической единицы (*дескриптора* — слова или словосочетания) интерпретируется через связи с другими дескрипторами.

Тезаурус в печатной форме обычно включает две части:

- систематический указатель дескрипторов;
- алфавитный указатель.

Любая часть тезауруса есть перечень *дескрипторных статей*, упорядоченных в первом случае по тематическим группам и подгруппам, а во втором — в алфавитном порядке.

На рис. 1.17 приведено по две дескрипторные статьи из тезаурусов INIS и INSPEC (информационная служба по физике, электронике и кибернетике — (б)). Началом каждой статьи является заглавный дескриптор или запрещенный термин (начинается с «—»).

<i>-cathodol</i>
<i>USE pyrocatechol</i>
CATIONS
<i>UF - positive ions</i>
<i>BT1 ions</i>
<i>BT2 charged particles</i>
<i>NT1 hydrogen 1 ions plus</i>
<i>NT2 protons</i>
<i>NT3 delayed protons</i>
<i>NT1 hydrogen ions 3 plus</i>
<i>RT chemical states</i>
<i>RT electrolysis</i>
<i>RT ion beams</i>
<i>RT ion exchange materials</i>

а)

INFORMATION RETRIEVAL
<i>see also information storage</i>
<i>UF document retrieval</i>
<i>BT information science</i>
<i>TT computer applications</i>
<i>RT information analysis</i>
<i>information storage</i>
INFORMATION RETRIEVAL
SYSTEM EVALUATION
<i>BT evaluation information</i>
<i>retrieval system</i>
<i>TT computer application</i>
<i>evaluation</i>

б)

Рис. 1.17. Фрагменты тезаурусов:

а — INIS; б — INSPEC

Дескрипторная статья начинается списком запрещенных понятий (связка «USE FOR» или «используемый вместо»), заменяемых при индексировании данным дескриптором. Далее, в ней содержатся сведения о связи заглавного с другими дескрипторами. В примерах на рис. 1.17 присутствуют следующие типы связей:

- *вышестоящие* термины (родовые — Broader Term) обозначены BT1, BT2, BTX и т. д. Число X означает номер уровня иерархии. Последовательность BT1 — BT2 — BT3... образует ветвь;
- *нижестоящие* термины (видовые — Narrower Term) обозначены NT1, NT и т. д., ветви и уровни определяются аналогично с BT;
- *ассоциативные* термины (Related Term), имеющие иерархическую связь с заглавным дескриптором. Обозначены с RT;
- *наивысший* родовой термин (Top Term — TT) — в тезаурусе INSPEC.

Правильная структура тезауруса должна удовлетворять следующим очевидным условиям:

- полнота — каждый термин, содержащийся в поле BTX, NTX, RT, UF и т. д., обязательно должен быть или заглавным дескриптором какой-либо статьи, или запрещенным термином;

- корректность — если в дескрипторной статье термина А в качестве ВТ1 присутствует термин В и в этой же ветви в качестве ВТ2 присутствует С, то в дескрипторной статье В термин С должен содержаться в качестве ВТ1. То же самое справедливо для отношения NTX и RT.

При составлении тезаурусов иногда используют также отношения «часть—целое», «причина—целое», «объект—применение» и т. д., однако в большинстве случаев достаточно ограничиться указанными ВТ, RT, NT, сводя к ним остальные типы.

Информационные языки

К информационным языкам относят *языки описания и манипулирования* данными. Сравнительный анализ некоторых ЯМД и ЯОД для конкретных систем приведен в гл. 2, табл. 2.2.

Языки манипулирования данными. ЯМД состоит из двух основных разделов:

- поиск данных;
- отображение данных.

Поиск данных предполагает наличие *критерия смыслового соответствия* (КСС) или решающего правила, определяющих факт формальной релевантности поискового образа документа (ПОД) поисковому образу запроса (ПОЗ). В общем случае КСС является некоторым предикатом (условным высказыванием), область истинности которого есть множество выдаваемых документов. Эта концепция лежит в основе большинства языков запросов (ЯЗ), некоторые из которых рассмотрены ниже (STAIRS, ADABAS, Foxpro, Irbis).

Отображение данных предполагает наличие языковых или иных (например, табличных) средств описания форматов (видов, подсхем) представления сгенерированных системой данных пользователю.

ЯМД включают в себя: *языки запросов* (ЯЗ — преимущественно для фактографических АИС) и *информационно-поисковые языки* (ИПЯ — преимущественно для документальных АИС).

Языки запросов. Следует отметить, что в настоящее время фактическим стандартом ЯЗ является SQL (Structured Query Language), который, например, подробно описан в одной из книг данной серии [6], поэтому мы здесь на нем останавливаться не будем. Ниже, в гл. 2 вкратце рассматриваются форматы альтернативных языков — Natural и Foxpro.

Информационно-поисковые языки. ИПЯ включает в себя лексику — словарь единиц текста, используемых для индексирования, и

грамматику — совокупность правил составления поисковых образов и уточнения смысла лексических единиц по контексту, позиции, квалификации.

В настоящее время в большинстве систем фактически используется упрощенная грамматика и ограниченная, но достаточно развитая лексика.

ПОД в типичной ИПС без грамматики есть неупорядоченный набор лексических единиц или же ограниченная совокупность именуемых неупорядоченных наборов (сегментов). Индексирование без грамматики разделяется на два типа:

- прекоординируемое;
- посткоординируемое.

К *первому типу* относятся уже упомянутые иерархические классификационные системы, такие, как УДК (см. рис. 1.17, б), МКИ — (см. рис. 1.17, а), INIS и т. д. Предполагается, что такая система заранее содержит все классы, к которым может быть отнесен любой документ, закодированный соответствующим индексом, например 681.5 — автоматика, техническая кибернетика (см. рис. 1.17, б).

Ко *второму типу* относятся дескрипторные, координатные ИПЯ, позволяющие приписать каждому документу несколько дескрипторов, каждый из которых является именем широкого класса понятий, терминов и, следовательно, помечает множество, в которое данный документ входит. Необходимо заметить, что иерархические классификационные системы также в ограниченных масштабах используют дескрипторные принципы.

В основе дескрипторных ИПЯ находятся тезаурусы (рассмотренные выше). Очевидно, использование словосочетаний или составных дескрипторов в качестве лексических единиц компенсирует отсутствие грамматики в дескрипторном ИПЯ.

Ниже приводится пример распечатки диалогового поиска в файлах службы STN International, на котором хорошо видны форматы поиска и выдачи информации, типичные для различных ЯЗ/ИПЯ (рис. 1.18).

Языки описания данных. Выше уже отмечалось, что одной из первых попыток создания ЯОД был язык DL/1 (Data Language # 1) фирмы IBM. В настоящее время в связи с широким распространением уже упомянутого SQL, в котором предусмотрена компонента описания БД (см. гл. 2, табл. 2.2) стандартом ЯОД является данная компонента. Поскольку этих описательных возможностей, тем не менее, обычно оказывается недостаточно, и SQL не является единственным средством разработки АИС, существуют и другие подходы, которые обычно базируются на понятии *словаря данных* (файл

*****WELCOME TO STN
INTERNATIONAL*****

=>FILE NTIS

FILE 'NTIS' ENTERED AT 10:45:36 ON 16 SEP 2000
COPYRIGHT (c) 2000 NATIONAL TECHNICAL INFORMATION SERVICE (NTIS)
FILE LAST UPDATED:04 SEP 2000 <20000904/UP>
=>S (FIBRE# OR FIBER#) (2A) OPTIC?

1159 FIBRE#
16323 FIBER#
47516 OPTIC?
L1 3947 (FIBRE# OR FIBER#) (2A) OPTIC?

=>D 1-3 TI

L1 ANSWER 1 OF 3947

TI Untersuchungen zum Einsatz von Uebergangsstrahlung zur Teilchenidentifikation bei ZEUS. (Studies on the application of transition radiation for the particle identification at ZEUS). (Diss.(Dr.rer.nat.))

L1 ANSWER 2 OF 3947

TI Optische Sender und Sendemodule. Schlussbericht. (Optical transmitters and transmitter modules. Final report)

L1 ANSWER 3 OF 3947

TI Wellenlaengenmultiplexkomponenten. Schlussbericht. (Wavelength multiplexing components. Final report)

=>D 2 BIB AB

L1 ANSWER 2 OF 3947

AN 92(18):2167 NTIS Order Number : TIB/A2000-91109/XAD

TI Optische Sender und Sendermodule. Schlussbericht. (Optical transmitters and transmitter modules. Final report)

AU Mettler, K.; Gier, J., Spaeth, W.

CS Siemens A.G., Munich (Germany, F.R.). Forschungslaboratorium; Bundesministerium fuer Forschung und Technologie, Bonn, (Germany, F.R.).

NC Contract: TK0191 S/1

NR TIB/A92-81109/XAD

108 p. NTIS Prices : PC E07

Notes: With 41 refs., 6 tabs., 52 figs

PD 1990

LA German CY Germany, Federal Republic of

OS GRA&I92818

AB The increasing importance of fiber-optic communication creates a growing demand for practical optical transmission equipment, especially for optoelectronic transmitter and receiver components. The main objective of this project was the development of optical transmitter devices for communication systems employing digital signals and transmission wavelength around 1300 nm. . . .

(TIB: FR716. (Copyright (c) 2000 by FIZ Citation no. 2000:1109.) =>FILE INSPEC

FILE 'INSPEC' ENTERED AT 10:50:52 ON 16 SEP 2000
COPYRIGHT (c) 2000 INSTITUTION OF ELECTRICAL ENGINEERS
FILE LAST UPDATED: 11 SEP 2000 <200011/UP>
=>S L1 37724 FIBRE#

23466 FIBER#
233687 OPTIC?

L2 26120 (FIBRE# OR FIBER#) (2A)OPTIC?

=>D 1-3 TI

L2 ANSWER 1 OF 26120

TI Dynamics of thermal processes in the drawing of quartz optical fibers.

```

L2 ANSWER 2 OF 26120
TI Fiber-optic ICs accelerate communication over data links.
L3 ANSWER 2 OF 26120
TI A fiber optic gyro strapdown reference system for guided weapons.
->FILE CIMPENDEX
FILE 'COMPENDEX' ENTERED AT 10:52:06 ON 16 SEP 2000
COPYRIGHT (c) 2000 ENGINEERING INFORMATION, INC.
FILE LAST UPDATEED:10 SEP 2000 <20000910/UP>
FOR CONFERENCE PAPERS SEE FILE MEET
=>S L1

              7328      FIBRE#
              40417     FIBER#
              90630     OPTIC?
L3 12735      (FIBRE# OR FIBER#) (2A) OPTIC?
=>D1-5 TI
L3 ANSWER 1 OF 12375
TI FIBER OPTIC AND LASER SENSORS IV
L3 ANSWER 2 OF 12375
TI FEMTOSECOND PULSE PROPAGATION IN OPTICAL FIBERS:HIGHER
ORDER EFFECTS.
L3 ANSWER 3 OF 12375
TI FIBER-OPTICTRIGGERED HIGH-POWER LOW-PRESSURE GLOW DISCHARGE
SWITCHES.
. . .

```

Рис. 1.18. Пример поиска в Базах данных STN International -- NTIS и COMPENDEX:

команды пользователя и сообщения системы: S (search) — команда поиска; D (display) — команда выдачи результатов; L1 — множество (количество) найденных документов. *Поля документов из баз данных:* TL — заголовок; AN — номер документа в БД; AU — автор документа; CS — корпоративный автор (организация — источник документа); NC — регистрационный номер контракта; NR — регистрационный номер отчета; PD — дата публикации; LA — язык публикации; OS — связанные источники информации; AB — реферат; CY — страна публикации; BIB — групповое поле, состоящее из AN, TI, AU, CS, NC, NR, PD, LA, CY, OS

или таблица БД), который содержит описания данных и типов их обработки.

В настоящем пособии приводится ряд примеров таких словарей:

- БД STAIRS (DBD, см. гл. 2);
- АИС ЮРИУС (IXDDM, см. гл. 4).

Поэтому мы здесь ограничимся примером описания полей БД INSPEC хост-системы Dialog (см. табл. 1.14).

В словаре данных администратором системы задается тип обработки каждого поля при загрузке БД: *построчный (Phrase Indexing)* или *пословный (Word Indexing)*. В первом случае поле рассматривается как целое и полностью помещается в частотный словарь БД, во втором. — осуществляются выделение отдельных слов (с использованием символов-разделителей и словарей запрещенных слов) и их загрузка в словари и индексы.

1.7. Структуры баз данных

Рассмотрим вкратце обобщенные логическую и физическую структуры БД.

Логическая структура БД

Логическая структура (рис. 1.19) предполагает следующие уровни рассмотрения БД:

- база данных (database) — включает одну или несколько подбаз (файлов, таблиц, массивов), каждая из которых состоит из агрегатов данных (записей, документов) — record. Запись иденти-

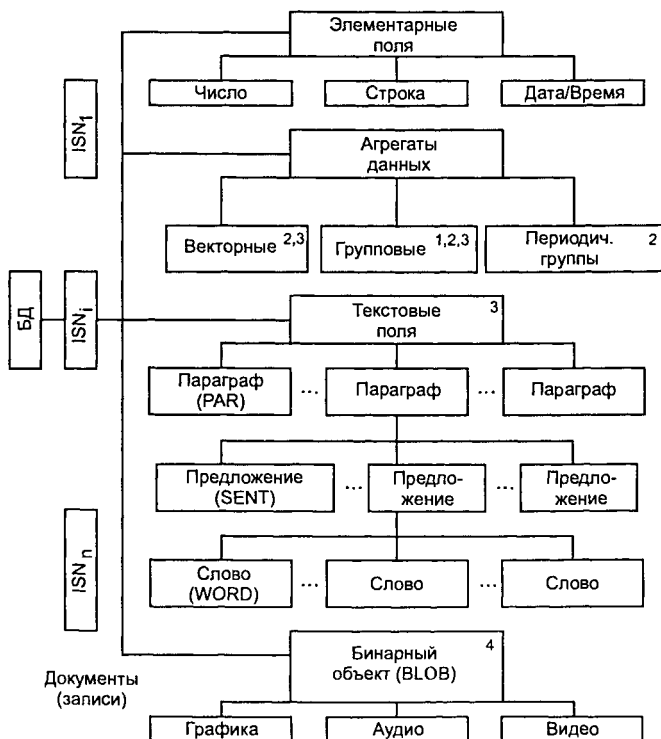


Рис. 1.19. Основные элементы логических структур данных в информационных системах:

1 — поля записей табличных (реляционных) баз данных (ORACLE, Foxpro, Access); 2 — поля записей (документов) постреляционных БД (ADABAS); 3 — поля документов ИПС (STAIRS, Dialog, IRBIS, ISIS); 4 — данные, которые могут быть связаны с полями базы данных

фицируется *внутренним номером* (ISN — internal sequential number, ВНЗ — внутренний номер записи, SDN — sequential document number и пр.);

- *запись (документ)* — совокупность разнотипных и разноструктурных данных, описывающих (относящихся к) объект реального мира, элемент предметной области АИС. Запись состоит из *полей* (field);
- *поле* — именованный элементарный или составной фрагмент записи (документа), содержащий информацию об определенном аспекте (аспектах) элемента (элементов) предметной области.

Возможны следующие структуры полей (рис. 1.19, табл. 1.14):

- *элементарные* — имеют фиксированную или ограниченную длину и не содержат входящих в них структур данных;
- *составные* (групповые) — образуются как агрегаты элементарных и также имеют фиксированную и ограниченную длину (реже — переменную или неопределенную, что связано с количеством вхождений элемента в агрегат);
- *текстовые* — поля переменной (неопределенной) длины и сложной внутренней структуры (обычно это иерархическая последовательность типа РАЗДЕЛ — ПОДРАЗДЕЛ — ПРЕДЛОЖЕНИЕ — СЛОВО);
- *бинарные* — данные, интерпретируемые как поля, однако обычно физически не входящие в состав записей БД. Необходимо отметить, что поля данного типа (BLOB — Binary Large Object) фактически являются данными, до обработки которых рассматриваемая конкретная СУБД еще «не доросла» и поэтому работа с ними возлагается на пользователя (прикладные программы). В частности, в системах Foxpro и Clipper большие текстовые (так называемых MEMO) поля также не обрабатываются системой и фактически оказываются в статусе BLOB.

Таблица 1.14. Описания некоторых полей БД INSPEC в хост-системе Dialog

Поле	Наименование поля	Тип индексирования	Примеры поиска
Основной индекс			
/AB	Реферат	Слова	S CLADDINGS(W)LAYER/AB
/DE	Дескрипторы	Слова и строки	S FEEDBACK(W)LASER?/DE; S III-V SEMICONDUCTORS
/ID	Идентификатор	Слова и строки	S LASER(W)DIODE?/ID; S MESA-STRIPE GEOMETRY/ID
/PI	Заглавие документа	Слова	S DISTRIBUTED(W)FEEDBACK/PI

Окончание табл. 1.14

Поле	Наименование поля	Тип индексирования	Примеры поиска
Дополнительные индексы			
AN=	Номер INSPEC	Строки	S AN=A23476198
AU=	Автор	Строки	S AU=OHATA, T.
CS=*	Организация	Слова	S CS=(AMERICAN(W)CYANAMID)
JN=	Журнал	Строки	S J. CRYST. GROWTH (NETHERLANDS)
TC=	Тип применения	Строки	S TC=EXPERIMENTAL
Химические индексы			
CI=	Вещество, элемент	Слова и строки	S CI=GAAS; S CI=AS
Числовые индексы			
AM=	Сила тока (А)	NI=CURRENT	S AM=0.005
BI=	Объем памяти (бит)	NI=STORAGE CAPACITY	S BI=4.4E6(S)NI=STORAGE CAPACITY
BS=	Скорость передачи (бит/с)	NI=BIT RATE	S BS=64000
EV=	Энергия (электрон-вольт)	NI=ELECTRON VOLT ENERGY	S EV=0.5:1.0
HZ=	Частота (Гц)	NI=FREQUENCY	S HZ=10:100
KE=	Температура (К)	NI=TEMPERATURE	S KE=271
MT=	Длина волны (м)	NI=WAVELENGTH	S MT=1E-2(S)NI=WAVELENGTH
VL=	Напряжение (В)	NI=VOLTAGE	S VL=-5
YR=	Возраст (лет)	NI=AGE	S YR=1E9:E40

Физическая структура БД

Данный тип структуры в общем случае имеет вид, приведенный на рис. 1.20, и включает следующие компоненты:

- *файл (файлы) исходных (первичных) данных* (текстов, бинарных данных) — содержит собственно объекты, подлежащие поиску, обработке и пр.;
- *файл (файлы) вторичной (справочной) информации* (регистрационные карты, библиографические реестры и пр.) — содержит описания исходных элементов (объектов). Важным видом справочных файлов являются *классификаторы, кодификаторы,*

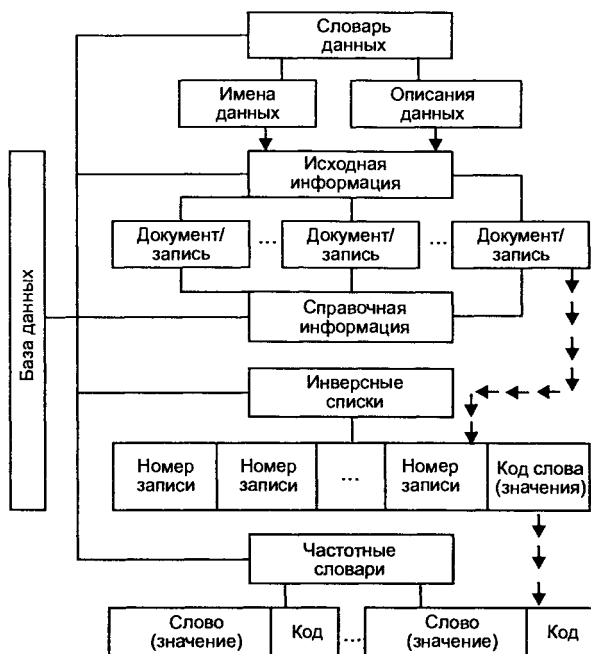


Рис. 1.20. Обобщенная физическая структура данных в БД

тезаурусы, обеспечивающие полноту и компактность представления информации в БД;

- *индекс* — файл (файлы), связывающий адрес (номер) объекта с его содержанием (значением атрибута объекта), обычно состоит из инверсного списка и частотного словаря, который облегчает составление запросов на поиск и повышает обзорность БД;
- *словарь данных* — файл, содержащий составленное с необходимой степенью подробности описание состава БД, документов, записей, агрегатов данных, их имена, типы и структуры, способы интерпретации и обработки.

Изменение содержания БД может осуществляться как в режиме конечного пользователя (диалоговый ввод или коррекция записей/документов по полям) — обычный для СУБД и редкий для АИПС, так и в режиме администратора БД (обычный для АИПС и редкий для СУБД), при этом происходит массовый ввод или загрузка записей/документов.

При любом виде добавления документа/записи для каждого поля осуществляется анализ, обработка и согласованное помещение документа и его фрагментов в соответствующие физические файлы БД. В конкретных случаях возможна неполная комплектность приведенной физической схемы:

- в фактографических (табличных) БД вторичный файл может являться основным накопителем информации, а текстовые и бинарные данные — фигурировать в качестве необязательного приложения;
- в справочно-библиографических БД текстовые данные могут находиться во вторичном файле, а первичный — отсутствовать;
- в БД с полнотекстовым поиском может отсутствовать вторичный файл, а индексирование (построение частотных словарей и инверсных списков) проводится по первичному файлу (страницы или абзацы полных текстов);
- может отсутствовать частотный словарь или инверсный список.

Надо отметить также вариативность физической реализации и взаимосвязи лингвистического и информационного обеспечения АИС:

- словарь данных может физически входить в информационные файлы (первичный или вторичный);
- классификаторы, кодификаторы, тезаурусы могут быть оформлены как физическими файлами (файлами ОС), так и входить в состав БД в виде отдельных таблиц (файлов БД, массивов и пр.) на логическом уровне и т. п.

1.8. Организационное обеспечение и пользователи АИС

В состав организационного обеспечения АИС принято включать *структурные подразделения* организации, осуществляющие управление технологическими процессами и поддержку работоспособности системы, а также *совокупность документации*, на основании которой обслуживающий персонал, разработчики и пользователи осуществляют эксплуатацию и развитие системы.

В состав *пользователей АИС* в общем случае входят следующие группы лиц, контактирующие с системой:

- администратор БД (АБД): лицо или группа, отвечающая за проведение данных, назначение уровней доступа, включение/исключение пользователей. Обычно АБД участвует в проектировании и определении структуры БД;

- прочие пользователи (не обладающие привилегиями доступа к данным), в том числе;
- операторы подготовки данных (ОПД) — персонал, осуществляющий ввод данных с рабочих листов или документов на основе соответствующих инструкций в среде специальных программных интерфейсов;
- интерактивные пользователи (ИП) — лица, имеющие доступ на ввод, коррекцию, обновление, уничтожение и чтение данных в рамках, как правило, ограниченной области БД;
- конечные пользователи (КП) — лица, использующие БД для получения справок и решения задач.

Иногда в круг пользователей БД включают также разработчиков приложений (программ, обращающихся к БД через ядро СУБД), написанных на языках прикладного программирования.

Работа пользователей с БД тесно связана с операциями определения (описания) (в основном АБД) и манипулирования данными (в основном прочие пользователи), которым соответствуют языковые средства описания — словари данных и таблицы определения полей и файлов, и манипулирования данными — языки запросов и программирования информационных систем.

Существует альтернативная точка зрения (хотя авторы считают ее несколько экстремистской, она будет вкратце изложена в гл. 3), суть которой состоит в том, что пользователь рассматривается как компонента информационной системы, находящаяся «внутри» нее. Эта точка зрения полезна при разработке некоторых типов систем, и включением здесь пользователей в состав «обеспечения АИС» мы частично ее принимаем. Однако очевидно, что человек никогда не станет придатком машин (пусть и самых «умных») и не будет реально входить в состав каких-либо систем в качестве «винтика». История человечества неоднократно отвергала подобные попытки.

Правовое обеспечение информационных систем

Внедрение автоматизированных систем различного назначения в определенной степени воздействует на виды, формы и характер функциональных взаимоотношений между существующими структурными подразделениями, поэтому четкая правовая регламентация прав, обязанностей и ответственности пользователей при автоматизации определенных видов работ является одним из самых сложных, ответственных и обязательных условий, которые необходимо учитывать при проектировании, использовании и развитии новых

методов обработки информации. Структура правового обеспечения информационной системы дана в табл. 1.15.

Таблица 1.15. Правовое обеспечение информационной системы

Разработка проекта	Функционирование системы
Приказы руководящего органа о начале работ по созданию и организации информационной службы	Положение о службах, обеспечивающих функционирование автоматизированных систем
Общепромышленные нормативные акты, регламентирующие отношения разработчика и заказчика	Должностные инструкции
Государственные и отраслевые стандарты на разработку информационного обеспечения	Нормативные акты о порядке создания и использования информации
Государственные стандарты на разработку проектной документации	Нормативные акты, регламентирующие технологический процесс автоматизированной обработки информации
Нормативные акты на получение и использование комплекса технических средств	Правовое положение отдельных видов автоматизированных систем управления и проектирования
Правовое положение службы АСУ ИПК	Нормативные акты на использование вычислительной техники

Стабильность данных, содержащихся в нормативной документации, относительна, т. е. практически все данные с разной периодичностью корректируются, уточняются и обновляются. Важным аспектом правового обеспечения АИС является проблема правового обеспечения защиты информации. Эти вопросы рассмотрены в [28].

Контрольные вопросы

1. В чем различие информационных технологий, систем и ресурсов?
2. Какие классы информационных технологий Вам известны? Приведите примеры.
3. Назовите основные основания для классификации АИС, приведите примеры.
4. Назовите основные классы информационных ресурсов.
5. Что такое службы-генераторы БД? Приведите примеры.
6. Что такое онлайн-службы? Приведите примеры.
7. Какие типы информации распространяются онлайн-службами?

8. Что такое форматная база?
9. Приведите примеры различных типов, форматов и структур данных.
10. Что такое коммуникативные форматы? Определите основные проблемы построения коммуникативных форматов.
11. Что такое идентификация данных и локализация описаний данных?
12. Назовите основные типы файлов.
13. Какие типы разметки текстовых файлов Вам известны?
14. Что входит в состав лексической (словарной) базы?
15. Каковы основные отличия тезаурусов от классификаторов и кодификаторов?
16. Перечислите основные типы информационных языков.
17. Какова структура дескрипторной статьи тезауруса?
18. Что такое организационное обеспечение АИС, и из каких компонентов оно состоит?

Глава 2

БАЗОВЫЕ ТИПЫ

ИНФОРМАЦИОННЫХ СИСТЕМ

Рассмотренная ранее классификация АИС является чрезмерно детальной с точки зрения наиболее распространенных и перспективных типов, к которым относятся (табл. 2.1):

- фактографические;
- документальные;
- интеллектуальные (экспертные);
- гипертекстовые.

Выбор именно этих типов АИС для более детального анализа определяется следующими факторами:

- системы появлялись и развивались именно в данной исторической последовательности;
- более ранние типы систем (фактографические, документальные) являются, как правило, платформой и средой для реализации более поздних (экспертные, гипертекстовые).

Перечисленные типы характеризуют следующие отличительные черты:

- распространенность (в статистике мировых информационных ресурсов документальные и фактографические БД занимают 1- и 2-е места — см. табл. 2.1);
- перспективность (интеллектуальные системы успешно осваивают новые области применения);
- гипертекстовые системы являются основой мировой информационной сети WWW (World Wide Web) — наиболее популярной составляющей Интернета.

При этом хотелось бы отметить, что в традиционном понимании выражение «информационная система» (особенно «автоматизированная информационная система» или «автоматизированная информационно-поисковая система» — АИПС) обычно ассоциируется с документальными системами (базами данных); термин же «база данных», как правило, ассоциируется с фактографическими, управленческими системами, задачами типа АСУ. Хотя, конечно же, и те

Таблица 2.1. Сравнительные характеристики основных типов АИС

Атрибуты систем	Типы систем			
	Фактографические	Документальные	Интеллектуальные	Гипертекстовые
Модель (структура) предметной области	Структура БД (логическая и физическая)	Содержание БД (структура стандартна)	Структура и содержание БД (БЗ)	Структура и содержание БЗ (ГБ)
Информационная совокупность	База данных (иногда файловая система ОС)	База данных	База знаний	Сеть связей (гипербаза – ГБ)
Единица информации	Запись (агрегат данных жесткой структуры)	Документ (агрегат данных диффузной структуры)	Факт (константа), высказывание (переменная)	Узел сети (запись, документ, факт), связь (ссылка, тематическая, смысловая, родовидовая, причинно-следственная)
Физическая среда хранения информации	Файловая система ОС	Файловая система или фактографическая БД	Файловая система ОС, фактографическая (или другого типа) БД	Файловая система, БЗ, БД
Ввод информации	Загрузка БД	Загрузка БД	Подсистема приобретения знаний (Knowledge acquisition)	Загрузка файлов или БД, или БЗ
Вывод информации (входной язык)	Пользовательский интерфейс (язык запросов)	Пользовательский интерфейс (информационно-поисковый язык)	Интерфейс пользователя, подсистема разъяснения результатов (explaining subsystem)	Навигатор сети или язык запросов
Обработка (поиск) информации	ОС или СУБД	СУБД или программная оболочка АМПС	СУБД или программная оболочка ЭС (машина логического вывода – Inference engine)	Гипермедиа-машина навигации
Программные средства реализации	Табличные СУБД и системы программирования с элементами СУБД – ADABAS, Foxpro, Oracle, MS SQL SERVER	Программные оболочки АМПС – DPS, STAIRS, ISIS, IRBIS	Оболочки ЭС (Shells) – EMYCIN, EXPERT, OPS, HEARSAY, EXSYS, Insight, ExpertEase	NoteCards, Tektronix Neptune, Гип-Си, ГиперМетод, ГИПЕРНЕТ, Мета Дизайн, АСАД

и другие типы систем являются *информационными* и включают *базы данных* в свой состав.

В этой традиционной интерпретации находит свое отражение то обстоятельство, что в фактографических системах *модель предметной области заключена в структуре БД*, и потому основное внимание сосредоточивается на проблеме проектирования БД, в документальных же системах *моделью является наполнение, содержание БД*, в том числе словарей, тезаурусов и т. д., поэтому основное внимание уделяется языковым, семантическим проблемам. (Эти и другие различия указанных типов систем описываются в табл. 2.1.)

2.1. Фактографические АИС

Основные признаки — простая структура данных и сложная система взаимосвязей между агрегатами данных.

В исторической последовательности развития данных систем сначала появились АИС, базирующиеся на *иерархических*, затем на *сетевых* и, наконец, на *реляционных и постреляционных* представлениях о структуре предметной области. В настоящее время наиболее распространенным подходом является *реляционный* (табличные БД), что не исключает, конечно, включения элементов иерархических и сетевых представлений при проектировании АИС.

Модели данных и структура БД

Поскольку в данном случае БД является информационной моделью определенной предметной области, существенной особенностью всякой БД является структура или, как принято говорить, модель данных (МД).

Рассмотрим некоторые наиболее известные (или «замечательные» [1, 5]) модели данных.

Иерархическая МД (ИМД). Впервые реализована в СУБД IBM — IMS (Information Management System), разработанной для поддержки банка данных по программе Apollo. При данном подходе предметная область представляется в виде совокупности структур иерархического типа (граф — «дерево»).

Основные понятия ИМД:

- *поле* — минимальная единица данных;
- *сегмент (узел)* — совокупность полей, являющаяся единицей обмена между БД и прикладной программой. Сегмент (узел

иерархического графа) более высокого уровня называется *исходным (родительским)* по отношению к ниже расположенному *порожденному (отпрыску)*. Может использоваться также терминология «узел, принадлежащий вышестоящему узлу».

Конкретные данные, входящие в сегмент называются *экземпляром сегмента*.

В ИМД существуют также следующие понятия:

- *брат* — узел, имеющий того же родителя, что и другой узел;
- *ветвь* — узел дерева вместе со всеми его отпрысками, отдаленными потомками и родительскими источниками;
- *лист* — узел, у которого нет отпрысков;
- *обход дерева* — процесс обследования по очереди каждого узла дерева в иерархической модели данных, и пр.

Преимущества IMS и реализованной в ней иерархической модели:

1. *Простота модели*. Принцип построения IMS легок для понимания. Иерархия базы данных напоминает структуру компании или генеалогическое дерево.

2. *Использование отношений предок/потомок*. СУБД IMS позволяла легко представлять отношения предок/потомок, например: «А является частью В» или «А владеет В».

3. *Быстродействие*. В СУБД IMS отношения предок/потомок были реализованы в виде физических указателей от одной записи к другой, вследствие чего перемещение по базе данных происходило быстро. Поскольку структура данных в этой СУБД отличалась простотой, IMS могла размещать записи предков и потомков на диске рядом друг с другом, что позволяло свести к минимуму количество операций записи-чтения.

Существенно то, что физическая организация БД в этом случае такова, что выбрать конкретные сведения об объектах можно, лишь пройдя всю цепочку групп (сегментов) сверху вниз (путь на иерархическом дереве). Данная схема наиболее проста, но не лишена очевидных недостатков.

В частности, в связи с полииерархичностью связей объектов в реальном мире в подобных БД необходимо создавать и поддерживать несколько иерархических отношений, что нарушает основную идею модели данных. Далее, рассматриваемая модель обладает рядом т. н. «парадоксов», наиболее очевидным из которых является «парадокс исключения». Удаление из БД некоторого вышестоящего сегмента приводит к автоматическому удалению и всех зависимых (порожденных сегментов).

Сетевая модель данных (модель CODASYL). В предложенной CODASYL модификации иерархической модели одна запись могла участвовать в нескольких отношениях предок/потомок. В сетевой модели такие отношения называются *множествами (set)*. В 70-е гг. независимые производители программного обеспечения реализовали сетевую модель в таких продуктах, как IDMS компании Cullinet, Total компании Cincom, которые приобрели большую популярность. Сетевые БД обладали рядом преимуществ:

1. *Гибкостью.* Множественные отношения предок/потомок позволяют сетевой БД хранить данные, структура которых сложнее обычной иерархии.

2. *Стандартизованностью.* Появление стандарта CODASYL.

3. *Быстродействием.* Вопреки своей сложности, сетевые БД достигали быстродействия, сравнимого с быстродействием иерархических БД. Множества были представлены указателями на физические записи данных, и в некоторых системах администратор мог задать кластеризацию данных на основе множества отношений.

Недостаток — жесткость БД, наборы отношений и структуру записей приходилось задавать заранее. Изменение структуры данных означало перестройку всей БД.

Реляционная модель данных (РМД). В то время как иерархическая модель в своей основе является формализацией и обобщением пользовательских свойств некоторой конкретной системы (IMS), в случае реляционной модели сначала были разработаны некоторые математические основы и лишь через 5—10 лет появились первые коммерчески эффективные системы. В рамках реляционной модели предметная область представлена совокупностью *таблиц (отношений, файлов)*.

Строки таблицы называются *экземплярами отношения*, столбцы — *атрибутами*; каждый атрибут имеет область значений, называемую *доменом*.

Важным отличием РМД от ИМД является возможность применения формального аппарата, описывающего преобразование и обработку данных в РМД — *реляционной алгебры*.

Операндами реляционной алгебры являются отношения, как постоянные, так и переменные.

Операции реляционной алгебры включают следующие преобразования отношений.

А. Теоретико-множественные операции над несколькими подобными (имеющими одинаковую структуру — число атрибутов, их имен, домен и т. д.), отношениями, в том числе *объединение, пересечение, разность*.

Б. Операции над одним отношением:

- *селекция*, или построение отношения-результата из отношения-источника путем отбора экземпляров, удовлетворяющих некоторому критерию отбора. Операция селекции соответствует поиску информации в БД по логическим условиям (см. табл. 2.2 — FIND ... WITH, FIND...WHERE, LOCATE);
- *проекция*, или построение результирующего отношения путем отбора части атрибутов всех экземпляров исходного отношения. Данной операции в реальных СУБД соответствует понятие пользовательской подсхемы и операции выдачи необходимых данных (см. табл. 2.2 — DISPLAY, VIEW, SET FORM TO, REPORT FORM...).

В. Операции над несколькими различными отношениями.

Рассмотрим только *естественное соединение* (в дальнейшем — соединение). Операция заключается в поиске в паре (или большем числе) отношений строк, содержащих общий атрибут, и создания из этих строк экземпляра результирующего отношения.

В СУБД соединению соответствует поиск связанных данных или логическое (физическое) связывание файлов (см. табл. 2.2 — FIND...COUPLED, SET RELATION TO, JOIN).

Реляционная алгебра позволяет рассматривать операции ввода, вывода, поиска коррекции и удаления данных в БД как вычисление отношений-результатов через исходные отношения. При этом исходным отношением может быть внешний (входной) формат данных, а результирующим — внутренний (хранимый) или, наоборот, исходным — внутренний, а результирующим — внешний (выходной).

Модель «сущность—связь» (Entity—Relationship, ER) [5], представляет собой обобщение РМД путем разделения отношений, описывающих предметную область на две группы — *сущностей* и *связей*.

Сущность (Entity) является первичным, устойчивым объектом, описываемым некоторой совокупностью атрибутов.

Связь (Relationship) является вторичным понятием, характеризующим взаимодействие в пространстве и времени двух или более сущностей, и также задается рядом атрибутов, среди которых присутствуют идентификаторы взаимосвязанных сущностей. При проектировании БД на основе ER-моделей используют *ER-диаграммы*. Модель ER является удобным средством описания предметной области перед тем, как перейти к ее представлению в реляционной модели данных.

Иерархическая МД в настоящее время представляет лишь исторический интерес, хотя ряд ее элементов и поддерживается некоторыми из рассматриваемых далее конкретными СУБД.

Наиболее распространенными являются подходы, базирующиеся на ER-модели и РМД.

Основные представления о структуре БД в рамках указанных моделей заключаются в следующем:

а) совокупность *сущностей* и *связей* образует *концептуальную схему* базы данных и отражает структуру предметной области. Элементами схемы являются типы (классы) сущностей и связей; типы состоят из экземпляров, описываемых значениями атрибутов.

На рис. 2.1 приведен пример фрагмента диаграммы «сущность—связь», описывающей учебный процесс вуза. Здесь сущностями являются **ФАКУЛЬТЕТ**, **ДИСЦИПЛИНА**, **СПЕЦИАЛЬНОСТЬ** (с возможными атрибутами, например, **НАИМЕНОВАНИЕ**, **ПРОДОЛЖИТЕЛЬНОСТЬ ОБУЧЕНИЯ**, **ЧИСЛО ЧАСОВ** и пр.). Связями являются **ВЫПУСКАЕТ**, **ВКЛЮЧАЕТ** (возможные атрибуты — **КВАЛИФИКАЦИЯ**, **СЕМЕСТР ОБУЧЕНИЯ** и пр.);

б) концептуальная схема трансформируется в *логическую схему*, в которой сущностям и связям соответствуют *отношения или логические файлы*, состоящие соответственно из экземпляров отношений и логических записей.

Логическая запись является более общим образом, чем отношение (строка данных), поскольку допускает появление групповых полей (или агрегатных данных), соответствующих некоторым зависимым сущностям (или связям).

В повторяющемся групповом поле экземпляр группы есть описание экземпляра сущности (связи) посредством соответствующих атрибутов. Групповые повторяющиеся поля представляют собой элемент иерархической модели данных, который при желании может применяться пользователями;

в) следующий уровень — *физическая реализация БД* в форме файлов операционной системы ЭВМ. При этом в различных кон-

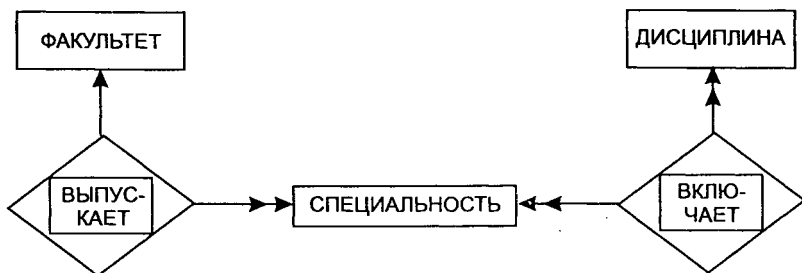


Рис. 2.1 Пример диаграммы «сущность—связь»

кретных системах логическому файлу может отвечать один или более физических файлов (или наоборот); физическая запись, как правило, включает одну или более логических записей;

г) уровень *представлений пользователя* описывает БД в виде совокупности пользовательских подхем, которые применяются для ввода/вывода информации. С представлениями пользователя связаны также понятия *маски редактирования* (преобразования данных при окончательном представлении пользователю), и *кодирования/декодирования (трансляции кодов)* — расширения кратких представлений данных и аббревиатур с помощью вспомогательных файлов и кодовых таблиц (по своей сути — операция соединения отношений в РМД).

Табличные базы данных

Рассмотрим пример БД, реализованной в рамках СУБД (точнее, системы программирования) FoxPro и состоящей из трех файлов данных (таблиц), описывающих некоторых граждан, их автомобили и финансовые учреждения, с которыми они связаны (рис. 2.2).

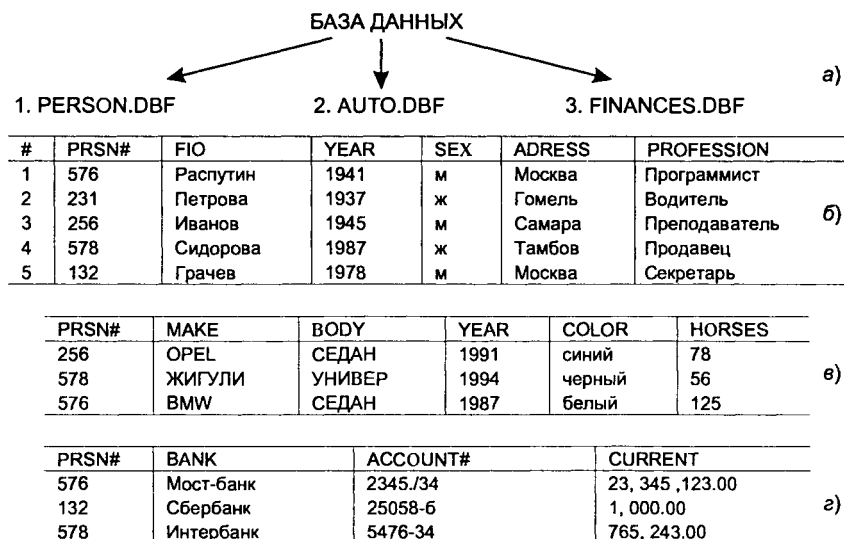


Рис. 2.2. Типичная структура простейшей табличной БД:

а — общая структура БД; б — структура PERSON.DBF; в — структура AUTO.DBF; г — структура FINANCES.DBF

Как уже отмечалось, подобные БД называются *табличными или реляционными* (от relation — «отношение»), и их теория рассматривается во многих источниках. Здесь мы ограничимся лишь базовыми понятиями:

- *файл* (file) соответствует совокупности однородных объектов и содержит их более или менее подробные описания в зависимости от приложений. Файл имеет имя (например, FINANCE.DBF и пр.);
- элементом файла является *запись* (record), или агрегат разнотипных данных, описывающих объект (точнее, экземпляр объекта). Записи имен не имеют, но им соответствуют физические номера в файле (колонка # на рис. 2.2, а);
- элементом записи (здесь — неделимым) является *поле* — данное, описывающее какой-либо аспект (или атрибут) объекта. Поля имеют имена (PRSN#, SEX и пр.). Разные файлы могут иметь поля с одинаковыми именами, но лучше этого избегать;
- при представлении файла в виде таблицы (рис. 2.2, а) столбцам соответствуют *атрибуты или поля*, строкам — *объекты или записи*. Иногда вводится понятие *домена*, или совокупности допустимых значений атрибута (например, поле SEX может иметь только два значения — «м», «ж», поле YEAR — только четырехразрядные числа, начинающиеся с 19 — если в БД речь идет о родившихся в XX в. и пр.);
- *открытый* (opened) *файл* — файл, доступный в данный момент данному приложению. Открытие файла создает в памяти буфер, в который с внешнего накопителя считываются записи. В разные моменты времени могут быть открыты различные множества файлов, количество открытых файлов обычно стараются ограничить, чтобы не расходовать оперативную память;
- *активный или текущий* (current, active) — тот из открытых файлов, который обрабатывается в данный момент времени. Все операции над файлами (добавление записи, удаление записи, редактирование записи) адресуются именно к активному файлу;
- *активная или текущая запись* — запись открытого файла (рис. 2.3), доступная для обработки в данный момент времени (редактирование, ввод полей, корректировка, удаление);
- *указатель текущей записи* — физический номер доступной записи. Текущая запись находится в оперативной памяти. При переходе к другой записи данного файла указатель записи изменяется, и содержание оперативной памяти замещается со-

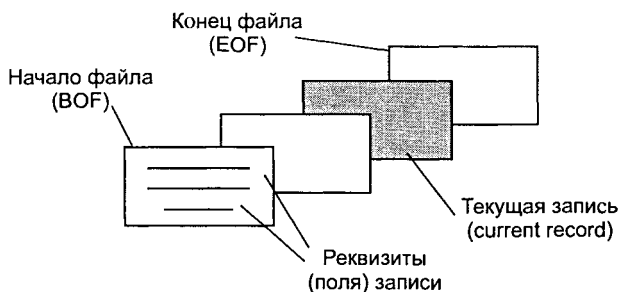


Рис. 2.3. Основные понятия, связанные с отдельным файлом табличной БД

держимым новой текущей записи. Подразумевается, что если в командах или программах фигурируют имена некоторых полей (в качестве аргументов функций или выражений), то их значения соответствуют содержанию текущей записи текущего файла;

- каждый файл и каждая запись могут в широких пределах обрабатываться независимо друг от друга (за исключением ситуаций проверки соответствия записей друг другу или *целостности БД*);
- *навигация в БД* — последовательность действий приложения (программы или пользователя в процессе диалога), при которой осуществляются изменения состояния файлов и записей (открытых, текущих файлов, активных записей). Изменение содержимого файлов при навигации не обязательно. В процессе навигации просматривается или редактируется содержимое БД.

Вид представления записей на экране может быть не только табличным (отчет, запись в строке), но и картотечным (форматированный экран, запись на экране).

В последнем случае каждая запись выводится в виде определенной формы. Структура формы одинакова для всех записей, причем название полей соответствует названиям столбцов табличной формы представления базы данных, а их расположение задается пользователем.

Занесенную в базу данных информацию можно обрабатывать, осуществляя следующие операции:

- сортировку по любому столбцу (по возрастанию/убыванию чисел, символьных строк, дат);
- поиск по любому столбцу с различными условиями (равно, больше, меньше и т. д.);

- соединение таблиц с целью построения выходных форм или отчетов, включающих связанную информацию по каждому объекту.

Операции над БД (навигация) осуществляются с помощью следующих команд СУБД (или СП):

- открыть (закрыть) файл (таблицу);
- выбрать одну из таблиц как активную (текущую);
- перейти к определенной (по номеру) записи (строке таблицы);
- перейти к предшествующей (последующей) строке таблицы;
- перейти в начало (конец) файла (таблицы);
- найти строку таблицы с определенным сочетанием значений атрибутов;
- добавить (редактировать, удалить) строку таблицы (запись файла).

Кроме того, в системе команд присутствуют операции администрирования (обслуживания) базы данных:

- создать файл (таблицу) БД;
- установить (снять) связь нескольких таблиц по логическим критериям;
- создать описание отчета (поколонная выдача информации на экран или печать из одной или нескольких взаимосвязанных таблиц);
- создать формат экрана (построчной выдачи информации на экран из одной или нескольких взаимосвязанных таблиц);
- изменить структуру элемента БД (таблицы данных, формата экрана, отчета и пр.).

Таблица 2.2. Сравнительный анализ средств реализации документальных и фактографических АИС

Объекты или операции	STAIRS	ADABAS	FoxPro	Jurius	Irbis	ORACLE/SQL
1	2	3	4	4	5	6
1. Физическая структура БД						
1.1. Прямые файлы	MASTR TEXTS	DATA	Файлы с расширением .dbf (dbf-файлы)	szdoc.dbf – полные тексты; sf1.dbf – регистрационные карты	Файлы с расширением .dat (dat-файлы)	Табличные данные (помещаются в табличные пространства)
1.2. Частотные словари	DICTN	ASSO	Нет	dclrv.dbf ixfrv.dbf	dic-файлы, alf-файлы	Нет
1.3. Инверсные списки	VOCAB	ASSO	idx-файлы	dcind.dbf ixind.dbf	inv-файлы, ptr-файлы	Индексные данные

Продолжение табл. 2.2

Объекты или операции	STAIRS	ADABAS	FoxPro	Jurius	Irbis	ORACLE/SQL
1	2	3	4	4	5	6
1.4. Описание структуры БД	DBD, ТОФП, ТОТП	FDT, DDM	dbf-файлы	ixddm.dbf	dbd-файлы	Словарь базы данных
2. Логическая структура БД						
2.1. Простые поля	Форматные	Элементарные	Все поля dbf-файлов	Короткие поля	Форматные	INT, REAL, CHAR, DATE
2.2. Сложные поля	Текстовые	Множественные MU групповые PE	MEMO-поля (без поиска)	Множественные (текстовые)	Текстовые	VARCHAR, TEXT, IMAGE, BINARY, VARBINARY
2.3. Инвертируемые поля	Текстовые	Любые (утилита INVERT)	Любые (команда INDEX)	Любые (описание в ixddm.dbf)	Все поля	CREATE INDEX ...
3. Пользовательские интерфейсы						
3.1. Открыть (закрыть сеанс, базу данных)	..DIAL INIS ..OFF	OP FILE, CL FILE, USE FILENAME, CLEAR ALL	DO, QUIT	Пункт меню БАЗА ДАННЫХ	START, QUIT	-
3.2. Инверсный поиск записей	..SEARCH	FIND WITH	SEEK FIND	DC-все поиски, IX – описано в ixddm	Все поиски	SELECT FROM... WHERE...
3.3. Сканирующий поиск	..SELECT	FIND WHERE	LOCATE (CONTINUE), EDIT, FOR, WHILE	Контекстный поиск в ПТ	Нет	SELECT FROM... WHERE...
3.4. Чтение в физическом порядке	..BROWSE	READ BY ISN READ PHYSICAL	SET INDE TO SKIP ,EDIT, зада- ется в запросе	Пункт меню СОРТИРОВАТЬ	ORDER BY	-
3.5. Чтение в логическом порядке, группирование записей (прерывание)	.. BROWSE	READ BY, READ LOGICAL, SORTED.. (AT BREAK...)	SET INDEX TO SORT BY	Задается в запросе	СОРТИРОВАТЬ	ORDER BY, (GROUP BY...)
3.6. Удаление записи, вставка записи, изменение записи	Для пользова- теля – нет	DELETE, STORE, UPDATE	DELETE, INSERT, REPLACE BY...	Для пользова- теля – нет	Для пользова- теля – нет	DELETE, INSERT, UPDATE
3.7. Вывод строки данных	Нет	WRITE	SAY, ?,??	Вывод страниц ПТ	ПЕЧАТЬ	SQL*REPORT
3.8. Задание формата выдачи	..BROWSE	DISPLAY	CREATE FORM, CREATE REPORT	Формат 1 (одна РК на экран) Формат 2 (одна РК на строку экрана)	СПИСОК ФОРМАТОВ (СХЕМ)	SQL*FORM, SQL*REPORT
3.9. Вывод отчета	..BROWSE	DISPLAY	REPORT	Вывод РК или текста	ПРОТОКОЛ	SQL*FORM, SQL*REPORT

Окончание табл. 2.2

Объекты или операции	STAIRS	ADABAS	FoxPro	Jurius	Irbis	ORACLE/SQL
1	2	3	4	4	5	6
3.10. Вывод/ввод экрана		INPUT/ REINPUT	EDIT		ПЕЧАТЬ, ВЫВОД В ФАЙЛ	SQL*FORM, SQL*REPORT
3.11. Просмотр словарей	..STEM	HISTOGRAM	Нет	Главное меню	СЛОВАРЬ	
3.12. Агрегатные функции	Подсчет числа выданных документов	MAX, MIN, AVER, COUNT	Задаются при создании формата	Нет	Нет	MAX, MIN, AVER, COUNT
3.13. Системные функции		*DATE, *TIME.	DATE(), TIME()	Нет		DAY, MONTH
4. Возможности администратора						
4.1. Создание файла, базы данных	READER	LOADER	CREATE STRUCTURE	Редактирование ixddm.dbf	Утилита IRBISDDM	CREATE DATABASE, TABLE
4.2. Удаление файла	Нет	DBMOD	Команда ОС DEL	Нет	- - -	DROP TABLE
4.3. Создание/модификация структуры	Нет	LOADER FILEMOD	CREATE/MODIFY STRUCTURE	Редактирование ixddm.dbf	- - -	ALTER TABLE (ALTER, DROP COLUMN)
4.4. Создание подсем	Команда ..BROWSE	MAINT	CREATE/MODIFY VIEW, SCREEN	Редактирование sf1.fmt	- - -	CREATE VIEW
4.5. Загрузка данных в БД	READER	LOADER, ADDLOAD	APPEND	Программы ixbinr6, ixmst16	- - -	SQL*LOADER
4.6. Внешний формат данных	Карточный формат, формат с разделителями	Системный формат ADAWAN	Системный формат или формат с разделителями	Файлы sf1 и szdoc во внешнем формате FoxPro	Карточный формат, или формат МЕКОФ	Стандартные файлы операционной системы

Система управления базами данных ADABAS

СУБД ADABAS (Adaptable Data Base System) разработана и развивается Американско-германским предприятием Software AG of North America (SAG) и представляет собой систему программирования ИС классической структуры [31].

ADABAS — *постреляционная* система управления базами данных компании, работает в приложениях, которые предполагают обработку больших объемов данных, а также большое количество одновременно работающих пользователей (Online Transaction Processing — OLTP:). ADABAS является одной из наиболее распространенных СУБД на мэйнфреймах (больших ЭВМ).

ADABAS работает на платформах Windows NT, Unix, Open VMS, AS/400, BS2000, MVS, VSE, VM, OS/390 и др. Он поддерживает раз-

личные механизмы параллельной обработки, в частности SMP (Symmetric Multi-Processing) и Sysplex-архитектуру фирмы IBM.

ADABAS разработан и оптимизирован таким образом, чтобы обеспечить максимальную производительность для критически важных приложений, используя относительно небольшое количество вычислительных ресурсов. ADABAS разработан для средних и больших предприятий, которым требуется СУБД, позволяющая работать тысячам пользователей при необходимости быть в рабочем состоянии 24 ч в сутки 7 дней в неделю (24×7) с минимальным временем простоя.

Иллюстрацией может служить Европейское правительственное агенство, где с одной ADABAS-системой *одновременно* работают 30 000 пользователей, используя мощное финансовое приложение, которое обрабатывает около 200 млн запросов в день.

Основные конструктивные особенности, обеспечивающие высокую эффективность СУБД ADABAS:

- вложенные отношения позволяют уменьшать схему базы данных и увеличивать количество данных, передаваемых за одну операцию ввода-вывода;
- автоматическое не зависящее от платформы сжатие данных требует меньшего объема памяти для хранения и позволяет оптимизировать процедуры ввода-вывода информации;
- обусловленное спецификой приложений использование различных типов памяти ЭВМ сокращает время обработки данных и уменьшает число операций ввода-вывода;
- использование многопроцессорных систем позволяет увеличить производительность;
- блокировки доступа к данным на уровне строки (записи) в многопользовательском режиме снижают коллизии доступа к базам данных и улучшают условия для параллельной обработки данных.

Подобно PCУБД, ADABAS хранит данные в таблицах, где строкам соответствуют отдельные записи, а столбцам — поля.

Отличия ADABAS от PCУБД заключаются в следующем:

- отношения могут храниться как вложенные отношения/таблицы, что приводит к сокращению использования ресурсов ЭВМ по сравнению с традиционными PCУБД;
- поддерживаются иерархические поля с возможностью иметь до 200 экземпляров значений такого поля внутри одной записи.

Способы хранения и доступа к данным в ADABAS отделены от особенностей конкретных физических носителей, что делает его гибким и эффективным.

Технология, лежащая в основе ADABAS, прекрасно соответствует новым требованиям, например, в области мультимедиа-данных, в области управления документами и других. ADABAS поддерживает различные модели и структуры данных, которые могут быть специально спроектированы для работы с разными видами приложений. Модуль ADABAS Text Retrieval (TRS) предназначен для хранения и поиска свободных, неформатированных текстов. ADABAS SQL Server дает возможность использовать ADABAS в рамках стандарта ANSI SQL. Доступ к данным, хранящимся в ADABAS, можно осуществлять с помощью:

- Natural (4GL — язык 4-го поколения) в случае OLTP-приложений (основной метод);
- SQL в случае не-OLTP-приложений;
- ODBC в случае Windows-приложений;
- JDBC для Java-приложений.

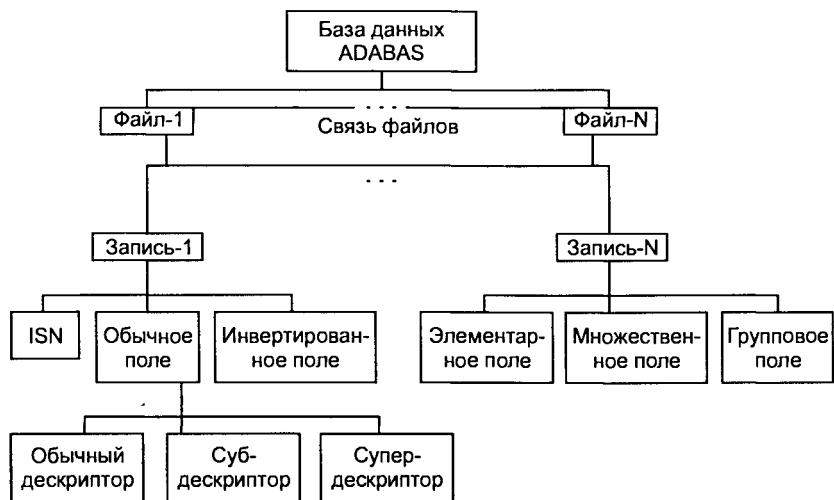
Приложения, разработанные на Natural, легко масштабируются и переносятся на другие платформы. Natural является прекрасным средством для разработки распределенных приложений. С помощью Natural удобно работать не только с ADABAS, но и со всеми основными СУБД на мэйнфреймах, Unix и Windows-системах. Natural поддерживает стандарт SQL2 и имеет возможность использовать диалекты основных РСУБД.

Для упрощения интеграции серверов приложений с персональными компьютерами NaturalX поддерживает технологию DCOM фирмы Microsoft, что делает возможным использование компонентного подхода при разработке информационных систем. Созданные ранее приложения могут быть снабжены DCOM-интерфейсами и использованы в качестве компонентов при разработке новых информационных систем. Кроме того, к приложениям, написанным на Natural, можно получить доступ через Web и Интернет.

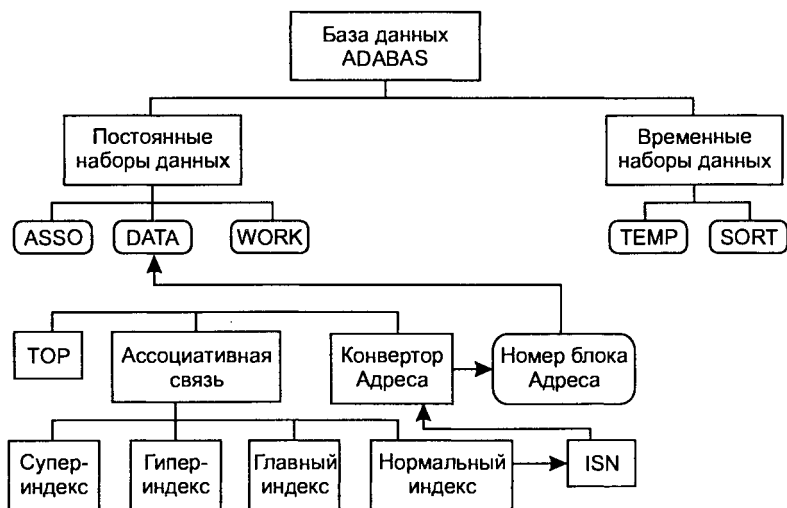
Логическая структура БД ADABAS. База данных ADABAS может рассматриваться как в логическом, так и в физическом аспектах (рис. 2.4, табл. 2.2). На логическом уровне БД представляет собой совокупность логических файлов, содержащих записи, включающие поля различных типов.

Логический файл (ADABAS-файл) — набор данных, состоящий из записей одинакового формата, в совокупности с описанием этих данных и их взаимосвязей. Файлу соответствует уникальный номер.

Запись — совокупность логически связанных данных, ISN (Internal Sequence Number, или ВНЗ — внутренний номер записи) — уникальный для данного файла номер записи, который СУБД назначает всякой новой записи.



а)



б)

Рис. 2.4. Структура БД:

а — логическая; б — физическая

Поле записи — низший уровень, на котором данные определяются в записях логических файлов.

Имя поля (внутреннее имя поля) — двухсимвольный идентификатор, уникальный для файла.

Групповое поле — последовательность нескольких рядом расположенных полей, определенных как целое, с одним именем.

Периодическая группа — групповое поле, допускающее повторение. Каждая реализация группы (либо элементарного поля, входящего в группу) идентифицируется индексом.

Периодическая группа может включать множественные поля.

Множественное поле — поле, содержащее более чем одно значение.

Дескриптор (обычный дескриптор) — поле, содержание которого инвертируется, запоминается в Ассоциаторе и используется в командах FIND, READ LOGICAL, HISTOGRAM.

Субдескриптор — определенная часть, порция поля, содержимое которой инвертировано и хранится в Ассоциаторе.

Супердескриптор — соединенные вместе определенные порции нескольких полей, образующие некоторое значение, которое инвертируется. Поля, не являющиеся дескрипторами, могут являться источниками супердескрипторов и субдескрипторов.

Связанные файлы — в системе может быть реализовано физическое связывание ADABAS-файлов, после осуществления которого появляется возможность «связанного поиска», при котором в поисковом критерии используются поля нескольких файлов (см. FIND ... COUPLED в описании Natural).

Перечисленные выше понятия относятся к физическому уровню доступа к данным по номеру файла и внутреннему имени поля, осуществляемому обращением к БД через ядро из программы пользователя. Средства обрамления ядра используют в своей работе те или иные разновидности словаря данных.

Словарь данных — описание на логическом уровне файлов, полей и форматов с целью управления представлением данных и обращения к ним в средствах обрамления ядра.

Физическая структура БД ADABAS. Физически БД включает пять наборов данных на МД: три обязательных (ASSO, DATA, WORK) и два вспомогательных, используемых утилитами системы.

Ассоциатор (ASSO) — часть физической БД, используемая ядром системы для определения адресов записей.

Ассоциатор включает следующие компоненты:

1. Ассоциативную сеть, или группу инвертируемых списков, каждый из которых представляет собой совокупность внутренних номеров записей (ISN), построенную для одного значения поискового инвертированного поля (дескриптора), т. е. перечень записей данного файла, имеющих определенное значение некоторого дескриптора. Доступ к инвертируемым спискам осуществляется по имени и

значению дескриптора посредством многоуровневой индексной таблицы. Индексная таблица ассоциатора ведется автоматически, корректируясь при каждом изменении значений дескрипторов в файлах, добавлении и удалении записей;

2. Конвертор адресов, вычисляющий по ISN физический адрес блока хранения записей в области DATA.

3. Таблицу описания полей (ТОП — FDT) — средство определения структуры записи файла СУБД, такая информация о полях, как имя, длина, тип, вводится и запоминается при первоначальной загрузке файла.

3. Область хранения данных (DATA) — часть БД, используемая для размещения хранимых записей. Записи хранятся в сжатом формате, в виде строк переменной длины. Сжатие данных скрыто от пользователя, поскольку СУБД выдает поля в представлениях пользователя. Принцип упаковки данных: исключение замыкающих пробелов в текстовых полях и ведущих нулей в текстовых полях.

4. Рабочую область (WORK) — используется для размещения промежуточных результатов при работе Ядра системы.

5. Временные необязательные наборы (TEMP и SORT) используются для сортировки дескрипторов и размещения промежуточных данных при работе утилит.

Форматы языка Natural. Пользователю Natural предоставляется в распоряжение язык программирования (часто именуемый также Natural), позволяющий готовить исходные программные модули, работающие с БД и реализуемые в пакетном или диалоговом режимах. Пользователь может выполнять отдельные команды (операторы) в режиме транзакций.

Конструкции языка Natural строятся с использованием операндов и операторов.

Операндами могут быть:

- файлы и поля БД;
- системные переменные: *ISN (номер обрабатываемой записи), *NUMBER (число записей, выбранных операторами FIND, HISTOGRAM) и др.;
- стандартные системные функции: MAX, MIN, COUNT, AVER (максимальное, минимальное, среднее значение арифметического поля или переменной по всем или части прочитанных записей), *DATE, *TIME (текущая дата, время) и др.;
- числовые, символьные, шестнадцатеричные строки и константы;
- числовые, символьные, шестнадцатеричные переменные, объявленные программистом.

Операторы языка разделяются на следующие группы:

- спецификация режима работы — SET GLOBALS;
- выбор данных из БД — FIND, READ, HISTOGRAM;
- ввод/вывод — INPUT, REINPUT, WRITE, DISPLAY, NEWPAGE, SKIP и пр.;
- управление программой — ACCEPT/REJECT, END, FOR, LOOP, DO/DOEND;
- условия — AT BREAK OF, AT START/END OF DATA, IF, AT TOP/END PAGE;
- арифметика/назначения — COMPUTE, COMPRESS, RESET, MOVE и пр.;
- модификация БД — DELETE, UPDATE, STORE, а также ряд других.

Ниже приведены краткие описания форматов языка с использованием IBM-нотации: квадратные скобки [] означают необязательность присутствия указанного в них элемента конструкции; фигурные { } — обязательность выбора одной из альтернатив; многоточие ... — повторение элемента.

Оператор SET GLOBALS управляет установкой глобальных опций программы. Основные из них:

- LS (LineSize) — размер строки выходного отчета (1—132);
- MT (MaxTime) — максимальное процессорное время, используемое программой (с);
- SF (SpaceFactor) — расстояние между колонками выходного отчета;
- PS (PageSize) — размер страницы отчета (1—250);
- LT (Limit) — максимальное число записей, считываемых командами READ, FIND или HISTOGRAM и др.

Выборка записей:

$$\text{READ}(\text{"ограничение"}) \text{"файл"} \text{BY} \left\{ \begin{array}{l} \text{ISN} \\ \text{" дескриптор"} \end{array} \right\} = \left\{ \begin{array}{l} \text{" константа - 1"} \\ \text{" поле - 1"} \end{array} \right\}$$

$$\text{THRU} \left\{ \begin{array}{l} \text{" константа - 2"} \\ \text{" поле - 2"} \end{array} \right\}$$

Оператор READ задает считывание записей из «файла» в физическом (по возрастанию ISN) или логическом (по возрастанию значений «дескриптора») порядке в интервале от «поле-1» до «поле-2»; при необходимости число обрабатываемых записей задается «ограничением».


```
HISTOGRAM("огр.") "файл" "дескр." STARTING { " конст - 1" } THRU
{ " конст - 2" }
{ " поле - 2" }
```

Оператор считывает значения «дескриптора» из «файла» и определяет частоты этих значений (при необходимости они могут быть извлечены и обработаны); количество значений регулируется так же, как и в READ:

```
FIND("ограничение") "файл-1" WITH "критерий поиска"
[ ( AND
  OR ) COUPLED "файл - 2" WITH "критерий поиска..." ]
[ SORTED BY "дескр-1" "дескр-2" ... ] [ DESCENDING ]
[ WHERE "логический критерий" ]
[ IF NO RECORD FOUND "оператор" ]
```

Форматы «критерия-поиска» являются стандартными для всех интерфейсов ADABAS и отражают логику работы Ядра:

$$\begin{aligned}
 \text{"крит-поиска"} &::= [\text{NOT}] \text{"крит-поиска"} \left\{ \begin{array}{l} \text{AND} \\ \text{OR} \end{array} \right\} [\text{NOT}] \text{"крит-поиска"} \\
 \text{"крит-поиска"} &::= \text{"дескр"} \left\{ \begin{array}{l} \text{NE} \\ \text{EQ} \\ > \\ < \\ \text{GE} \\ \text{GT} \end{array} \right\} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\} \\
 &= \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\} \text{THRU} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\} [\text{BUT NOT} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\}] \\
 &= \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\} \text{OR} = \left\{ \begin{array}{l} \text{" поле" } \\ \text{" конст." } \end{array} \right\} [\text{OR} = \dots]
 \end{aligned}$$

Оператор инициирует цикл обработки, заканчивающийся либо с исчерпанием всех записей «файла», удовлетворяющих «критерию поиска» и/или «логическому критерию» как в основном («файл-1») так и в зависимом («файл-2») файле (если в БД существуют связанные файлы). При необходимости записи сортируются по 1—3 дескрипторам «файла-1» по возрастанию или убыванию значений. Если в «файле-1» нет соответствующих значений, может предусматриваться выполнение «оператора».

«Логический критерий» является обобщением «критерия-поиска» и разрешает присутствие в правой части выражения не только

«дескриптора», но и простого «поля» или переменной пользователя. Кроме того, этот тип логического выражения допускает сканирование или маскирование (поиск в левой части выражения строки, содержащейся в левой части).

Кроме оператора FIND «логический критерий» используется в конструкциях IF.

Ввод/вывод данных (рассматриваются только некоторые операторы вывода и управление ими).

Оператор FORMAT устанавливает опции отчета:

- ES (EmptySuppress) — подавление печати пустых строк;
- IS (IdenticalSuppress) — заполнение пробелами идентичных значений в столбце;
- MP (MaxPages) — максимальное число страниц в отчете;
- ZP (ZeroPress) — подавление печати нулей;
- AL, NL (AlphabeticLength, NumericLength) — задание ширины колонки отчета для символьных и численных данных.

Оператор DISPLAY имеет следующий основной синтаксис:

$$\text{DISPLAY} \left\{ \left\{ \begin{array}{l} \text{NOTITLE} \\ \text{NOHDR} \end{array} \right\} \right\} (\text{"опции отчета"}) \text{"элемент вывода"} \dots$$

$$\text{"элемент вывода"} = \left\{ \left\{ \begin{array}{l} nX \\ nT \\ / \end{array} \right\} \left[\text{'текст'} \right] \left\{ \begin{array}{l} \text{" поле"} \\ \text{" переменная"} \end{array} \right\} \left\{ \text{"опции поля"} \right\} \right\}$$

Здесь — «опции-отчета» и «опции-поля» в основном совпадают с опциями оператора FORMAT, однако они распространяются на данный отчет и/или конкретные поля; Операторы NOTITLE, NOHDR запрещают печать заголовка отчета и столбцов соответственно; nX, nT управляют пропуском n символов и табуляцией до позиции n;

«текст», помещенный непосредственно перед «полем», переопределяет заголовок столбца (HDR);

«/» задает перевод строки и соответственно вертикальную печать значений полей в столбце. Модификация WRITE отличается от оператора DISPLAY тем, что разрешает переполнение строки отчета и не обеспечивает автоматическую генерацию заголовка. WRITE TITLE/TRAILER задает печать выводимых строк в начале/конце каждой страницы отчета (рис. 2.5).

Управление программой. Оператор END задает конец программы и закрывает все незакрытые циклы, LOOP — закрывает

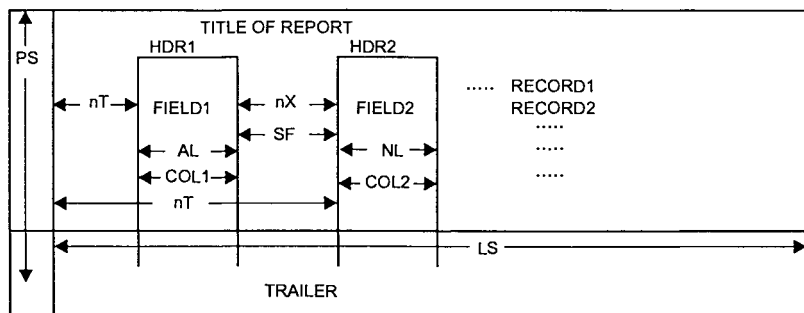


Рис. 2.5. Основные понятия, связанные со структурой страницы отчета. Здесь HDR — или «текст» из оператора DISPLAY, или название переменной или имя поля из Словаря Данных

соответствующий цикл, группа DO/DOEND задает блок связанных операторов:

```
DO "оператор -1" , ..... , "оператор-N" DOEND.
```

ACCEPT/REJECT IF "логич. критерий" — продолжение/прекращение обработки записи.

Оператор FOR имеет следующий формат:

$$\text{FOR} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" переменная" } \end{array} \right\} \left[\begin{array}{l} \left\{ \begin{array}{l} = \\ \text{FROM} \\ \text{[STEP]} \end{array} \right\} \left\{ \begin{array}{l} \text{" константа" } \\ \text{" поле" } \\ \text{" целая - конст. " } \end{array} \right\} \left\{ \begin{array}{l} \text{TO} \\ \text{THRU} \end{array} \right\} \left\{ \begin{array}{l} \text{" целая - конст. " } \\ \text{" поле!" } \end{array} \right\} \end{array} \right] \text{" оператор"}$$

Операторы условий:

IF "логич. критерий" [THEN] "оператор" — основной условный оператор;

AT $\left\{ \begin{array}{l} \text{START} \\ \text{END} \end{array} \right\}$ OF DATA "оператор" — выполнение оператора

перед считыванием первой записи цикла или после считывания последней записи;

AT $\left\{ \begin{array}{l} \text{TOP} \\ \text{END} \end{array} \right\}$ OF PAGE "оператор" — выполнение действий в начале (конце) страницы отчета;

AT BREAK OF $\left\{ \begin{array}{l} \text{" поле" } \\ \text{" перем." } \end{array} \right\}$ "оператор" — выполнение действий

при прерывании, т. е. изменении значения «поля» или «переменной» при сортировке записей или чтении в логическом порядке.

При прерывании, т. е. изменении значения «поля» или «переменной» при сортировке записей или чтении в логическом порядке.

Арифметика и пересылки:

RESET — инициализация полей;

COMPUTE "переменная" = "выражение" — вычисление и присвоение значения «выражения» «переменной». При построении выражения используются знаки (,), +, -, * (умножение), ** (возведение в целую степень), и функция SQRT (извлечение квадратного корня).

Оператор COMPRESS имеет формат:

$$\text{COMPRESS} \left\{ \begin{array}{l} \text{" конст. " } \\ \text{" поле" } \\ \text{" перем. " } \end{array} \right\} \dots \text{INTO} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" перем. " } \end{array} \right\} [\text{LEAVING NO SPACES}]$$

и осуществляет пересылку (сжатие) нескольких операндов в алфавитно-цифровое поле. Опция LEAVING NO SPACES подавляет разделение операндов пробелом.

Оператор REDEFINE:

$$\text{REDEFINE} \left\{ \begin{array}{l} \text{" поле" } \\ \text{" перем. " } \end{array} \right\} \left\{ \left\{ \begin{array}{l} \text{nX} \\ \text{" поле" } \\ \text{" перем. " } \end{array} \right\} \left\{ \begin{array}{l} \text{nX} \\ \text{" поле" } \\ \text{" перем. " } \end{array} \right\} \dots \right\}$$

используется для переопределения левого оператора по байтам слева направо с помощью оператора списка правой части. Для обозначения символов-заполнителей, которые в переопределяемом поле должны пропускаться, используется запись nX.

Системы программирования dBase, Clipper, FoxBase, FoxPro

Система программирования FoxPro предназначена для разработки приложений, являющихся открытыми или замкнутыми документальными или фактографическими информационными системами, и впервые предложена в 1988 г. Предшествующими аналогичными системами являлись — dBaseIII/IV и Clipper (1984—1987); последующей — FoxPro (1990 г.), в настоящее время распространены FoxPro for Windows и Visual FoxPro.

Аналогичность здесь подразумевает преемственность, основанную на том, что различие команд, системы функций не превосходит 10—15%. Ранее разработанные приложения не требуют существенных переработок, навыки программирования, полученные на Clipper имеют силу для FoxPro и т. п.

Основные понятия, связанные с работой в FoxPro:

- *файл данных* — файл операционной системы (ОС), имя которого содержит расширение .dbf, содержащий как описание структуры записей, так и собственно записи с информацией (см. также рис. 2.2);

- *запись файла* — совокупность связанных данных, состоящая из полей;
- *указатель записи* — текущий номер активной записи, доступной для чтения и обновления;
- *текущий файл БД* — выбранный для обработки один из открытых файлов;
- *индексный файл (индекс)* — файл ОС с расширением .idx, .tdx, .ndx, предназначенный для управления порядком обработки файла БД. Одному файлу БД может соответствовать несколько существующих и/или активных индексных файлов;
- *текущий индекс* — открытый индексный файл, выбранный для управления текущим файлом БД; выборка данных из БД осуществляется по возрастанию ключа (индексного выражения, вычисляемого по полям текущего файла), соответствующего текущему индексу;
- *формат экрана* — файл с расширением .fmt, содержащий описание порядка выдачи данных на экран монитора (чтения данных с экрана);
- *формат отчета* — файл ОС с расширением .fmt, содержащий описание отчета, выдаваемого на экран или принтер (длина строки, ширина страницы, имена выдаваемых полей, заглавие отчета, имена колонок, печать итогов при прерывании и т. д. (см. также рис. 2.5);
- *командный (пакетный, программный) файл* — файл с расширением .prg, содержащий в каждой строке одну из команд языка FoxPro;
- *команда* — элемент языка манипулирования данными (близкого к языкам программирования);
- *функция* — встроенный оператор, осуществляющий преобразование переменных, либо получение справочной информации.

Команды и функции языка FoxPro (см. табл. 2.2, а также прил. 2).

А. Команды, позволяющие реализовать функции администрирования БД (создание, загрузку, преобразование файла) или изменения структуры файла, создания или модификации соответственно файла, экрана, отчета:

- $\left. \begin{array}{l} \text{CREATE} \\ \text{MODIFY} \end{array} \right\} \left\{ \begin{array}{l} \text{STRUCTURE} \\ \text{SCREEN} \\ \text{REPORT} \end{array} \right\}$ — создание или модификация соответственно файла БД, формата экрана, отчета — открывают

экранный интерфейс, посредством которого система вступает с пользователем в диалог;

- INDEX ON "выражение" TO "файл" — создание индекса для текущего файла БД, который записывается в «файл».idx; ключом доступа является «выражение» (чаще всего имя поля);
- JOIN, SORT — создание нового файла путем слияния двух существующих по некоторому критерию или сортировки одного файла;
- APPEND FROM "файл" — загрузка (импорт) файла БД из файла ОС, подготовленного в формате записей фиксированной длины, либо же с указанными в команде разделителями;
- COPY TO "файл" — выгрузка (экспорт) данных из файла БД в файл ОС (операция, обратная APPEND FROM).

Б. Команды ввода/редактирования записи (средства интерфейса оператора):

- GET — ввод отдельного поля, строки или пользовательской переменной;
- EDIT — экранная команда ввода/вывода/исправления записи в текущем файле с использованием некоторого формата, описание которого хранится в файле формата экрана;
- BROWSE — экранная команда ввода/исправления записей в текущем файле, при этом используется стандартный формат экрана и режим системного меню;
- APPEND — создание пустой записи в конце текущего файла и заполнение ее с применением пользовательского формата экрана;
- REPLACE — замещение в текущей записи значений некоторых полей на новые;
- UPDATE — массовое обновление значений записей файла БД.

В. Команды вывода данных:

- SAY — выдача на экран или принтер поля, данного или строки;
- BROWSE, EDIT — просмотр данных на экране;
- REPORT — выдача поколонного отчета на экран или принтер с использованием файла формата отчета.

Г. Команды навигации в БД (открытия, закрытия файлов, перемещения по записям файлов);

- USE, SELECT — выбор текущего файла;

- SET INDEX TO, SET ORDER TO — задание текущего индекса при работе с файлом БД;
- GO, SKIP — изменение положения указателя текущей записи путем прямого перехода к записи или путем пропуска некоторого числа записей;
- FIND, SEEK — поиск записи в индексированном файле по ключу и установка указателя на найденную запись;
- LOCATE, CONTINUE — поиск записи по сложному критерию без использования индексов путем сканирования полей записей;
- SET RELATION TO — установление связи двух файлов БД так, что при перемещении указателя записи в основном файле синхронно перемещается указатель в зависимом;
- CLOSE, CLEAR — закрытие всех или части файлов БД, экранов, отчетов.

Д. Команды управления вычислительным процессом:

- DO WHILE "условие" . . . ENDDO — описание блока, выполнение которого осуществляется циклически, пока соблюдается логическое «условие»;
- LOOP — переход из некоторой точки блока DO . . . ENDDO в его начало;
- EXIT — выход из блока DO . . . ENDDO;
- DO, RETURN — вызов командного файла и возврат в вызывающую программу;
- DO CASE . . . CASE . . . ENDCASE — оператор условного ветвления программы на несколько путей;
- IF . . . ELSE . . . ENDIF — оператор условного ветвления на две ветви.

Е. Команды установления режимов и опций:

- SET FORMAT TO — открытие файла формата ввода-редактирования текущего файла;
- SET COLOR TO — определение атрибутов цветов экрана.

Язык программирования FoxPro допускает использование также значительного числа функций, обычных для прикладных языков (MIN, MAX, SUBSTR, SQRT), так и специфических (RECNO — номер текущей записи, RECCOUNT — число записей в текущем файле, EOF — признак конца текущего файла и др.).

Возможности программирования в FoxPro расширяются тем, что многие из процедурных команд обработки и выдачи информации EDIT, REPLACE, REPORT, COPY и др. включают в свою структуру кон-

струкции, уточняющие свойства записей файла, на который эти команды распространяются:

КОМАНДА ["интервал"] WHILE ["условие-1"] FOR ["условие-2"]

где "интервал" = $\left\{ \begin{array}{l} \text{NEXT } n \\ \text{REST} \\ \text{ALL} \end{array} \right\}$ – следующие n записей;
 – оставшиеся до конца файла записи;
 – все записи.

Выполнение КОМАНДЫ прекращается, если нарушается «условие-1». КОМАНДА распространяется при этом только на записи, удовлетворяющие «условию-2».

Разработка приложений в среде FoxPro состоит в создании совокупности взаимосвязанных файлов БД, форматных, индексных, командных, образующих в итоге среду пользователя, АБД, оператора подготовки данных.

2.2. Документальные системы

Документальные системы (предназначенные для обработки, поиска, представления полнотекстовых документов или справочно-реферативной информации) ведут свое происхождение от библиотечно-реферативных служб или информационных центров, выпускающих реферативную информацию (обзоры, экспресс-информацию, реферативные журналы).

Происхождение документальных АИС

Первоначально, когда системы научно-технической информации (СНТИ) начали выпускать *указатели, справочно-библиографические издания* и др., все они предназначались собственно для удовлетворения потребностей информационных работников, облегчения оперирования большими массивами информации. По мере осознания полезности данных вторичных документов системы стали использовать научно-технические работники, что привело к необходимости развития индустриальных методов выпуска указателей, реферативных журналов, каталогов.

Подготовка данных для ввода в фотонаборные машины при выпуске изданий в СНТИ с помощью ЭВМ привела к накоплению массивов вторичных документов на машиночитаемых носителях.

Следующим естественным шагом развития явилось непосредственное использование таких массивов в АИПС.

В перечне традиционных форм информационного обеспечения выделяют реферативный журнал (РЖ) как один из важнейших, по крайней мере, самый массовый продукт СНТИ. На рис. 2.6 приводится пример РЖ Всероссийского института научной и технической информации РАН (ВИНИТИ). Аналогичную структуру имеют и другие реферативные издания, например РЖ INIS (см. рис. 1.4).

Годовой комплект РЖ ВИНИТИ «Техническая кибернетика» состоит из 12 выпусков собственно РЖ и двух указателей. Каждый выпуск содержит около 500 статей-записей, описывающих первоисточники либо их фрагменты (журнальные статьи и др.). Здесь указаны: автор, заглавие, библиографические данные (год, том, номер, классификационный индекс, место издания и др.), реферат; запись имеет уникальный внутренний номер, идентифицирующий внутренний данный документ (реферат) также в авторском и предметных указателях. Латинские символы в скобках на рис. 2.6 соответствуют меткам полей при автоматизированном поиске (см. рис. 2.10).

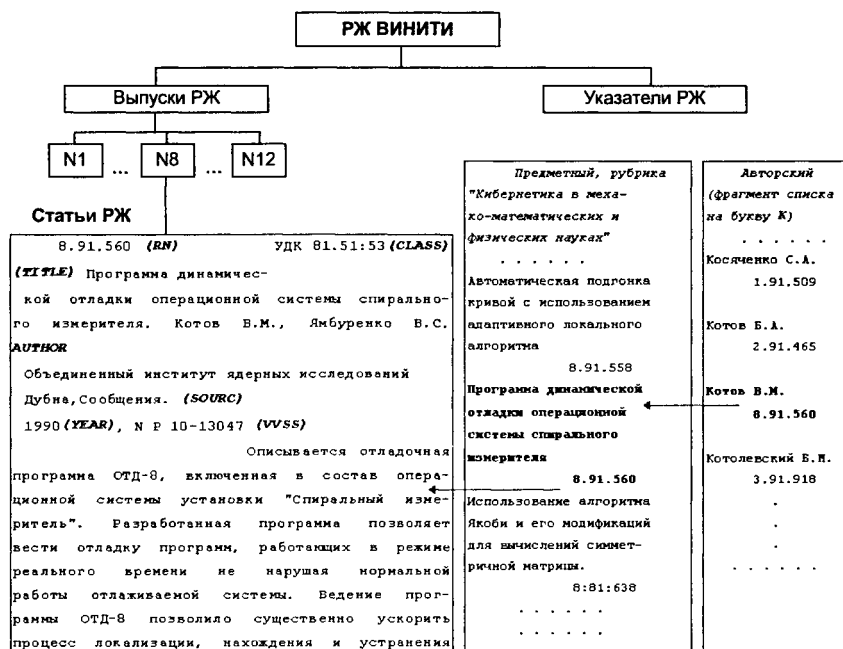


Рис. 2.6. Структура РЖ ВИНИТИ

Системой НТИ является совокупность организаций, осуществляющих обработку *первичных* и создание потоков *вторичных документов*. В РФ Всероссийские органы НТИ (более 20) осуществляют сбор и обработку информации по определенной широкой видовой или тематической группе первоисточников. Наиболее типичными являются:

- ВИНТИ — осуществляет координацию СНТИ РФ, выпуск РЖ, экспресс-информации, обзоров «Итоги науки и техники». С 1979 г. РЖ выпускается в виде машиночитаемых баз данных;
- ИНИОН (Институт Научной информации по общественным наукам РАН) — осуществляет выпуск реферативной информации и баз данных по истории, философии, филологии и другим гуманитарным отраслям знания, базы данных ИНИОН доступны на сервере www.inion.ru;
- ГПНТБ (Государственная научно-техническая библиотека) — осуществляет библиотечное обслуживание, межбиблиотечный абонемент, выпускает обзоры, осуществляет библиографическую обработку зарубежных книг;
- ВНИЦентр (Всероссийский научно-технический информационный центр) — осуществляет сбор, регистрацию и микрофильмирование открытых непубликуемых первоисточников (отчетов, диссертаций и т. д.), организует и обслуживает массивы информационных и регистрационных карт.

Навигация при ручном поиске документальной информации

Рассмотрим деятельность потребителя информации (исследователя, группы) в аспекте поиска и потребления информации. Задачей потребителя является создание массива *релевантных* (потенциально необходимых) сообщений. Это достигается с помощью возможностей, создаваемых системой НТИ, а также ее продуктов — каталогами, указателями, РЖ и т. д.

В обобщенном виде поиск осуществляется по принципам навигации, вкратце описанным на рис. 2.7. Проследим одну из ветвей подобного поиска. Располагая наименованием тематики релевантной информации (например, *кибернетика в механико-математических и физических науках*, см. рис. 2.6), потребитель по предметному указателю определяет номер и заглавие реферата.

В принципе, имея заглавие, можно адресоваться к алфавитному каталогу библиотеки, однако, во-первых, это, как правило, заглавие статьи, а не название журнала, которое собственно и является входом в каталог, а во-вторых, целесообразно по реферату уточнить релевантность статьи. По номеру реферата (8.91.560, см. рис. 2.6)

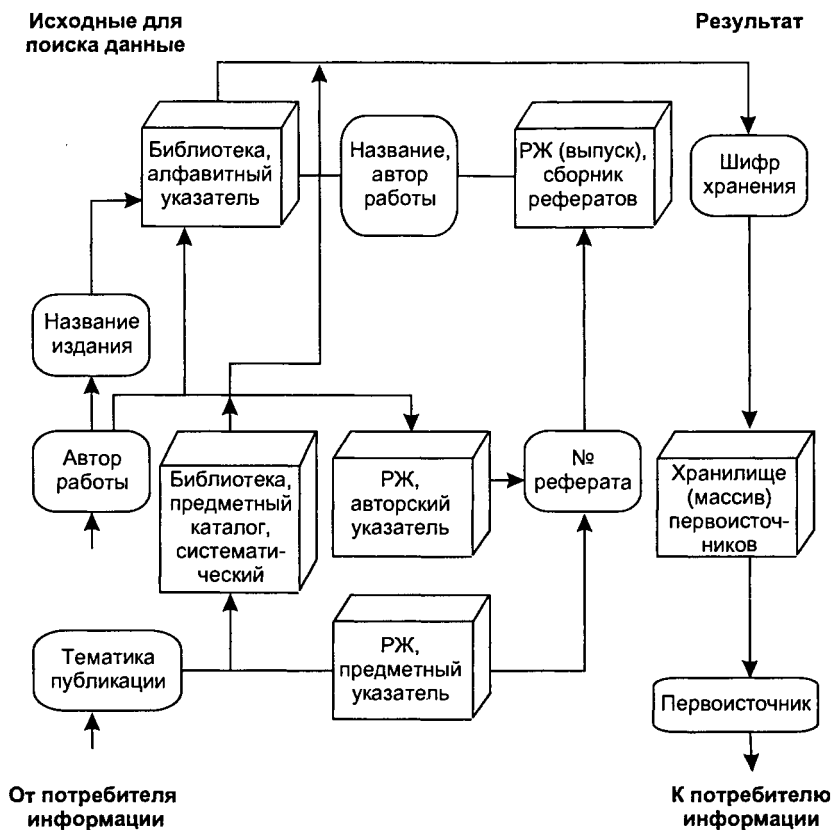


Рис. 2.7. Обобщенная схема навигации при создании информационного обеспечения

пользователь в соответствующем выпуске отыскивает статью, извлекает название издания («Сообщение ОИЯИ». 1990. № P10-13047, см. рис. 2.6), по библиотечному каталогу определяет шифр хранения, заказывает первоисточник в фонде библиотеки.

Очевидно, данная траектория поиска, как и другие, очень трудоемка, и при продвижении по ней накапливаются ошибки, возможны потери и отказы.

Поэтому возникает необходимость автоматизации информационной деятельности как в рамках СНИ (подготовка продуктов и услуг), так и потребителя (улучшение условий их использования). С этой целью создаются *автоматизированные информационно-поисковые системы*.

Исследования в области информационного поиска начались уже более тридцати лет назад. За это время из узкоспециализированной тематики информационный поиск превратился в одну из ключевых областей информатики.

Классические схемы функционирования документальных АИПС

Это режимы ретроспективного поиска (РП) и избирательного распределения информации (ИРИ). Данные схемы являются упорядоченными совокупностями функциональных подсистем реальных АИС, но не отображают структуру потоков и компоненты какой-либо из реальных современных систем и потому относятся к абстрактным АИПС.

Ретроспективный поиск. В АИПС режима РП могут быть выделены три функциональных блока: контура обработки документов, контур обработки запросов, блок поиска и выдачи документов (рис. 2.8).

Обработка первичного документа состоит в присвоении ему уникального номера (аналогичного шифру хранения, но не имеющего классификационных возможностей) и построении поискового

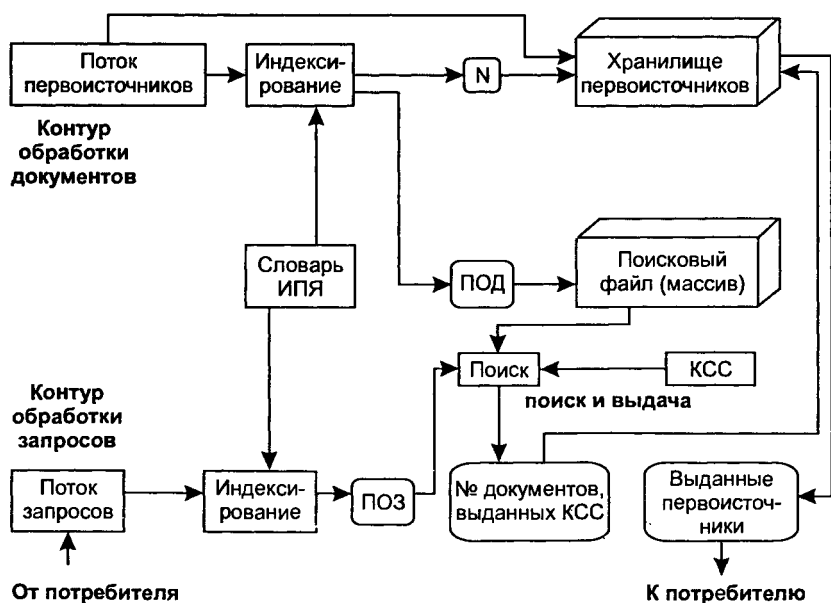


Рис. 2.8. Процессы в абстрактной АИПС (режим ретропоиска)

образа документа (ПОД) — формализованного отображения смысла сообщения на искусственном (информационно-поисковом, ИПЯ) языке. Множество ПОД образует поисковый массив. Прохождение запроса через систему аналогично: поисковый образ запроса (ПОЗ) подвергается обработке совместно с поисковым массивом. Каждый ПОД сопоставляется с данным ПОЗ с помощью критерия смыслового соответствия (КСС). В результате система выдает список номеров формально релевантных документов, соответствующих адресам хранения и являющихся основанием для поиска и окончательной выдачи первоисточников потребителю.

Очевидно, данный режим аналогичен библиотечному поиску (см. рис. 2.7 — тематика — предметный каталог — шифр хранения — первоисточник), а поисковый массив соответствует каталогу; кроме того, здесь поток не сохраняющихся разовых запросов сопоставляется с массивом постоянно хранящихся ПОД.

Избирательное распределение информации. В АИПС режима ИРИ (рис. 2.9), наоборот, поток не сохраняющихся ПОД взаимодействует с массивом постоянно действующих запросов (профилей ИРИ, поисковых образов абонентов — ПОА). В отличие от схемы

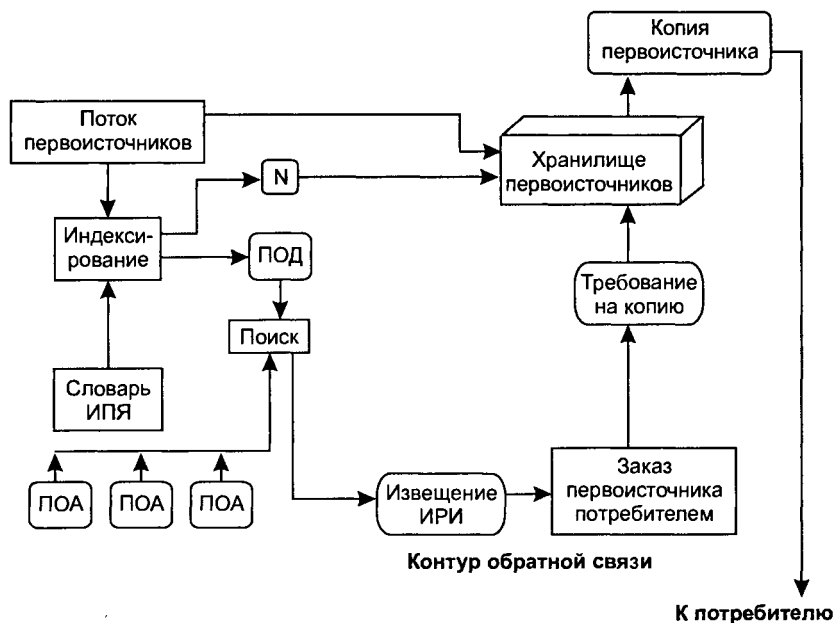


Рис. 2.9. Процессы в АИПС (режим ИРИ)

рис. 2.8, результаты автоматизированного поиска здесь непосредственно преобразуются в выдачу первоисточников (по номеру документа).

Современный период развития технических и программных средств начался в середине 60-х гг., когда корпорация IBM выпустила серию вычислительных установок (ЭВМ) System/360 (затем 370, 375). Программное обрамление ЭВМ составила классическая триада: OS/360, IMS/360, DPS/360:

- *OS (Operating System)* — операционная система, функционально полная и с точки зрения сегодняшнего дня;
- *IMS (Information Management System)* — система управления базами данных, породившая понятие *иерархической модели данных*;
- *DPS (Document Processing System)* — программная оболочка документального поиска.

Через несколько лет система DPS была сменена более совершенным продуктом — STAIRS (STorage And Information Retrieval System — система хранения и выдачи информации). STAIRS унаследовала основные положительные черты DPS и стала прототипом для многих аналогичных разработок вплоть до 1990—2000-х гг.

Логическая структура документальной БД

Ниже описана логическая, физическая структура БД, ЯЗ и навигация в некоторой упрощенной системе, соединяющей черты DPS и STAIRS, которую будем условно называть STAIRS/DPS.

Основные принципы функционирования системы:

- каждый потребитель вступает в контакт с системой на определенный период времени, в процессе которого осуществляется обмен сообщениями, именуемый *диалоговой сессией или сеансом* связи пользователя с системой;
- потребители могут одновременно проводить поиск в нескольких различных БД (поисковых массивах разной тематики, обладающих различными структурами документов). БД, поддерживаемые в системе, могут создаваться как собственными силами организации-пользователя, так и путем загрузки внешних БД, подготавливаемых службами-генераторами на машиночитаемых носителях. Согласование внешних и внутренних форматов документов осуществляется с помощью *конверторов*;
- на период диалоговой сессии за каждым потребителем закрепляются временные наборы данных (ВНД), куда заносятся ре-

зультаты поиска, запросы и учетная информация, и часть содержимого которых может быть сохранена по директивам пользователя в постоянных наборах данных (ПНД) после окончания диалоговой сессии;

- работа системы проходит под наблюдением администратора, который имеет возможность путем вмешательства с главного терминала устранять помехи для нормального процесса работы.

Основной единицей информации в DPS/STAIRS является документ. Под документом здесь понимается описание некоторого объекта (научной статьи, патента, экономического факта или события, кадровой анкеты и т. д.) в определенной структуре. Типовая структура документа, подготовленного на основе БД INIS с точки зрения потребителя, приведена на рис. 2.10. Пример содержания документа приведен на рис. 1.4.

Каждый документ разделяется на форматные и текстовые поля. Форматные поля имеют ограниченную длину и жесткую структуру. Текстовые (называемые иногда параграфами или сегментами) имеют свободную структуру, каждый сегмент может содержать логически неопределенное количество предложений (в отличие от сегмен-

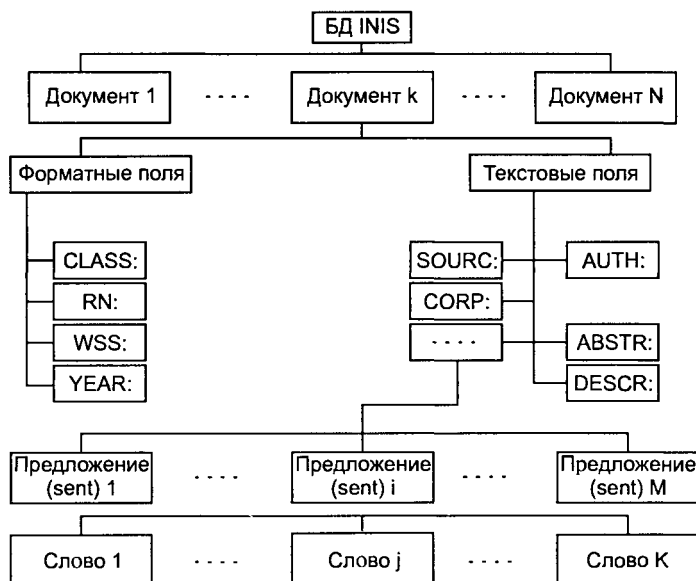


Рис. 2.10. Структура документа БД INIS с точки зрения пользователя системы STAIRS

тов предложения имен не имеют). Каждое предложение содержит неопределенное количество слов.

В рассмотренном примере доступны (см. рис. 2.10 и 1.4):

а) форматные поля:

- CLASS: предметная категория INIS (см. также рис. 1.14);
- RN: номер реферата;
- VVSS: номер тома и номер выпуска (для журналов);
- YEAR: год публикаций;

б) текстовые поля (сегменты, параграфы):

- AUTH: автор (авторы) публикаций;
- TITLE: заголовок оригинала;
- COPR: корпоративный автор, организация, представляющая публикацию;
- SOURC: источник публикации (издательство, журнал, сборник);
- ABSTR: реферат публикации;
- DESCR: список дескрипторов тезауруса INIS, образующих ПОД публикации.

Поисковые возможности АИПС STAIRS

Сеанс диалога состоит из последовательности фаз работы (рис. 2.11), основными из которых являются:

- поиск по текстовым полям (активизируется командой `..SEARCH`);
- поиск по форматным полям (`..SELECT`);
- просмотр найденных документов из текущего списка (`..BROWSE`).



Рис. 2.11. Допустимые переходы между фазами (состояниями) диалога

Текстовый поиск. Запрос, обрабатываемый системой в данном режиме, состоит из учетного номера (N), присваиваемого системой, и предиката (ПОЗ в предикатной форме). Предикат состоит из операторов (логических функций) и операндов.

1. Основные операторы:

- А СТХ В — операнды А и В находятся рядом в заданном порядке в одном предложении;
- А SENT В — А и В — в одном и том же предложении независимо от взаимного расстояния и порядка;
- А SEGM В — А и В в одном сегменте;
- А AND В — А и В в одном документе;
- А OR В — А или В в одном документе;
- А XOR В — А или В, но не А и В (исключающее ИЛИ).

2. Основные виды операндов:

- слова (THERMAL AND REACTOR);
- слова с ограничителями (квалификаторами) — указателями принадлежности (.) или ее отсутствия (..) к сегменту (COOPER.AUTH AND ALUMINIUM AND FERRUM или COOPER..ABSTR AND FERRUM.TITLE.);
- слова с расширителями (THERMAL AND REACT\$); в этом случае второй операнд эквивалентен цепочке REACTION OR REACTIVE OR..., содержащей все термины словаря БД системы, начинающиеся с основы REACT\$;
- учетный номер запроса N, ранее сформулированного пользователем и обработанного системой, — обратная ссылка.

В режиме ..SEARCH после обработки каждого запроса потребителю сообщается статистика встречаемости операндов и их расширений в словаре БД и общее число найденных документов.

Пример:

```
Система: ENTER THE QUERY
000001>
Пользователь: THERMAL AND REACT$
Система:
THERMAL 62 СЛОВОПОЯВЛЕНИЙ 39 ДОКУМЕНТОВ
REACT 103 СЛОВОПОЯВЛЕНИЙ 88 ДОКУМЕНТОВ
REACTION 21 СЛОВОПОЯВЛЕНИЙ 19 ДОКУМЕНТОВ
REACTIONS 12 СЛОВОПОЯВЛЕНИЙ 10 ДОКУМЕНТОВ
REACTIVE 32 СЛОВОПОЯВЛЕНИЙ 26 ДОКУМЕНТОВ
REACTOR 16 СЛОВОПОЯВЛЕНИЙ 11 ДОКУМЕНТОВ
REACTORS 22 СЛОВОПОЯВЛЕНИЙ 12 ДОКУМЕНТОВ
ИТОГ 31 СЛОВОПОЯВЛЕНИЙ 25 ДОКУМЕНТОВ
000002> Пользователь: 00001 AND REACTIVE..TITLE
Система: и т. д.
```

Форматный поиск. Запрос, обрабатываемый при поиске по форматным полям, также состоит из номера и предиката. Операндами предиката являются форматные поля и некоторый номер ранее обработанного запроса, в выдаче на который и проводится отбор. Операторами являются:

- связки EQ (равно), GT (больше), LT (меньше), NE (не равно), OL (вне интервала), WL (внутри интервала), NG (не больше), NL (не меньше);
- логические функции NOT, AND, OR.

Пример:

```
00003> 2 YEAR LT 1999 AND NL 2001 AND CLASS WL A11, A15.
```

На такой запрос (с учетным номером 0003) будут отобраны документы, изданные с 1999 по 2001 г. включительно, входящие в предметные рубрики A11 «Теоретическая физика»,.....A15 «Астрофизика и космология», ранее выданные (отобранные) на запрос с учетным номером 2.

Просмотр найденных документов. При этом пользователь указывает номер запроса (активного списка) и формат выдачи, например:

```
00004> 3 RN, TITL, AUTH (выдать по запросу 3 номер реферата, заголовок, авторов). Система выдает информацию в системном формате (выходной формат STAIRS):
```

```
DOCUMENT 1 OF 25
PAGE 1 OF 1
RN: 512012
AUTH: Tarvin J.A.; Sigler R.D. Busch G.E.
TITLE: Wave-front shearing interferometer for cryogenic
laser-fusion targets.
```

Ниже приводится полный набор команд STAIRS:

а) управляющие команды:

```
DIAL — подключение к системе;
..OFF — отключение от системы;
```

б) главные команды:

```
..CHANGE — выбор БД для последующей работы;
..SEARCH — поиск по текстовым полям документа (результаты заносятся в ВНД);
..RANK — текстовый поиск с ранжированием;
..SELECT — отбор по форматным полям в массиве документов, ранее найденных и хранящихся в ВНД;
..BROWSE — просмотр содержимого ВНД на экране;
```

.SAVE — сохранение отлаженного запроса с заданным учетным номером, под желаемым мнемоническим именем в ПНД системы;

.EXEC — исполнение сохраненного запроса (применяется для реализации ИРИ);

.SORT — сортировка содержимого ВНД по форматным полям;

.HELP — обучение пользователя (для перехода в особую фазу обучения в произвольный момент диалога);

в) служебные команды:

.DISPLAY — отображение запроса и его компонент;

.MAIL — печать результатов поиска;

.PRINT — печать документов просматриваемого списка;

.STEM — просмотр фрагмента частотного словаря системы.

Кроме этого, в любой момент диалоговой сессии пользователь может получить копию содержимого интересующей его части экрана с помощью локального печатающего устройства (аппаратурная возможность).

Физическая структура и навигация в документальной БД

Физическая структура БД рассматриваемой нами «условной» системы приведена на рис. 2.12. и включает в себя четыре файла операционной системы:

DICTN — файл частотного словаря, устанавливающий соответствие между словом, встречающимся в БД, его кодом и частотой (используется при текстовом поиске);

VOCAB — инверсный (инвертированный, обратный) для каждого слова БД список его содержащих документов (используется при текстовом поиске);

TEXTS — текстовый файл, содержащий собственно документы, используется при выдаче (просмотре) документов;

MASTR — прямой, последовательный файл, содержащий «собранные» с одну строку фиксированной длины форматные поля и список двухбайтовых кодов слов, находящихся в тексте данного документа и прошедших входную обработку (см. ниже). При необходимости в соответствующих местах находятся разделители сегментов и/или предложений. Файл используется при форматном поиске и при наличии в запросах конструкций SENT, SEGM, CTX.

Вопрос о происхождении содержания файлов рассматривается ниже, здесь же рассмотрим взаимодействие указанных файлов при проведении поиска (навигация в физической БД).



Рис. 2.12. Физическая структура БД STAIRS/DPS

Простой текстовый поиск:

000001>THERMAL AND REACTOR;

- обращение к файлу DICTN и определение кодов слов запроса: $C_T(THERMAL)$ и $C_R(REACTOR)$;
- обращение к файлу VOCAB и извлечение списков документов: $L_T(THERMAL)$ и $L_R(REACTOR)$;
- пересечение списков, образование результирующего списка: $L^C = L_T \cap L_R$;
- обращение к файлу TEXTS, извлечение документов списка L^C и предоставление пользователю.

Текстовый поиск с расширением:

000002>THERMAL AND REACT\$;

- обращение к файлу DICTN и определение кодов слов запроса: $C_T(THERMAL)$, $C_{R1}(REACTION)$, $C_{R2}(REACTIONS)$, $C_{R3}(REACTIVE)$, $C_{R4}(REACTOR)$, $C_{R5}(REACTORS)$;
- обращение к файлу VOCAB и извлечение списков документов: $L_T(THERMAL)$, $L_{R1}(REACTION)$, $L_{R2}(REACTIONS)$,

$L_{R3}(REACTIVE), L_{R4}(REACTOR), L_{R5}(REACTORS);$

- объединение списков, относящихся к $REACT\$:$

$$L_R = L_{R1} \cup L_{R2} \cup L_{R3} \cup L_{R4} \cup L_{R5};$$

- пересечение списков, образование результирующего списка

$$L^C = L_T \cap L_R;$$

- обращение к файлу TEXTS, извлечение документов списка L^C и предоставление пользователю.

Текстовый поиск с учетом контекста:

000001>THERMAL CTX REACTOR;

- обращение к файлу DICTN и определение кодов слов запроса:

$C_T(THERMAL)$ и $C_R(REACTOR);$

- обращение к файлу VOCAB и извлечение списков документов:

$L_T(THERMAL)$ и $L_R(REACTOR);$

- пересечение списков, образование списка $L_P^C = L_T \cap L_R;$

- обращение к файлу MASTR, извлечение документов, входящих в список $L_P^C;$

- сканирование закодированных текстов, выявление документов, в которых коды C_T и C_R находятся *рядом*. Составление результирующего списка $L^C;$

- обращение к файлу TEXTS, извлечение документов списка L^C и предоставление пользователю.

Аналогично строятся траектории навигации при сканировании форматных полей (в режиме `..SELECT`).

Обработка входного потока документов (загрузка БД)

Формирование содержания физической БД на основе продуктов служб-генераторов БД (или *загрузка* документов, рис. 2.13) является весьма важной операцией, за которую ответственность несет *администратор* базы данных (АБД). Основным средством администратора при проектировании логической структуры документальной БД является совокупность таблиц, описывающих внутренний формат документа и его связь с внешним (входным) форматом. Такие таблицы получили обобщенное название DBD (database definition — определение базы данных).

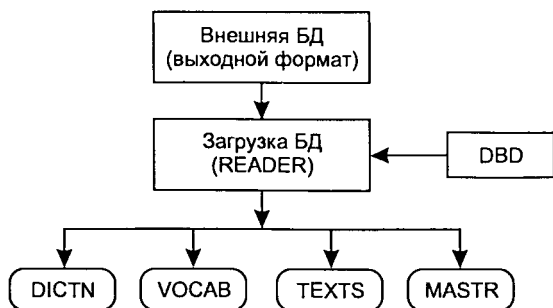


Рис. 2.13. Информационные потоки при загрузке БД

Входным форматом STAIRS является *карточный формат* (см. рис. 1.11, БД INIS), в котором поток документов представляет собой последовательность 80-байтовых строк, разделенных на три зоны — номер документа, метку поля (полей) документа, содержание.

Первая колонка содержит номер документа (последовательность строк с одинаковыми номерами образует документ), вторая колонка — метки полей (длинное поле может содержаться в нескольких строках с одинаковой меткой). Наоборот, несколько коротких (форматных) полей могут содержаться в одной строке.

Определение БД для случая, представленного на рис. 1.4, может иметь следующий вид и состоять из таблиц:

- описания форматных полей (ТОФП, табл. 2.3.);
- текстовых полей (ТОТП, табл. 2.4).

Описание форматных полей. Каждая строка таблицы содержит описание поля: имя с точки зрения пользователя, метка во входном потоке, начальная позиция в строке содержания, длина поля. Фактически это полная информация для извлечения подстроки поля из входного потока с целью помещения в определенные позиции записи файла MASTR.

Таблица 2.3. Описание форматных полей для БД INIS

Имя поля	Метка	Смещение	Длина	Пример значения (рис. 1.11)
RN:	000	23	6	000339
VID:	001	0	30	ОТЧЕТ
YEAR:	330	0	4	1984
ISSN:	350	0	20	ISSN 0056-2971
CLASS:	600	0	3	A12
CTRY:	700	0	2	US

Описание текстовых полей. Строки таблицы содержат информацию для анализа содержания документа и заполнения файлов DICTN и VOCAB. Поясним содержание таблицы:

- 1) имя поля — то же, что и для форматных полей;
- 2) метки поля — допускается указание нескольких меток, с целью объединения нескольких текстовых полей, например в данном примере (500) — первоначально приписанные дескрипторы (IAD, Initially Added Descriptors), в то время как (600) — автоматически приписанные дескрипторы, на основе тезауруса INIS (Computer Added Descriptors) может использоваться групповая нотация типа 50\$;
- 3) разделители (*терминаторы*) слов — символы, позволяющие выделить слова из входного потока — пробел и знаки препинания;
- 4) разделители (терминаторы) предложений — как правило, точка с пробелом; иногда учитываются разделители сегментов, если текстовое поле содержит внутренние сегменты (абзацы);
- 5) удаляемые символы (в данном случае — цифры);
- 6) транслитерация — используется при необходимости глобальной замены некоторых входных символов на другие;
- 7) максимальная и минимальная длина слова — используется для выделения из текста значащих слов. Очевидно, что короткие слова обычно представляют собой предлоги или артикли, а слишком длинные — ошибочные строки или «склеившиеся» слова. И те и другие не должны попадать в словарные файлы системы (DICTN, VOCAB);
- 8) ссылка на файл общих слов (Common Words, *отрицательный словарь*), или слов, которые хотя и имеют «нормальную» длину, но все же для данной предметной области не информативны или являются вспомогательными частями речи (*КОТОРЫЙ, ВСЕХ, ОДНАКО* и пр.). Включение подобных слов в файлы системы нецелесообразно;
- 9) ссылка на *положительный словарь* или список слов, которые могут не пройти через фильтр $L_{\min} - L_{\max}$, но являются информативными для данной предметной области (например, для юридических БД (см. Jurigus) в такой словарь могут войти: *СУД, ЛИЦ, ГОД, ЛЕТ, УК, ГК* и др., которые будучи короткими, все же существенны в очевидных контекстах!).

Таблица 2.4. Описание текстовых полей INIS

Имя поля	Метки	TERW	TERS	PURG	TRAN	$L_{\min} - L_{\max}$	CW	PW
1	2	3	4	5	6	7	8	9
ABSTR:	400	., /; : [] ()	._	1234 5678	НЕТ	3/35 ⁴	ФАЙЛ	ФАЙЛ
DESCR:	.500 600	;	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ	НЕТ

Опишем вкратце алгоритм загрузки некоторого документа в БД STAIRS.

Первоначальная фаза и загрузка форматных полей:

а) выделение очередного документа (с номером N_d) из входного потока и помещение его в буфер;

б) создание новой записи в файле TEXTS (N_d) и размещение в ней текста документа;

в) создание новой записи в файле MASTR (N_d), в соответствии с ТОФП извлечение из документа форматных полей и размещение их в фиксированной строке записи файла MASTR.

Загрузка текстовых полей:

г) извлечение очередного слова (W) из текущего текстового поля в соответствии с разделителями слов;

д) поиск W в положительном словаре. Если найдено — переход к (з);

е) проверка на $L_{\min} - L_{\max}$. Если W не входит в интервал — переход к (г);

ж) поиск в отрицательном словаре. Если W найдено — переход к (г);

з) начало загрузки слова в физические файлы БД. Поиск W в файле DICTN. Если найдено — переход к (л);

и) включение W в словарь (создание новой записи файла DICTN, назначение слову кода, обнуление частоты слова);

к) создание новой записи инверсного файла VOCAB, занесение кода W и номера текущего документа N_d ;

л) увеличение частоты слова на единицу, извлечение кода слова W , занесение кода в текущую запись MASTR;

м) поиск по коду в файле VOCAB, занесение в найденный список N_d ;

н) переход к (г).

АИПС локального и удаленного доступа Irbis

Система Irbis разработана в РГГУ (при участии таких организаций, как ВИНТИ, ВНТИЦЕНТР, ИНИОН, осуществлявших тестирование и опытную эксплуатацию промежуточных версий ИПС). Программный комплекс Irbis ориентирован на работу как в *полностью локальном* режиме (в том числе для поддержки баз данных, выпускаемых на CD ROM, — WinIrbis, рис. 2.14), так и в сетевых режимах, в том числе в локальной сети, в составе BBS и в составе Web-сервера для доступа по каналам Интернет (WebIrbis — см.

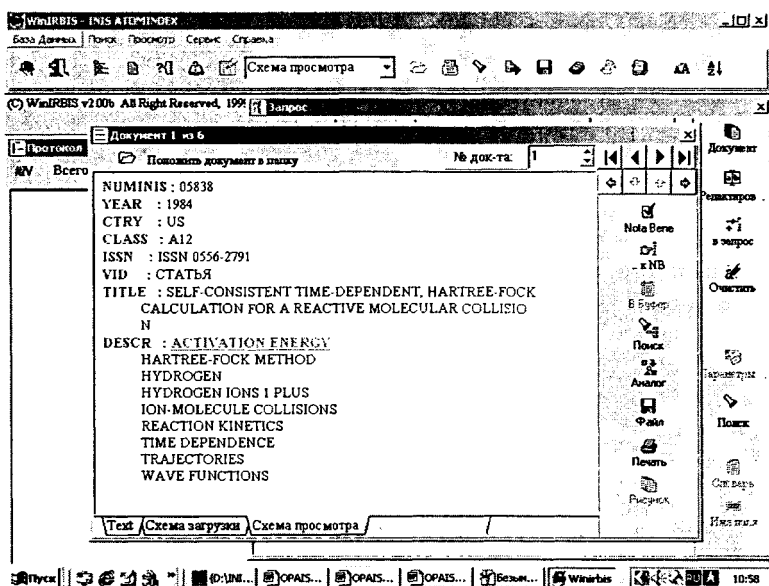


Рис. 2.14. Документ БД INIS в АИПС WinIrbis

рис. 2.16—2.19). АИС Irbis предназначена для многоцелевой обработки больших, в том числе полнотекстовых баз данных разнородных документов неограниченной длины с нерегулярной структурой. Система имеет развитые средства поиска, сортировки и вывода информации, обеспечивая гибкость и эффективность технологий информационного поиска.

Логическая структура данных IRBIS. База данных документальной ИПС IRBIS — это именованная совокупность массива документов и структурированных справочников, обеспечивающих эффективность поиска. Логическая структура БД представлена на рис. 2.15.

Документ базы данных есть структурированная форма представления информации, определяется своим уникальным (в массиве документов базы данных) идентификатором и составом полей.

Поле как часть документа представляет собой однозначно идентифицируемый в информационном массиве фрагмент, для которого определены тип, имя и характер обработки.

Слово как фрагмент поля, выделяемый по формальным (заданным в схеме представления документа) правилам, является единицей информации в операциях поиска.

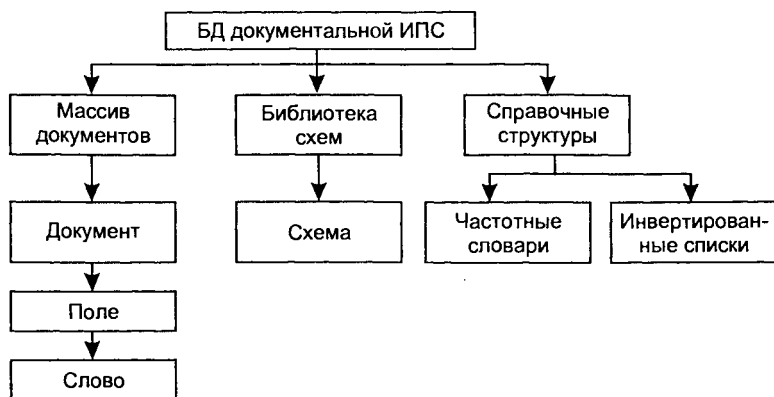


Рис. 2.15. Логическая структура БД документальной ИПС¹

Логическая связь именования, физического размещения и наполнения полей, образующих документ, а также стратегия поиска определяются *схемой* представления документа. Для одной базы данных может быть определено несколько разных схем, причем их количество не ограничено. Все схемы, используемые для работы с конкретной базой данных, в принципе равноправны. Каждая из них имеет свое имя, хранится в библиотеке схем и может быть одновременно использована для определения документов в нескольких базах данных.

Документ характеризуется в схеме перечислением описаний отдельных полей, для каждого из которых задается:

- идентификация материала в базе данных, обеспечивающая пользователю доступ средствами документального поиска;
- представление материала при вводе и выводе (формат и длина поля, размещение и оформление материала при отображении и т. д.);
- спецификация стратегии документального поиска (прямое сканирование данных или использование справочников). Для полей, специфицированных как ключевые, т. е. имеющих поисковые справочники, дополнительно определяются правила формирования дескрипторов (заданием списков символов-разделителей слов и списков стоп-слов).

¹ Голицына О. Л. Моделирование и разработка средств и технологий поиска документальной информации. 05.25.05 — Информационные системы и процессы, правовые аспекты информатики. Диссертация на соискание ученой степени канд. техн. наук / Научный руководитель — доктор техн. наук, проф. И. И. Попов. Москва, 2004 г.

Логически непрерывный массив документов в общем случае размещается в нескольких физических файлах, данные в которых связаны через указатель логического следования. Справочник базы данных размещается отдельно от массива документов и имеет специализированную структуру (частотный словарь, алфавитный индекс и инвертированные списки). Поисковые справочники являются производным по отношению к массиву документов.

Физическая структура базы данных Irbis. В ИПС Irbis используется следующая иерархия понятий:

База данных — некоторый объем файлового физического пространства для размещения данных, принадлежащих одной логической базе.

Файлы БД. Каждая база данных состоит не менее чем из двух типов файлов:

- файлов данных;
- файлов инвертированных структур.

Отдельный файл может принадлежать только одной базе данных.

Экстент. Пространство для хранения данных в БД выделяется блоками (*экстентами*) по 8 следующих друг за другом страниц размером 8 Кб. Экстент является единицей выделения пространства.

Страница. Файлы делятся на страницы размером по 8 Кб каждая. Логический номер страницы складывается из номера файла и номера страницы в файле (в простейшем случае логический номер равен номеру страницы в файле). В рамках БД файлы нумеруются, начиная с 1, и так же нумеруются страницы в рамках файла.

Словарные инвертированные структуры БД хранятся в отдельной области и включают три типа страниц:

- индексные страницы;
- страницы текстового представления словарных структур;
- страницы инвертированных списков.

Информационно-поисковый язык документальной ИПС Irbis. Информационно-поисковый запрос документальной ИПС представляет собой совокупность отдельных предложений запроса, в общем случае синтаксически и семантически не связанных между собой. Однако, само понятие «Запрос» предполагает объединенную общей тематикой последовательность поисковых действий, направленных на получение обобщенного результата, что позволяет разрешать ссылки на результаты отдельных предложений в рамках текущего запроса, объединять поисковые результаты, выделять общее множество релевантных документов и т. п.

Предложение запроса. Структурной единицей «Запроса» в рассматриваемом ИПЯ является *Предложение запроса*. Синтаксис *Предложения запроса* в нотациях Бэкуса-Наура следующий:

```
<Предложение запроса> ::= <Условие поиска> |
    <Предложение запроса><Логическая
операция><Предложение запроса>|
    (<Предложение запроса><Логическая
операция><Предложение запроса>
    <Логическая операция> ::= И | AND | _ | ИЛИ | OR | , | НЕ |
NOT | ^
```

Предложение запроса в общем случае состоит из произвольного числа *Условий поиска*, связанных логическими операциями И (AND, «пробел»), ИЛИ (OR, «,») и НЕ (NOT, «^»). Внутри предложения допускается использование скобок, задающих дополнительные приоритеты выполнения условий поиска.

Условие поиска. *Условие поиска* устанавливает критерии соответствия поисковых дескрипторов запроса некоторой области поиска, представляющей собой совокупность структурных единиц документа — полей.

```
<Условие поиска> ::= <Область поиска><Оператор
критерия><Выражение условия> | <Результат поиска>
```

Область поиска внутри документа задается именем отдельного поля или логическим выражением, объединяющим имена нескольких полей.

Выражение условия — набор терминов (поисковых дескрипторов), объединенных с помощью булевых или контекстных операторов в логическое выражение.

Оператор критерия задает условие включения или сравнения дескрипторов запроса и терминов, содержащихся в указанных полях документов.

В простейшем случае предложение запроса состоит из имени поля, оператора вхождения и одного дескриптора, например:

```
kw : РОССИЯ
```

Область поиска. Область поиска задается именами структурных единиц документа — полей.

```
<Область поиска> ::= <Имя поля>|
    (<Область поиска> <Логическая операция> <Область поиска>)
```

Из нотации видно, что допускается использование логических операций при формировании области поиска. Например:

```
(AB OR TI) : (РОССИЯ NOT СССР)
```

означает, что в результат поиска включаются все документы, в которых хотя бы в одном из заданных полей (или в обоих) встречается дескриптор РОССИЯ, но не встречается дескриптор СССР.

Отличительная особенность ИПЯ Irbis — возможность использования логического выражения как в правой, так и в левой части условия поиска.

Оператор критерия. Для связи области поиска с терминами запроса используются следующие операторы критерия (вхождения, сравнения):

<Оператор критерия> ::= : | = | EQ | <> | NE | > | GT | >= | GE | < | LT | <= | LE

«>» (условие вхождения) — позволяет найти документы, которые содержат в указанной области поиска результат вычисления выражения условия;

«=» (условие «равно», или EQ) — позволяет найти документы, для которых указанная область поиска равна результату вычисления выражения условия;

«<>» (условие «не равно», или NE) — позволяет найти документы, которые не содержат в указанной области поиска результат вычисления выражения условия;

«>» (условие «больше», или GT) — позволяет найти документы, которые содержат в указанной области поиска значения большие, чем результат вычисления выражения условия;

«>=» (условие «больше или равно», или GE) — позволяет найти документы, которые содержат в указанной области поиска значения, большие или равные результату вычисления выражения условия;

«<» (условие «меньше», или LT) — позволяет найти документы, которые содержат в указанной области поиска значения, меньшие, чем результат вычисления выражения условия;

«<=» (условие «меньше или равно», или LE) — позволяет найти документы, которые содержат в указанной области поиска значения, меньшие или равные результату вычисления выражения условия.

Выражение условия. Синтаксис выражения условия в ИПЯ следующий:

<Выражение условия> ::= <Дескриптор> |
 <Выражение условия> <Операция> <Выражение условия> |
 (<Выражение условия> <Операция> <Выражение условия>)

<Операция> ::= <Логическая операция> <Контекстная операция>

<Контекстная операция> ::=

STX|CTX[N]||+|NEAR|NEAR[N]|SENT|CON[N]

При использовании в запросе нескольких дескрипторов они должны быть связаны контекстными или логическими операторами и помещены в круглые скобки.

Для формулировки запроса в ИПС Irbis используются контекстные операторы STX[N], NEAR[N], SENT. Параметр N может принимать значения от 0 до 255 (по умолчанию N равно 0). Отсутствие параметра означает следование терминов в поле непосредственно друг за другом (идентично значению 0).

Оператор STX позволяет найти документы, в заданной области поиска которых в одном предложении присутствуют поисковые дескрипторы, расположенные в указанном порядке на расстоянии не более N слов друг от друга. Выражение условия имеет вид:

<дескриптор1> STX[N] <дескриптор2>

Оператор NEAR позволяет найти документы, в заданной области поиска которых в одном предложении присутствуют поисковые дескрипторы на расстоянии N слов друг от друга (в произвольном порядке). Выражение условия имеет вид:

<дескриптор1> NEAR[N] <дескриптор2>

Оператор SENT позволяет найти документы, в заданной области поиска которых поисковые дескрипторы находятся в одном предложении. Выражение условия имеет вид:

<дескриптор1> SENT <дескриптор2>

Контекстный оператор пересечения полей CON[N] позволяет использовать в выражении условия имена полей (выступающие в данном случае в роли дескрипторов), содержимое которых сравнивается на предмет отыскания общих терминов.

Оператор пересечения полей служит для отбора документов, в заданных полях которых имеется не менее N одинаковых поисковых терминов. Выражение условия запроса имеет вид:

<имя поля1> CON[N] <имя поля2>

Синтаксис и семантика использования дескрипторов. Поисковые дескрипторы могут быть заданы одним из следующих способов:

- выбор из частотного словаря;
- ввод с клавиатуры;
- отметка ключевых слов в тексте документа;
- выбор терминов из рубрикаторов, словарных или тезаурусных структур.

При задании поисковых дескрипторов допускается использование операторов (символов) маскирования, алгоритма нормализации и ссылок на ранее полученные результаты поиска.

Маскирование. ИПЯ предполагает применение двух методов:

- маскирование (или замена) произвольного числа рядом стоящих символов дескриптора (символы «*» или «\$»);
- маскирование одного (непустого) символа дескриптора (символ «%»).

Символы маскирования могут использоваться вместо любого символа дескриптора, и их количество не ограничено.

Параметризованные символы маскирования произвольного количества символов (например, «*(N)») означают, что в дескрипторе на месте символа маскирования может стоять произвольная последовательность длиной не более чем N символов (где N изменяется от 0 до 255).

Нормализация. Для расширения возможностей дескрипторного языка на этапе сопоставления ПОД и ПОЗ может быть использован аппарат нормализации дескрипторов.

Используются следующие правила нормализации дескриптора ПОЗ:

1. Три первые буквы дескриптора остаются без изменения.
2. Все следующие гласные буквы заменяются символом маскирования произвольного числа рядом стоящих букв.
3. Конечные буквы *в*, *г*, *м*, *х* в дескрипторе заменяются символом маскирования произвольного числа рядом стоящих букв.
4. В конце дескриптора проставляется символ маскирования произвольного числа рядом стоящих букв (если после всех преобразований конечный символ дескриптора не является символом маскирования).

Лингвистическое обоснование такой замены заключается в том, что смысловозначительная роль согласных во много раз больше, чем гласных. Начальная часть слова включается в новый дескриптор без изменения, поскольку информативность первых трех букв в слове велика. Согласные *в*, *г*, *м*, *х* могут попадать в дескриптор из окончаний существительных и прилагательных, поэтому их исключение из дескриптора и замена символом маскирования ведет к отсечению окончаний.

Нормализованный таким образом дескриптор ПОЗа позволяет обеспечить более полный дескрипторный поиск с использованием только лишь частотного словаря БД.

Рассмотрим, например, запрос, который на естественном языке представляет собой предложение: «Частотный анализ терминов словаря». Такой запрос в системе (с применением правил нормализации) автоматически преобразуется в следующий ПОЗ:

част\$тн\$ AND анал\$з\$ AND терм\$н\$ AND слов\$р\$

Нормализованный таким образом ПОЗ обеспечивает поиск по логическому выражению с разрешением символов маскирования:

част\$тн\$ = частотность, частотности, частотный, частотные, частотных, частотного, частотной;

анал\$з\$ = анализ, анализа, анализе, анализу, анализируется, анализируются;

терм\$н\$ = термин, термина, термину, термином, термины, терминов, терминах, терминология, терминологии, терминологию, терминологические, терминологическим, терминологических, терминологической, терминологический, терминосистем, терминологичности;

слов\$р\$ = словарь, словаря, словаре, словарем, словарей, словарные, словарными, словарных, словарного, словоформа, словоформе, словоформы, словоформ, словарные, словарно-грамматический, словоупотреблений.

Использование ранее полученных результатов поиска. В качестве операнда условия поиска в предложении запроса может использоваться ранее полученный *результат поиска*:

<Результат поиска> ::= # <Идентификатор результата поиска>

Для включения в предложение поискового запроса результатов ранее проведенного поиска используются ссылки на номер предложения в текущем запросе.

Например, запрос может иметь вид:

#2 and ((KW or AB) : Россия),

где #2 — ссылка на результат второго предложения запроса.

Символ «#» является индикатором ссылки. За ним указывается номер одного из предыдущих предложений текущего запроса или имя сохраненного запроса, результат поиска по последнему предложению которого используется для уточнения.

Средства формирования запросов. Поисковые механизмы построены на основе ИПЯ, однако технология и средства формирования пользователем запроса не требуют от него обязательного знания и навыков построения выражений алгебраического вида.

Поисковые интерфейсные средства системы Igbis условно можно разделить на два класса.

Первый класс (сценарии типа «укажи и выбери») — это конструкторы запросов, которые позволяют, используя термины поисковых словарей или других поисковых структур (тезаурусов, рубрикаторов, словников), в режиме диалога построить выражение той или иной сложности.

Второй класс — это средства, реализующие простейший сценарий типа «укажи и получи». В этом случае пользователь выделяет в отображаемом объекте (документе или множестве документов) значимые с его точки зрения элементы (термины в документе или словаре; документы в выборке или протоколе) и, используя механизмы поиска по сходству (поиск аналогов, эвристический поиск, поиск с использованием обратной связи), получает выдачу, минуя этап составления поискового выражения.

В основу формирования поискового запроса по технологии «укажи и выбери» в системе положено три различных подхода, ориентированных на разные степени подготовленности пользователя.

Конструктор запроса «по образцу» реализует традиционный для библиографического поиска форматно-ориентированный интерфейс, имеет жестко фиксированную модель поискового условия, предполагающую обязательное выполнение частных условий, связанных с полями, выбираемыми из предопределенного списка. По умолчанию предполагается, что отдельное условие — это список терминов (синонимов), обычно выбираемых из словаря и обозначающих одно и то же понятие.

Конструктор формирования запроса «по шагам» характеризуется большей гибкостью. Здесь поисковые термины также выбираются из словаря, но могут связываться любыми отношениями. Причем, построенные таким образом лексические выражения, относимые к отдельным полям, в свою очередь могут связываться операторами, выбираемыми из списка. Такой конструктор позволяет формировать достаточно сложные предложения запроса последовательным наращиванием либо выражения условия (путем добавления очередного термина), либо всего предложения (путем добавления нового условия поиска). Необходимо отметить, что для сложных предложений запроса необходима достаточно хорошая предварительная структуризация.

Конструктор формирования логического выражения запроса путем непосредственного набора выражения запроса с возможностью обращения в произвольном порядке к словарям, спискам имен полей и т. д.

На рис. 2.16—2.19 приведен пример последовательности экранов при простейшей формулировке запроса — один термин из об-

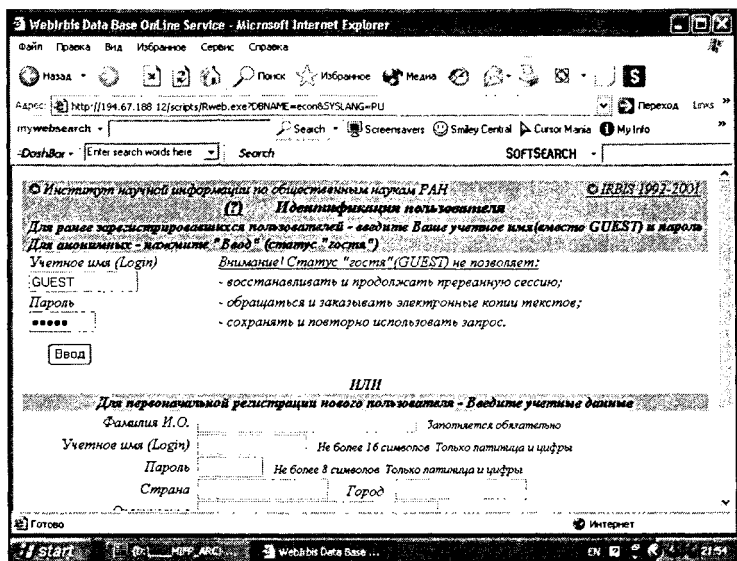


Рис. 2.16. Исходная страница при доступе к БД по экономике и демографии ИНИОН посредством ППП WebIrbis

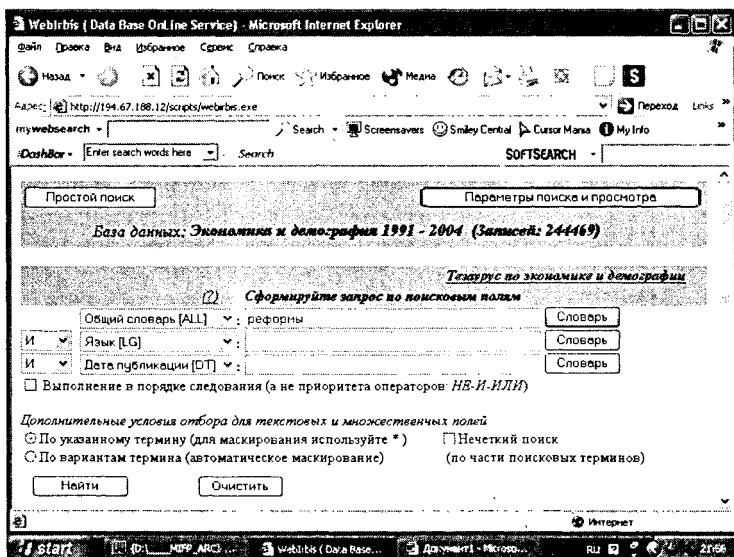


Рис. 2.17. Запрос на поиск в частотном словаре БД Irbis. Ключевое слово — РЕФОРМЫ

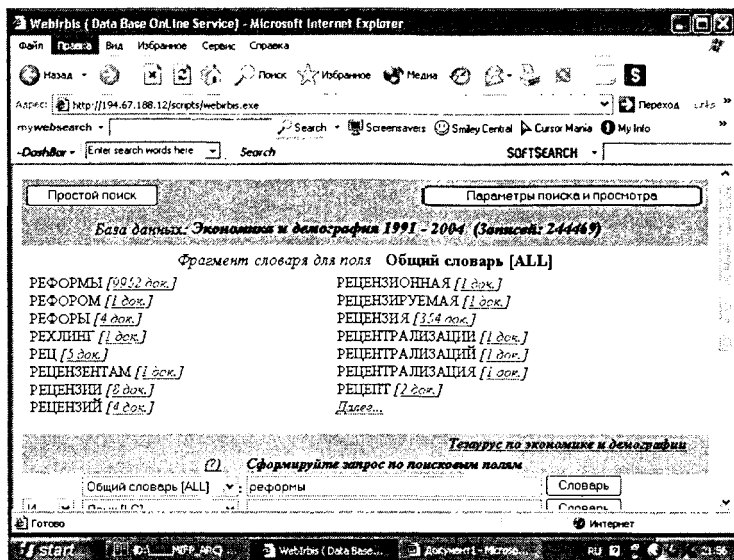


Рис. 2.18. Просмотр фрагмента частотного словаря БД начиная со слова РЕФОРМЫ

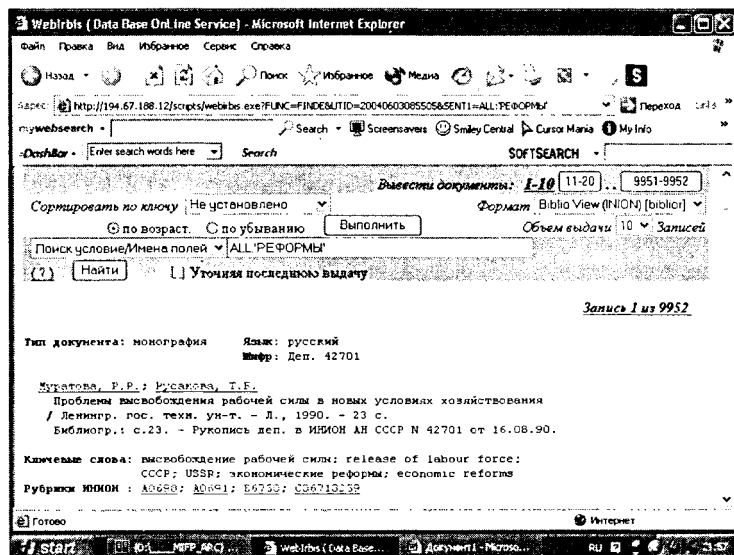


Рис. 2.19. Экран просмотра результатов поиска в WebIrbis (библиографическая запись, содержащая словосочетание ЭКОНОМИЧЕСКИЕ РЕФОРМЫ)

шего (ALL) частотного словаря. Поиск производится в базе данных «Экономика и демография» ИНИОН РАН, Web-адрес — www.inion.ru

Документальный информационный поиск в сети Интернет

Информационно-поисковые системы Интернет могут быть разделены по функционально-структурному принципу на следующие классы (см. [23, 32]):

- полностью распределенные системы, где реализуются принципы распределенных вычислений и хранения данных;
- частично распределенные — распределенные данные и локализованная обработка;
- локальные системы — локализованные данные и их обработка (см. выше, WebIrbis).

Поисковые задачи в сети Интернет имеют следующие особенности [19]:

1. Огромный объем доступной информации. За последние годы было предпринято множество попыток оценить размер ресурса Интернет, и, хотя оценки не полностью совпадают, все они единогласны в том, что в Интернет содержится более миллиарда страниц и их число увеличивается экспоненциально. Если учесть, что размер среднестатистической страницы составляет 5—10 Кб, объем оценивается в десятках терабайт (10×2^{40} байт).

2. Высокий процент временной информации. Информация в Интернет очень динамична, информационные ресурсы непрерывно появляются, пропадают, перемещаются, обновляются. В среднем, ежемесячно изменяется около 40 % информации, среднее время жизни половины страниц в Интернет не превышает 10 дней.

3. Неконтролируемое качество информации. Отсутствие редакторского контроля над публикуемой информацией в Интернет обуславливает проблему ее качества — информация может быть некорректной (например, устаревшей), ложной, плохо сформулированной, содержать ошибки (опечатки, грамматические ошибки, ошибки оцифровки и т. п.). Так, по некоторым оценкам, одна опечатка встречается в среднем в каждых двухстах часто употребляемых словах или в трех иностранных фамилиях.

4. Разнородность информации. Кроме различных форматов представления информации, используется также множество различных языков и алфавитов. Около 30 % информации в Интернет составляют точные или приблизительные копии других документов.

В Интернет изменяется понятие *типичного пользователя*. Имеются следующие различия:

- *плохо сформулированные запросы*. Никто специально не обучает пользователей АИПС в Интернет формулировать запросы, и, как следствие, немногие из них используют расширенные возможности поиска, такие, как логические выражения. Более того, типичные запросы очень коротки — более 60 % поисковых запросов в Интернет состоит из 1—2 слов, в то время как в классических АИПС — из 7—9 слов;
- *разнородный контингент*. Разнообразие в знаниях, потребностях и ожиданиях пользователей очень велико. Большинство же работающих с классическими ИПС обычно имеет много общих черт;
- *поведение пользователей*. Обычно они не готовы долго ожидать результата или даже искать его в предоставленной системой выборке (58 % не идет дальше первого экрана, а 67 % не предпринимает попыток модифицировать свой первоначальный запрос).

Структура процессов в поисковых системах WWW. Ключевым отличием данных систем от классических АИПС является наличие *сетевых роботов — программных модулей-агентов*, занимающихся сбором информации о доступных WWW-ресурсах. Они осуществляют *сканирование* и, начиная с некоторого множества ссылок (URL) на WWW-страницы, рекурсивно обходят ресурсы Интернет, извлекая ссылки из получаемых документов до тех пор, пока не будет выполнено некоторое условие остановки.

Собранная информация помещается в хранилище, содержимое которого определяет набор документов, по которым идет поиск.

Как и в классических ИПС, для достижения приемлемой эффективности поиск производится не напрямую по документам в хранилище, а по индексным структурам, за создание которых отвечает *модуль индексирования*.

Получение и выполнение запросов пользователей — это задача модуля *поисковой машины*.

При исследовании задачи сканирования возникает ряд вопросов.

1. Выбор WWW-ссылок для индексирования. Обычно робот имеет информацию о существовании множества еще не отсканированных ресурсов и может выбирать, какой из них «посетить» следующим. Этот выбор осуществляется согласно используемой роботом стратегии сканирования, которая напрямую определяет множество страниц, которое будет обработано роботом и, как следствие, какие страницы будут известны поисковой системе.

2. Частота обновления индекса. В связи с высокой динамикой изменений в WWW собранная информация о многих посещенных страницах довольно быстро перестает отражать их реальное текущее содержание.

3. Минимизация нагрузки на WWW-серверы. Робот не должен перегружать сервер своими запросами или сканировать информацию против воли владельца сервера. Для выражения этих условий может использоваться файл `robots.txt`, располагающийся в корневом каталоге сервера.

4. Организация параллельного сканирования. Одним из способов повышения производительности является параллельный запуск нескольких роботов на разных машинах, но при этом необходим хорошо масштабируемый механизм координации их действий (например, чтобы они одновременно не сканировали один и тот же адрес).

Индексы поисковых систем. Важнейшим отличием поисковых систем для поиска в Интернет от классических систем информационного поиска является необходимость обслуживания всех запросов без реального доступа к ресурсам на момент их выполнения, иначе необходимо либо хранить свежую локальную копию всех ресурсов (что слишком накладно), либо посещать их во время выполнения запроса (что слишком медленно).

Поэтому в системах поиска в Интернет все запросы обслуживаются на основе индекса, содержащего описание известных данной поисковой системе ресурсов.

Хранилище. Хранилище содержит большое количество объектов данных (страниц WWW) и в этом смысле очень похоже на СУБД или файловую систему. Однако многие возможности последних в данном случае не нужны (например, поддержка транзакций или иерархия директорий), зато очень важны такие, как масштабируемость и реализация двух режимов доступа:

- произвольного — для того, чтобы быстро найти конкретную страницу по ее идентификатору (например, для создания копии страницы);
- потокового — для того, чтобы извлечь значительную часть всей коллекции (например, для индексирования или анализа).

Кроме того, важными являются эффективная поддержка обновлений, сборка «мусора» (устаревших страниц).

Модуль индексирования. Задачей этого модуля является построение необходимых индексов. Кроме текстовых; часто используются индексы, описывающие структуру графа WWW, а также вспомогательные индексы (например, индекс для доступа к страницам по их длине или по количеству используемых графических изображений).

Специфика WWW определяет свои особенности построения текстового индекса для ИПС. В дополнение к традиционным целям — минимизации времени доступа и размера индекса, также важно минимизировать время его создания и обеспечить возможность эффективного обновления.

Поисковая машина. Расширенные возможности поиска имеют малый спрос у пользователей ИПС для WWW. Как следствие, за исключением предикатов, позволяющих наложить условия на входящие/исходящие ссылки, в языках запросов, применяемых в системах в WWW, нет существенных нововведений по сравнению с языками запросов классических ИПС.

Однако простота запросов влечет их низкую селективность, и поэтому очень важной задачей является упорядочивание результатов, так чтобы первыми оказались те результаты, которые, вероятнее всего, интересны пользователю. Классические подходы к ранжированию опираются на меру схожести текстов запроса и документа, но «расплывчатые запросы» пользователей и огромное количество документов значительно понижают эффективность таких подходов в контексте WWW.

Некоторые поисковые возможности и характеристики систем Yandex и Rambler.

Стандартный поиск Yandex. Рассмотрим общий вид стандартной поисковой формы Yandex (рис. 2.20).

1. *Основная поисковая форма.* Главный ее элемент — строка запроса. При желании можно искать только в результатах предыдущего поиска («искать в найденном»). Можно также задать запрос по любому из источников информации («Рунет», «Новости», «Товары», «Энциклопедии», «Картинки»).

2. *Статистика поиска* — число найденных документов и частота заданных в запросе слов. По ссылке «страниц» можно перейти на несгруппированную по серверам выдачу.

3. *Рубрики каталога.* По нажатию на ссылку происходит переход в соответствующий раздел каталога (выводится только при точном соответствии запросу).

4. *Результаты поиска по WWW.* Для каждого документа выдается следующая информация: заголовок — ссылка на ресурс, [Показать найденные слова] — переход на активный документ, в котором контекст запроса выделен стрелочками, поиск похожего документа и, если найденный сайт описан в каталоге, переход в рубрику каталога.

5. *Переход на следующие страницы результата.* Сортировка по убыванию дат.

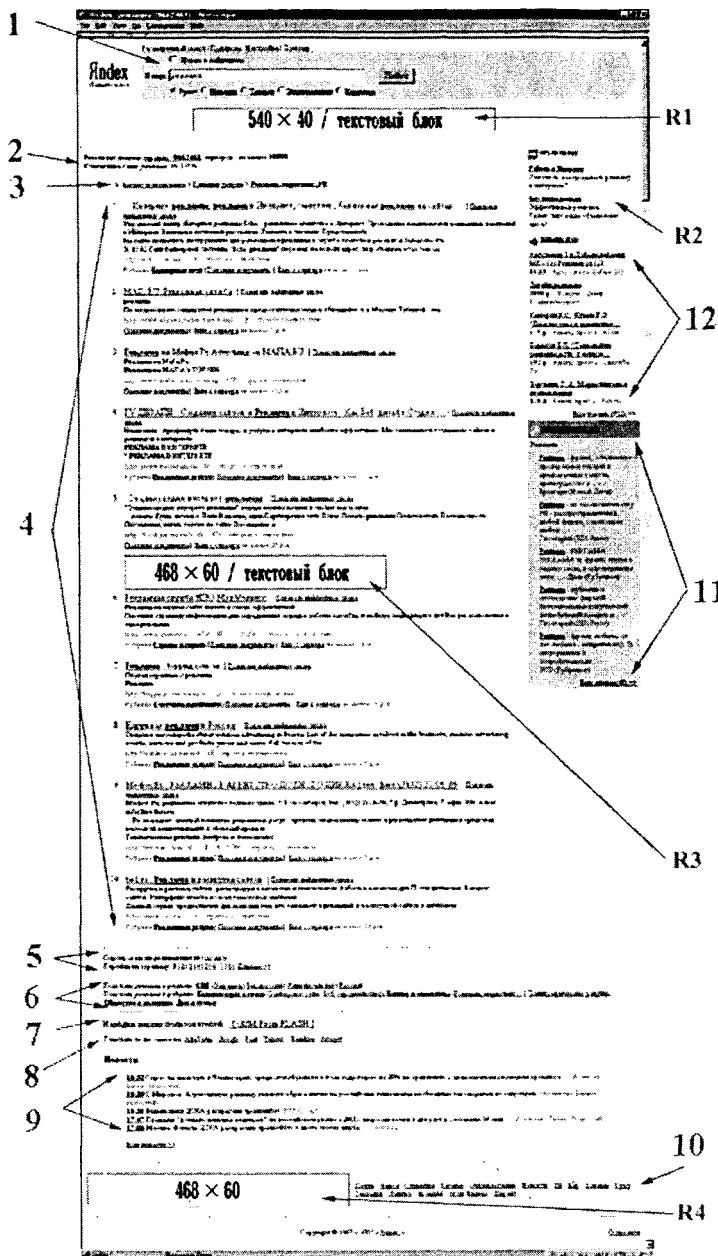


Рис. 2.20. Формат обычного поиска в Yandex

6. *Мастер запросов.* Здесь даются советы и предложения по уточнению запроса. Можно сузить поиск по рубрике каталога или по региону.

7. *Популярные находки пользователей* — ссылки на документы, выбранные пользователями по этому запросу (выводится только при точном соответствии).

8. *Возможность поиска* в других русскоязычных поисковых машинах.

9. *Результат поиска в новостных лентах* информационных агентств (выводится только при точном соответствии запросу). При высоком соответствии и актуальности эта секция может оказаться над результатами поиска по WWW-серверам.

10. *Ссылки на прочие службы* Yandex'a.

11. *Результат поиска* в энциклопедических статьях (выводится только при точном соответствии запросу).

12. *Результат поиска* в базе товарных предложений магазинов (выводится только при точном соответствии запросу).

Реклама на странице:

R1. *Верхний баннер* (текстовый блок);

R2. *Yandex.Директ* — реклама без посредников;

R3, R4. *Текстовая строка* (баннер);

R5. *Нижний баннер.*

Расширенный поиск Yandex. Форма расширенного поиска приводится на рис. 2.21.

Искать

(используйте минус перед словом для его исключения)

Искомые слова

расположены относительно друг друга

расположение на странице

употреблены в тексте

подряд

в одном предложении

не очень далеко

на одной странице

где угодно

в заголовке

в тексте ссылки

на сайт

в любой форме

точно так, как в запросе

Страницы

Найденные страницы должны обладать следующими свойствами.

язык

русский

английский

французский

немецкий

дата

любая

последние 2 недели

последний месяц

7 4 2004

формат

HTML

PDF

DOC

RTF

содержит ссылку

В результатах поиска показывать ссылок на странице

находится на сайте

похожи на страницу

Итого: искать на все слова расположены не очень далеко, в любой части страницы.

Искать

Рис. 2.21. Расширенный поиск в системе Yandex

Словарный фильтр.

Здесь пользователь может указать, какие слова обязательно должны встретиться в документе, каких быть не должно, а какие желательны (т. е. могут и отсутствовать). Поле «все формы» или «точная форма» указывает системе Yandex, надо ли учитывать при запросе все словоформы. «Точная форма» обычно требуется только для поиска цитат.

Зоной поиска слова (табл. 2.5) может быть как текст документа (слова находятся в одном предложении или во всем документе), так и его заголовок, аннотация (тэг HTML *description* — см., например, [24]), ссылка (подпись URL) или адрес (сам URL). Вариант «во фразе» означает необходимость искать слова в том порядке, в котором они введены; может быть задано несколько слов через запятую.

Таблица 2.5. Синтаксис языка запросов (строгий поиск)

Синтаксис	Значение оператора	Пример запроса
Пробел или &	Логическое И (в пределах предложения)	Лечебная физкультура
&&	Логическое И (в пределах документа)	Рецепты && (плавленный сыр)
	Логическое ИЛИ	Фото фотография снимок фотоизображение
+	Обязательное наличие слова в найденном документе (работает также в применении к стоп-словам)	+быть или +не быть
()	Группирование слов	(Технология изготовление) (сыра творога)
-	Бинарный оператор И–НЕ (в пределах предложения)	Банки - закон
-- или -	Бинарный оператор И–НЕ (в пределах документа)	Путеводитель по Парижу -- (агентство тур)
/ (п м)	Расстояние в словах (-назад, +вперед)	Поставщики /2 кофе музыкальное /(-2 4) образование вакансии - /+1 студентов
" "	Поиск фразы	"Красная шапочка" (эквивалентно красная /+1 шапочка)
&&/ (п м)	Расстояние в предложениях (-назад, +вперед)	Банк && /1 налоги

Дата.

Ограничение выдачи документов по дате. Документы с неизвестной датой в этот список не включаются.

Сайт/вершина.

Запрос идет только по страницам указанного сайта или поддиректории (вершины) сайта. Поиск будет осуществлен среди всех поддиректорий. Здесь же (в соседнем поле) можно исключить из поиска страницы определенного сайта, а также внести несколько адресов, перечислив их через пробел.

Изображение.

Поиск документов, содержащих картинку с определенным названием или подписью. Если файл картинки называется, например, `applegreen.jpg`, то найти такие файлы можно запросом `apple` (или `apple*. *`). Для поиска в подписи к изображению (тэг `alt`) следует вписать запрос в соседнее поле.

Специальные объекты.

Поиск страниц, содержащих файлы объектов (скрипт, объект, апплет, `java`). В поле указывается имя объекта (табл. 2.6).

Язык.

Yandex определяет язык документа, который может быть указан при поиске (кириллица или другой). В базе Yandex находятся только документы русскоязычного Интернета (по умолчанию в поисковую машину вносятся серверы в доменах `su`, `ru`, `am`, `az`, `by`, `ge`, `kg`, `kz`, `md`, `tj`, `ua`, `uz`), а также зарубежные сайты, представляющие интерес для русскоязычного поиска.

Формат выдачи.

«Краткая выдача» показывает только список заголовков документов. «Только URL» — адреса найденных страниц.

Таблица 2.6. Поиск в элементах страниц

Синтаксис	Что означает оператор	Пример запроса
<code>\$title</code> (выражение)	Поиск в заголовке	<code>\$title (CompTek)</code>
<code>\$anchor</code> (выражение)	Поиск в тексте ссылок	<code>\$anchor (CompTek Dialogic)</code>
<code>#keywords=(выражение)</code>	Поиск в ключевых словах	<code>#keywords=(поисковая система)</code>
<code>#abstract=(выражение)</code>	Поиск в описании	<code>#abstract=(искала поиск)</code>
<code>#image="значение"</code>	Поиск файла изображения	<code>#image="tort*"</code>
<code>#hint=(выражение)</code>	Поиск в подписях к изображениям	<code>#hint=(lenin ленин)</code>
<code>#url="значение"</code>	Поиск на заданном сайте (странице)	<code>#url="www.comptek.ru*"</code>
<code>#link="значение"</code>	Поиск ссылок на заданный URL	<code>#link="www.yandex.ru*"</code>
<code>#mime="значение"</code>	Поиск в документах данного (pdf или rtf) типа	<code>#mime="pdf"</code>

Возможности расширенного поиска в системе Rambler (рис. 2.22).

Форма расширенного поиска дает возможность:

- задавать дополнительные параметры поиска;
- редактировать параметры поиска и поля, заданные по умолчанию;
- выбирать наиболее удобную форму показа результатов поиска.

В форме расширенного поиска указываются следующие параметры.

Поиск по тексту.

- всего документа — поиск осуществляется по всему документу, включая его название и заголовки; включено по умолчанию;

Расширенный поиск в Интернете

Что искать:

Поиск по тексту:

всего документа названия (<title>) гиперссылок

Искать слова запроса:

все ("и") хотя бы одно ("или") точную фразу

Расстояние между словами запроса:

не ограничивать ограничивать

Исключить документы, содержащие следующие слова:

Язык документа:

своей русский английский украинский

Дата документа (в формате "24/04/2001"):

начиная с по

Искать документы только на следующих сайтах:

Примеры: *science.rambler.ru, www.lenta.ru, www.hosting.ua/~name*
(после "*" - только каталоги, начинающиеся с "*" - " ")

Вывод результатов

Сортировать:	Выдавать:	Форма вывода:
<input checked="" type="radio"/> сайты по релевантности	<input checked="" type="radio"/> по 15	<input checked="" type="radio"/> стандартная
<input type="radio"/> страницы по релевантности	<input type="radio"/> по 30	<input type="radio"/> краткая
<input type="radio"/> страницы по дате (сначала новые)	<input type="radio"/> по 50	<input type="radio"/> детальная
<input type="radio"/> страницы по дате (сначала старые)		
Связанные запросы:		
<input type="radio"/> показывать отдельной колонкой		
<input checked="" type="radio"/> показывать внизу страницы		

Рис 2.22. Форма расширенного поиска системы Rambler

- названия — учитываются только названия документов (тэг HTML <title>);
- заголовков — учитываются только заголовки документов (тэги <h1>, <h2>, <h3>, <h4>).

Искать слова запроса.

- все («и») — документ выдается только в том случае, если в нем присутствуют все слова запроса (включено по умолчанию);
- хотя бы одно («или») — документ выдается, если в нем встретилось хотя бы одно слово из запроса;
- точную фразу — документ выдается, если в нем встретились все слова запроса, причем в том же порядке и формах, как и в запросе; выбор этой опции равнозначен заключению поискового запроса в двойные кавычки.

Расстояние между словами запроса.

- ограничивать — расстояние между словами из запроса в тексте документа не должно быть слишком большим; включено по умолчанию, поскольку повышает точность поиска;
- не ограничивать — расстояние между словами не играет роли; будут найдены все документы, содержащие слова запроса, вне зависимости от того, на каком расстоянии друг от друга они находятся.

Исключить документы, содержащие следующие слова.

Из списка найденного исключаются те документы, в которых есть слова, перечисленные в этом поле.

Язык документа.

- любой — независимо от языка (включено по умолчанию);
- русский — поиск только по «русскоязычным» (кириллица) документам;
- английский — поиск только по «англоязычным» документам (латиница).

Дата создания документа.

Позволяет отбирать только те документы, дата создания которых укладывается в заданный диапазон. В частности, можно ограничить выдачу только «новыми» (начиная с указанной даты) или «старыми» документами (до указанной даты). Все даты задаются в формате день/месяц/год, например 29/02/2000. По умолчанию находятся любые документы вне зависимости от даты. Внимание: если сервер не возвращает даты документа, то в качестве таковой про-

ставляется дата индексирования (день, когда документ был считан «роботом» Ramblera).

Искать документы только на следующих сайтах.

Позволяет отбирать только те документы, которые найдены на указанных сайтах. Под сайтом понимается либо уникальное DNS-имя (домен), либо DNS-имя с каталогом первого уровня, начинающимся с тильды. Например [24]:

```
top100.rambler.ru, www.lenta.ru, www.hosting.ua/~name —  
но не www.rambler.ru/domains/.
```

Можно указать несколько сайтов через запятые. По умолчанию в поиске участвуют документы со всех проиндексированных сайтов.

- Вывод результатов поиска.

Сортировать

- сайты по релевантности — найденные документы группируются по сайтам, так что одна позиция в списке результатов поиска может соответствовать нескольким документам; порядок выдачи сайтов определяется их релевантностью (степенью соответствия запросу документов с сайта); включено по умолчанию;
- страницы по релевантности — документы не группируются по сайтам, т. е. все документы с одного сайта выдаются по отдельности; порядок выдачи определяется релевантностью каждого отдельного документа;
- страницы по дате (сначала новые) — документы не группируются по сайтам; порядок выдачи — от более новых документов к более старым;
- страницы по дате (сначала старые) — то же, что и в предыдущем случае, но сначала выводятся самые старые из найденных документов.

Выдавать

- по 15 — на страницах результатов поиска выводится по 15 найденных документов (включено по умолчанию);
- по 30 — количество позиций на страницах результатов поиска увеличивается до 30;
- по 50 — количество позиций на страницах результатов поиска увеличивается до 50.

Форма вывода

- стандартная — включено по умолчанию;
- краткая — в результатах поиска показываются только заголовки найденных документов;

- детальная — выводится максимум информации о найденных документах: заголовок, аннотация, идентификатор документа, даты модификации и индексирования, размер, кодировка, адрес и т. п.

Связанные запросы.

- показывать — в левой части экрана выводится список запросов, «связанных» с данным, т. е. часто задаваемых теми пользователями, которые вводили данный запрос;
- не показывать — включено по умолчанию; колонка со списком «связанных» запросов не выводится (начало списка показывается внизу страницы под заголовком «У нас также ищут»).

В связи с постоянным ростом количества документов в сети, система должна быть *масштабируемой*. В *Ramblere* масштабируемость достигается за счет параллельного исполнения задачи произвольным количеством машин.

Сбором информации занимается *робот*, который обходит страницы с заданными URL [24], загружает их в базу данных, а затем архивирует и перекладывает в хранилище суточными порциями. Роботы размещаются на нескольких машинах, и каждый из них выполняет свое задание. Так, робот одной машины может загружать новые страницы, которые еще не были известны поисковой системе, а другой — страницы, которые ранее уже были загружены (не менее месяца, но и не более года назад). При необходимости обработка распределяется другим способом, например, разбив список URL на 10 частей и распределив их по 10 машинам. Параллельная работа программ позволяет увеличивать нагрузку — при возрастании числа страниц, которые нужно обойти роботу, достаточно разместить задачу среди большего числа машин.

Собранная в хранилище информация (в сжатом виде) разбивается на фрагменты по 50 Мбайт, распределяемые между 70 машинами, осуществляющими индексирование. Как только индекатор на одной из машин заканчивает обработку порции страниц, он обращается за следующей. В результате на первом этапе формируется ряд небольших индексных баз, каждая из которых содержит информацию о некоторой части Интернета.

После того как все части информации обработаны, осуществляется объединение результатов. Благодаря тому, что частичные индексные базы и основная БД, к которой обращается поисковая машина, имеют одинаковый формат, процедура слияния является быстрой операцией, не требующей никаких дополнительных моди-

фикаций частичных индексов. Основная база участвует в анализе как одна из частей нового индекса.

Кроме того, единый формат позволяет проводить тестирование частичных баз еще до объединения их с основной, и обнаруживать ошибки на более раннем этапе.

Индексная база поисковой системы Rambler состоит из восьми частей. Весь Интернет разделен на семь секторов, каждый условно обозначается цветом: *красным, оранжевым, желтым, зеленым, голубым, синим, фиолетовым*.

Сайт компании Rambler относится к *голубому сектору*. Информация о ресурсах каждого сектора хранится в соответствующей части индексной базы. Восьмая часть — «быстрая база» — включает в себя страницы, включенные в указатель *Top 100* и которые еще не успели попасть в основную индексную базу.

Все части собираются и обновляются по отдельности. Например, сегодня происходит переиндексация и обновление *красного* сектора, завтра — *оранжевого* и *желтого*, послезавтра — *зеленого* и т. д. Благодаря такому ступенчатому алгоритму в поисковой машине регулярно появляется свежая информация. Полный цикл обновления занимает около недели. При этом сбор информации происходит параллельно, а непосредственно на изготовление индекса документов одного сектора затрачивается несколько часов.

Разделение Интернета на семь секторов является условным. При необходимости он может быть разбит на 10, 20 или 40 секторов, каждый из которых будет обрабатываться автономно. В такой системе заложена возможность значительного увеличения нагрузки. С ростом объема информации в сети растет и индексная база поисковой машины. Постепенно переиндексация и сборка базы начинают занимать все больше времени, а процесс обновления индекса становится более громоздким. Поступление новых данных затягивается, информация начинает терять свою актуальность. Возможность «передела» Интернета на большее число секторов позволяет удерживать размер каждой части базы в оптимальном диапазоне, контролировать время ее сборки и обновления.

«Быстрая база» отличается от остальных частей индекса меньшим объемом и более оперативным обновлением — время ее построения занимает около двух часов. В базе содержится информация о страницах, на которых был установлен счетчик *Top 100*. Участниками рейтинга *Top 100* являются новостные порталы, сайты крупных компаний, Интернет-магазины, форумы — наиболее популярные ресурсы сети. Каждый раз при установке счетчика на новую страницу сайта, зарегистрированного в *Top 100*, информация пере-

дается в поисковую систему. Страница ищется во всех «цветах» основной базы и, если она еще не известна поисковой системе, отправляется в очередь на обработку. Перед обработкой страницы дополнительно фильтруются, и из них отбираются самые популярные (посещаемые).

Скорость поиска тесно связана с его чувствительностью к нагрузкам. В среднем в рабочие часы на поисковую машину Rambler в секунду поступает около 60 запросов. Такая загруженность требует сокращения времени обработки отдельного запроса.

Схематично обработка поискового запроса изображена на рис. 2.23.

Запрос поступает в поисковую систему через маршрутизатор Cisco 6000. Маршрутизатор передает его наименее загруженной машине первого уровня, например frontend-серверу 1.3. Frontend-сер-

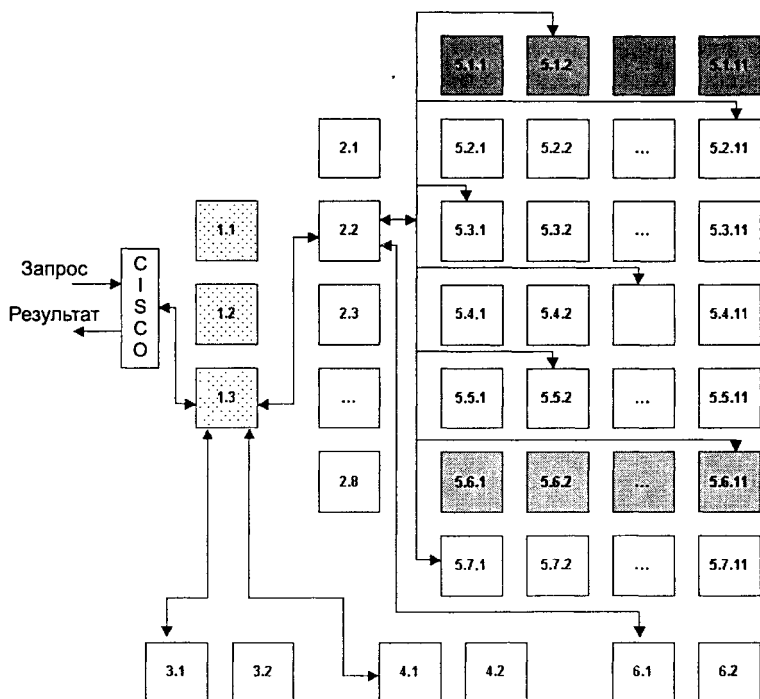


Рис. 2.23. Обработка поискового запроса в Rambler:

1.1—1.3 — Frontend-сервера; 2.1—2.8 — Proxy-сервера; 3.1—3.2 — поиск по товарам; 4.1—4.2 — поиск по Top 100; 5.1.1—5.7.11 — Backend-сервера, содержащие основную индексную базу; 6.1—6.2 — Backend-сервера, содержащие быструю базу

вер, в свою очередь, отправляет запрос дальше, на один из восьми проху-серверов, также выбирая наиболее свободный сервер (например, машине 2.2). Одновременно frontend-сервер отправляет запрос на машины, осуществляющие поиск по товарам (в данном случае — машине 3.1) и по базе *Top 100* (машине 4.1). На проху-сервере проводится поиск по ссылочному индексу, и его результаты вместе с поисковым запросом передаются на машины, которые содержат основную индексную базу, — backend-серверы (машинам 5.1.2, 5.2.11, 5.3.1 и т. д.) Та же информация отправляется на машины с «быстрой базой» (в данном случае — 6.1).

В поиск включено 77 backend-серверов. Они сгруппированы по 11 машин, и каждая группа содержит копию одной из частей поискового индекса. Таким образом, информация о сайтах, условно входящих в *красный* сектор Интернета, находится на backend-серверах первой группы (5.1.1—5.1.11), *оранжевый* сектор — на backend-серверах второй группы (5.2.1—5.2.11) и т. д.

Проху-сервер выбирает наименее загруженный backend-сервер в каждой группе машин и отправляет на него поисковый запрос с результатами ссылочного поиска. На backend-серверах осуществляется поиск по частям индексной базы и ранжирование с учетом результатов поиска по ссылочному индексу. При ранжировании для всех найденных документов высчитываются веса по конкретному запросу.

После того как запрос обработан на backend-серверах, информация о результатах и ранжировании отдается обратно на проху-сервер. Туда же поступают отсортированные результаты с машин «быстрой базы». Проху-сервер интегрирует данные, полученные с восьми машин: клеит дубли, объединяет зеркала сайтов, ранжирует документы в общий список по весам, рассчитанным на backend-серверах. Полученные результаты отдаются на frontend-сервер.

Помимо информации с проху-сервера, frontend-сервер получает результаты поисков по товарам и из базы *Top 100* (отсортированные, с цитатами и подсветкой слов запроса). Frontend-сервер осуществляет окончательное объединение результатов, генерирует выходной html-документ со списком найденного, вставляет баннеры и перевязки (ссылки на различные разделы Rambler'a) и передает HTML-документ маршрутизатору, который переадресовывает информацию пользователю.

Для повышения скорости поиска используется также «кэширование» (сохранение информации о запросах и результатах поиска в буфере). Многие пользователи обращаются с одинаковыми поисковыми запросами и «вычислять» их заново было бы неразумной тра-

той времени. Поэтому если аналогичный запрос обрабатывался до истечения некоторого интервала времени, результаты поиска отдаются пользователю из «кэша».

2.3. Экспертные системы

Идея подобных систем, способных сделать знания специалистов (экспертов) достоянием пользователей, зародилась в 50—60-е гг. XX в. По мнению многих исследователей, *экспертные системы (ЭС)* являются наиболее перспективным и быстроразвивающимся направлением *систем искусственного интеллекта (СИИ)*. Их применение в странах Запада распределяется следующим образом: 27 % — техническая диагностика, 18 % — военное дело, 18 % — финансовое планирование, 15 % — производственное планирование, по 5 % — медицина, электротехника, маркетинг, 7 % — прочие сферы. Ожидается расширение использования ЭС в управлении предприятием и техническом проектировании (особенно в проектировании АИС).

К СИИ относятся также

- управление роботами (робототехника);
- решение сложных математических задач (в том числе автоматическое доказательство теорем);
- автоматизация разработки программ (CASE-технологии);
- компьютерные игры;
- распознавание образов;
- распознавание речи.

Существует следующее принципиальное различие между классическими АИС и СИИ:

- в первом случае рассматриваются структуры данных, которые обрабатываются с помощью программ, основанных на машинной логике;
- во втором — системы работают с символическими представлениями знаний, причем ход обработки заранее не определен.

Развитие СИИ традиционно шло в двух направлениях:

- общеконцептуальном, предполагающем наличие некоторого единого принципа, на котором строится система в целом и из которого выводятся все частные решения;
- эвристическом, предполагающем наличие большого количества правил и ограничений, относящихся к конкретным ситуациям и образующих образ поведения СИИ в целом.

Примером первого подхода является применение языка Пролог (распространен преимущественно в Западной Европе), второго — языка Лисп (преимущественно в США). Интеграция подходов привела к появлению концепций *экспертных систем (ЭС)* и *баз знаний (БЗ)*.

Разработкой СИИ занимаются:

- крупные полиотраслевые фирмы — Texas Instruments, Digital Equipment, IBM, Hewlett Packard, General Motors, Xerox;
- небольшие специализированные — Artificial Intelligence, Computer Thought, Intellicorp, Machine Intelligence, Symbolics, Teknowledge и пр.

ЭС известны также как системы баз знаний (Knowledge Based Systems) — программные комплексы, способные решать узкопрофессиональные задачи наравне с человеком-экспертом. ЭС комбинируют программные технологии СИИ и БЗ, содержащие профессиональный опыт экспертов.

Подсистемы ЭКС

Подсистемами ЭКС являются (рис. 2.24):

- база знаний (БЗ) — Knowledge Base;
- механизм логического вывода (МЛВ) — Inference Engine;
- подсистема приобретения знаний (ППЗ) — Knowledge Acquisition Subsystem;
- подсистема разъяснения результатов, интерпретации ответов (ПРР) — Explaining Subsystem;
- интерфейс с человеком-пользователем (ИП) — Human Interface.

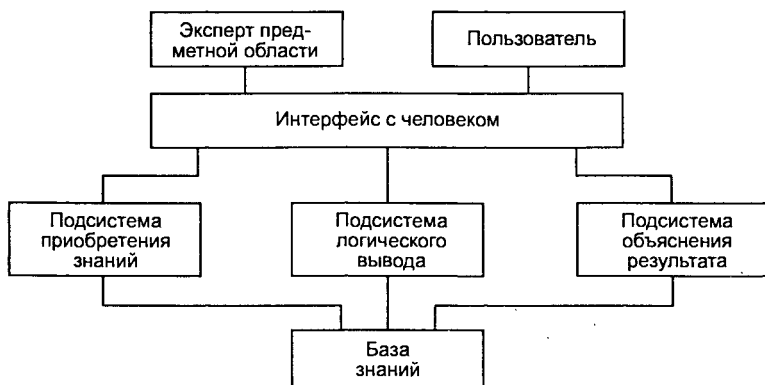


Рис. 2.24. Программно-аппаратурные компоненты экспертных систем

База знаний. Существуют различные методы представления знаний, из которых наиболее распространенным является *метод продукций или правил* — база правил, база продукций (БП). В отличие от обычных БД, где данные хранятся в точной, фактической форме, БЗ содержит как *факты* (безусловные утверждения), так и *условные высказывания* (правила вывода — Inference Rules — типа if <условие> then <действие>).

Факты, выводимые из правил, могут использоваться в качестве новых правил и пр. Важным свойством ЭС является возможность обработки неопределенной информации, которая лежит в основе всякой интеллектуальной деятельности. Существуют различные теории принятия решений, позволяющие осуществлять рациональный выбор в условиях неопределенности. Они, как правило, состоят из двух разделов: общего, связанного с раскрытием понятия «неопределенность», и нормативного, оценивающего меру неопределенности полученных выводов.

Это, в частности — теория вероятностей, методы размытых множеств, теория возможности, метод очевидности (Dempster-Shafer), формализм «групп знания» (связан с *сигнатурными таблицами Самуэля* и впервые использован в медицинской диагностической системе MDX) и пр. Фактор неопределенности может включаться в систему правил. Например, ЭС MYCIN игнорирует заключения с определенностью меньше 0,2.

Альтернативой БП являются системы фреймов (фреймы — структуры, аналогичные записям в реляционных БД) — полезны при обработке объектов и их классов. Кроме того, известны *средства логики предикатов, семантические сети, гибридные структуры*.

Лаборатории СИИ Массачусетского Технологического института и фирмы Хехох разрабатывают системы с базами моделей, которые могут делать выводы «по аналогии» (специалисты считают, что такого рода системы лучше соответствуют процессам мышления человека, чем ЭС с правилами).

Механизм логического вывода интерпретирует правила и выполняет логический вывод. Он независим и может использоваться с различными БЗ. Известны следующие типы МЛВ:

- управляемый данными (data driven, forward chaining) — пользователь вводит совокупность фактов и система делает выводы;
- управляемый целями (goal driven, backward chaining) — движение от целей, результатов к причинам, факторам. Этот процесс проще, так как не относящиеся к целям факты могут быть игнорированы;

- смешанные типы — пользователь выбирает тип задачи, а система может задавать вопросы, чтобы приобрести необходимые данные.

Подсистема разъяснения результатов. ЭС должна быть способной разъяснить пользователю линию рассуждений, если он запросит об этом, а также сообщить ему, какие сведения есть, а какие отсутствуют в БЗ.

Существуют следующие типы ПРР:

- объяснение по запросам пользователя;
- разъяснение стратегий системы при выводе заключений;
- критический анализ решений, предложенных пользователем;
- сообщение пользователю, какую дополнительную информацию он должен ввести в БЗ или указать в запросе.

Подсистема приобретения знаний. ЭС должна приобретать новые знания, удалять или модифицировать устаревшие — т. е. осуществлять редактирование БЗ. Поскольку БЗ отделена от МЛВ, она поддается автономному редактированию, что дает возможность развития ЭС в интерактивном режиме. Сначала создается прототип ЭС, который обрабатывает простые проблемы, затем БЗ расширяется. Идеальной ППЗ является автоматическая подсистема, действующая без вмешательства человека.

Интерфейс с человеком. Обеспечивает взаимодействие с непрограммирующим пользователем на языке, близком к естественному. ИП должен обеспечить как приобретение знаний при работе с экспертом, так и ввод запросов конечного пользователя и режим объяснения результатов.

Развитие ЭС

В 1970—1980-е гг. был разработан ряд прототипов ЭС (Knowledge Engineering Framework) — EMYCIN, ROSIE, KAS, EXPERT, OPS, AGE, HEARSAY — программные системы, настраивающиеся на предметную область путем загрузки в БЗ правил (продукций). Большинство таких систем в состоянии обрабатывать нечеткие правила и утверждения. Некоторые системы создавались без участия человека-эксперта, например TAXMAN была построена на основе справочников.

Существуют оболочки экспертных систем (Shell), реализующие ИП и МЛВ — EXSYS, Insight, ExpertEase и пр. БЗ и подсистема приобретения знаний должны разрабатываться пользователем. Эти системы позволяют обрабатывать до 5000 правил.

ЭС могут использоваться в следующих ролях: консультант, контролер, учитель, средство уточнения экспертизы, средство коммуникации, причем одну и ту же роль могут играть разные ЭС и наоборот;

ЭС могут использоваться для решения следующих классов проблем:

- проведение экспертизы, если специалист недоступен (*дубликат эксперта*);
- объединение мнений разных специалистов в одной БЗ (*объединенный эксперт*). В ЭС GINESIS использован механизм согласования знаний (правила подтверждения). Если правила БЗ, сформулированные в результате общения с разными экспертами, совпадают, то они принимаются. В противном случае включаются вероятностные механизмы согласования для разрешения конфликта;
- запись и использование знаний по проблеме для обучения и консультирования (*документирование*).

Пределы использования ЭС

- слишком большое число правил (более 10 тыс.) требует много времени на ввод БЗ и поиски ответа;
- ЭС применяется скорее для глубоких и узких, нежели для широких и поверхностных задач;
- ЭС не приспособлены для эффективного использования имеющихся БД, построенных по обычной технологии АИС;
- ЭС не приспособлены для проблем, по которым нет согласия экспертов;
- дружественные ИП для ЭС только разрабатываются.

Кроме обычных ЭС, обеспечивающих передачу знаний от экспертов менее подготовленным пользователям, разрабатываются также системы *разделения знаний*, которые рассчитаны на использование самими экспертами, вкладывающими в них свой опыт, и предназначены для предметных областей, включающих несколько различных систем знаний, не связанных друг с другом логически. В качестве механизма координации разнородных фрагментов знаний могут использоваться интуитивные суждения пользователей.

Системы передачи знаний предназначены для стимулирования работы экспертов путем обмена суждениями и мнениями о них. Добавление интуитивных суждений экспертов к их коллективному опыту позволяет выработать общий взгляд на предметную область и найти более эффективные решения отдельных проблем.

Применение ЭС

ЭС находят применение в следующих проблемных областях.

При разработке географических информационных систем (ГИС). Существуют четыре основные проблемные области ГИС, для которых ведется разработка ЭС: составление карт, выделение топографических признаков, географические СУБД и географические СОПР. Использование ЭС для управления географическими БД позволяет создать дружественный интерфейс в ГИС (напр., ORBI и SRAS) и ускорить процесс поиска (KBGIS-II). Отмечается тенденция использовать наиболее проверенные средства ИИ, такие, как Пролог, Лисп и командный язык ЭС EXSYS. Однако недостаточная производительность многих из них препятствует их широкому внедрению в ГИС. К фундаментальным проблемам разработки экспертных ГИС относится формализация географических знаний и неопределенностей. Предполагается, что в большинстве разработок ГИС в будущем будет использоваться методика ЭС.

Строительная индустрия обладает рядом существенных с точки зрения информационного обеспечения специфических особенностей: сочетание элементов науки, техники, искусства, строго определенных и «расплывчатых» постановок задач; многоотраслевой характер применяемых знаний; каждый проект требует создания специальной информационной базы «с нуля», ее непрерывного роста и обновления; значительная часть информации относится только к данному проекту и теряет смысл после его осуществления.

Разведывательная деятельность и стратегические исследования являются подходящей областью для применения ЭС, основанных на БЗ, поскольку связаны с обработкой все возрастающего количества недостоверных, противоречивых, а иногда дезинформирующих данных. Основные преимущества применения ЭС в данной области — сокращение среднего времени анализа данных и быстрое обнаружение противоречивых или подозрительных ситуаций; ускорение обучения и повышения квалификации оперативного состава в процессе взаимодействия с ЭС.

Известны примеры применения ЭС для сопровождения проектно-конструкторских работ (ПКР):

- MENTOR — система анализа и поиска ошибок в проектах систем охлаждения и кондиционирования;
- WINDKS — система обработки запросов и выдачи справок по общим проблемам строительства и пр.

Особенности информационных систем, обеспечивающих ПКР:

- предоставление нужной информации только на соответствующие фазы проекта;
- обеспечение доступа к справочным, графическим, номенклатурным данным;
- оказание помощи в критическом анализе проектного решения;
- предоставление информации о готовых аналогичных решениях с указанием технологических и экономических сведений;
- широкий спектр предоставляемой информации (от общих вопросов ПКР до специфически конкретных данных).

В стадии разработки находятся проекты интеграции реляционных БД и систем логического программирования, связанные с созданием новых информационных технологий и вычислительных систем, однако уже сегодня существуют промышленные CPR-системы (Coupling Prolog to Relational databases) для связывания Пролога и реляционных БД. CPR-системы предназначены не только для построения интенциональной БД из экстенциональной, но также для создания БЗ с привлечением эвристик.

Методика построения ЭС

В настоящее время сложилась определенная технология разработки ЭС, основой которой являются следующие шесть этапов: идентификация, концептуализация, формализация, выполнение, тестирование и опытная эксплуатация (рис. 2.25).

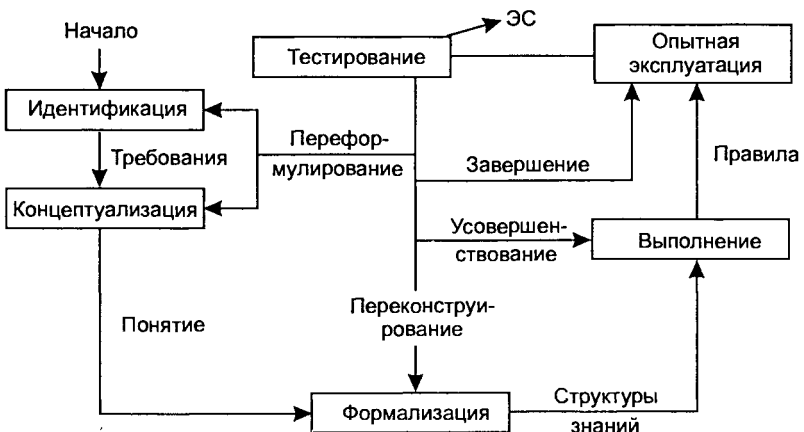


Рис. 2.25. Этапы разработки ЭС

Этап идентификации. Этап связан, прежде всего, с осмыслением задач, которые предстоит решить будущей ЭС, и формированием требований к ней. Результатами являются — идентификация задачи, определение участников процесса проектирования и их роли, выявление ресурсов и целей.

Обычно в разработке ЭС участвуют не менее трех-четырёх человек — один эксперт, один или два инженера по знаниям и один программист, привлекаемый для модификации и согласования инструментальных средств. Также к процессу разработки ЭС могут по мере необходимости привлекаться и другие участники. Например, инженер по знаниям может пригласить дополнительных экспертов, чтобы убедиться в правильности своего понимания основного эксперта, представительности тестов, демонстрирующих особенности рассматриваемой задачи, совпадения различных взглядов на качество предлагаемых решений. Кроме того, для сложных систем считается целесообразным привлекать к основному циклу разработки несколько экспертов. Однако в этом случае, как правило, требуется, чтобы один из них отвечал за непротиворечивость знаний, сообщаемых коллективом экспертов.

Идентификация задачи заключается в составлении неформального (вербального) описания, в котором указываются:

- общие характеристики задачи;
- подзадачи, выделяемые внутри данной задачи;
- ключевые понятия (объекты), их входные (выходные) данные;
- предположительный вид решения, а также знания, относящиеся к решаемой задаче.

В процессе идентификации задачи инженер по знаниям и эксперт работают в тесном контакте. Начальное неформальное описание экспертом используется инженером по знаниям для уточнения терминов и ключевых понятий. Эксперт корректирует описание задачи, объясняет, как и на основе каких рассуждений она решается. После нескольких циклов, уточняющих описание, эксперт и инженер по знаниям получают окончательное неформальное описание задачи.

При проектировании ЭС типичными ресурсами являются источники знаний, время разработки, вычислительные средства и объем финансирования. Для эксперта источниками знаний служат его предшествующий опыт, книги, известные примеры решения задач, а для инженера по знаниям — опыт в решении аналогичных задач, методы представления знаний и манипулирования ими, программные инструментальные средства. При определении длительности работ обычно имеется в виду, что сроки разработки и

внедрения ЭС составляют, как правило, не менее года (при трудоемкости 5 чел.-лет). Определение объема финансирования оказывает существенное влияние на процесс разработки (например, при недостаточном финансировании предпочтение может быть отдано не разработке оригинальной новой системы, а адаптации существующей).

При идентификации целей важно отличать те, ради которых создается ЭС, от задач, которые она должна решать. Примерами возможных целей являются: формализация неформальных знаний экспертов; улучшение качества решений, принимаемых экспертом; автоматизация рутинных аспектов работы эксперта (пользователя); тиражирование знаний эксперта.

Этап концептуализации. На данном этапе проводится содержательный анализ проблемной области, выявляются используемые понятия и их взаимосвязи, определяются методы решения задач. Этот этап завершается созданием модели предметной области (ПО), включающей основные концепты и отношения.

На этапе концептуализации определяются следующие особенности задачи:

- типы доступных данных;
- исходные и выводимые данные, подзадачи общей задачи;
- стратегии и гипотезы;
- виды взаимосвязей между объектами ПО, типы отношений (иерархия, причина — следствие, часть — целое и т. п.);
- процессы, используемые в ходе решения;
- состав знаний, необходимых для решения задачи;
- типы ограничений, накладываемых на процессы решения;
- состав знаний, используемых для обоснования решений.

Существует два подхода к процессу построения модели предметной области, которая является целью разработчиков ЭС на этапе концептуализации.

1. *Признаковый или атрибутивный* — предполагает как наличие полученной от экспертов информации в виде троек «объект — атрибут — значение атрибута», так и обучающей информации. Этот подход развивается в рамках направления, получившего название формирование знаний или «машинное обучение» (machine learning).

2. *Структурный (или когнитивный)*, осуществляется путем выделения элементов предметной области, их взаимосвязей и семантических отношений.

Для *атрибутивного подхода* характерно наличие наиболее полной информации о предметной области (объекты, атрибуты и их значения). Кроме того, существенным моментом является исполь-

зование дополнительной обучающей информации, которая задается группированием объектов в классы по тому или иному содержательному критерию.

Тройки *объект — атрибут — значение атрибута* могут быть получены с помощью метода *реклассификации*, который основан на предположении о том, что задача является объектно-ориентированной и они хорошо известны эксперту. Идея метода состоит в том, что конструируются правила (комбинации значений атрибутов), позволяющие отличить один объект от другого. Обучающая информация может быть задана на основании прецедентов правильных экспертных заключений, например с помощью метода извлечения знаний, получившего название «анализ протоколов мыслей вслух».

При наличии обучающей информации для формирования модели предметной области на этапе концептуализации можно использовать весь арсенал методов, развиваемых в рамках задачи распознавания образов.

Структурный подход к построению модели предметной области предполагает выделение следующих элементов знаний:

- понятия;
- взаимосвязи;
- метапонятия;
- семантические отношения.

Выделяемые *понятия* предметной области должны образовывать систему в виде совокупности, обладающей следующими свойствами:

- уникальностью (отсутствием избыточности);
- полнотой (достаточно полным описанием различных процессов, фактов, явлений и т. д. предметной области);
- достоверностью (валидностью — соответствием выделенных единиц смысловой информации их реальным наименованиям)
- непротиворечивостью (отсутствием омонимии).

Рассмотрим методы построения *системы понятий*.

Метод локального представления. Эксперта просят разбить задачу на подзадачи для перечисления целевых состояний и описания общих категорий цели. Далее для каждого разбиения (локального представления) эксперт формулирует информационные факты и дает им четкое наименование (название). Считается, что для успешного решения задачи построения модели предметной области число таких информационных фактов в каждом локальном представлении, которыми человек способен одновременно манипулировать, должно быть примерно равно семи.

Метод вычисления коэффициента использования основан на следующей гипотезе. Элемент данных (или информационный факт) может являться понятием, если он используется:

- в большом числе подзадач;
- с большим числом других элементов данных;
- редко совместно с другими элементами данных по сравнению с общим числом его использования во всех подзадачах (что и составляет коэффициент использования).

Полученные значения могут служить критерием для классификации всех элементов данных и, таким образом, для формирования системы понятий.

Метод формирования перечня понятий заключается в том, что экспертам (желательно, чтобы их было больше двух) дается задание составить список понятий, относящихся к исследуемой предметной области. Выделенные ими понятия включаются в систему, остальные подлежат обсуждению.

Ролевой метод состоит в том, что эксперту дается задание обучить инженера по знаниям решению некоторых задач предметной области. Таким образом, эксперт играет роль учителя, а инженер по знаниям — роль ученика.

Например, процесс обучения записывается на магнитофон, а затем третий участник прослушивает магнитофонную ленту и выписывает на бумаге все понятия, употребленные учителем или учеником.

Метод составления списка элементарных действий. При этом эксперту дается задание составить такой список при решении задачи в произвольном порядке.

Метод составления оглавления учебника. Эксперту предлагается представить ситуацию, в которой его попросили написать учебник. Необходимо составить на бумаге перечень предполагаемых глав, разделов, параграфов, пунктов и подпунктов книги.

Текстологический метод формирования системы понятий заключается в том, что эксперту дается задание выписать из руководств (книг по специальности) некоторые элементы, представляющие собой единицы смысловой информации.

Группа методов *установления взаимосвязей* предполагает установление семантической близости между отдельными понятиями. В основе определения взаимосвязей лежит психологический эффект «свободных ассоциаций», а также фундаментальная категория близости объектов или концептов.

Эффект свободных ассоциаций заключается в следующем. Испытуемого просят отвечать на заданное слово первым пришедшим на ум словом. Как правило, реакция большинства испытуемых (если

слова не были слишком необычными) оказывается одинаковой. Количество переходов в цепочке может служить мерой «смыслового расстояния» между двумя понятиями. Многочисленные опыты подтверждают гипотезу, что для двух любых слов (понятий) существует ассоциативная цепочка, состоящая не более чем из семи слов.

Метод свободных ассоциаций основан на психологическом эффекте, описанном выше. Эксперту предъявляется понятие с просьбой назвать как можно быстрее первое пришедшее на ум понятие из сформированной ранее системы, далее производится анализ полученной информации.

Метод обнаружения регулярностей основан на гипотезе о том, что элементы цепочки понятия, которые человек вспоминает с определенной регулярностью, имеют тесную ассоциативную взаимосвязь. Для эксперимента произвольным образом отбирается 20 понятий. Эксперту предъявляется одно из числа отобранных. Процедура повторяется до двадцати раз, причем каждый раз начальные концепты должны быть разными. Затем инженер по знаниям анализирует полученные цепочки с целью нахождения постоянно повторяющихся понятий (регулярностей). Внутри выделенных таким образом группировок устанавливаются ассоциативные взаимосвязи.

Кроме рассмотренных выше неформальных методов для установления взаимосвязей между отдельными понятиями применяются также формальные методы. Сюда в первую очередь относятся *методы семантического дифференциала и репертуарных решеток*.

Выделенные понятия предметной области и установленные между ними взаимосвязи служат основанием для дальнейшего построения системы *метапонятий* — осмысленных в контексте изучаемой предметной области системы смысловых группировок, для определения которых применяют как неформальные, так и формальные методы.

Последним этапом построения модели предметной области при концептуальном анализе является установление *семантических отношений* между выделенными понятиями и метапонятиями. Установить семантические отношения — значит, определить специфику взаимосвязи, полученной в результате применения тех или иных методов. Для этого необходимо каждую зафиксированную взаимосвязь осмыслить и отнести ее к тому или иному типу отношений.

Существует значительное множество базовых отношений, например «часть — целое», «род — вид», «причина — следствие», пространственные, временные и другие отношения (см. также гл. 1). Для каждой предметной области помимо общих базовых отношений могут существовать и специфические, уникальные отношения.

Прямой метод установления семантических отношений основан на непосредственном осмыслении каждой взаимосвязи. В том случае, когда эксперт затрудняется дать интерпретацию выделенной взаимосвязи, ему предлагается следующая процедура. Формируются тройки: *понятие-1 — связь — понятие-2*. Рядом с каждой тройкой записывается короткое предложение или фраза, построенное так, чтобы *понятие-1* и *понятие-2* входили бы в это предложение. В качестве связок используются только содержательные отношения и не применяются неопределенные связки наподобие *похож на* или *связан с*.

Для *косвенного метода* не обязательно знать взаимосвязи, а достаточно лишь наличия системы понятий. Формулируется некоторый критерий, для которого из системы понятий выбирается определенная *совокупность концептов*. Последняя предъявляется эксперту с просьбой дать вербальное описание сформулированного критерия. Концепты предъявляются эксперту все сразу (желательно на карточках). В случае затруднений эксперта прибегают к разбиению отобранных концептов на группы с помощью более мелких критериев. Исходное количество концептов может быть произвольным, но после разбиения на группы в каждой из таких групп должно быть не более десяти концептов. После того как составлены описания по всем группам, эксперту предлагают объединить эти описания в одно.

Рассмотренные выше методы формирования системы понятий и метапонятий, установления взаимосвязей и семантических отношений в разных сочетаниях применяются на этапе концептуализации при построении модели предметной области.

Этап формализации. Теперь все ключевые понятия и отношения выражаются на некотором формальном языке, который либо выбирается из числа уже существующих, либо создается заново. Другими словами, на данном этапе определяются состав средств и способы представления декларативных и процедурных знаний, после осуществления которого формируется описание решения задачи ЭС на предложенном (инженером по знаниям) формальном языке.

Сюда относится указание способов представления (фреймы, сценарии, семантические сети и т. д.) и определение способов манипулирования и интерпретации знаний (логический вывод, аналитическая модель, статистическая модель и др.).

Этап выполнения. Цель этого этапа — создание одного или нескольких прототипов ЭС, решающих требуемые задачи. Затем на данном этапе по результатам тестирования и опытной эксплуатации создается конечный продукт, пригодный для промышленного использования. Разработка прототипа состоит в программировании

его компонентов или выборе их из известных инструментальных средств и наполнении базы знаний.

Главное в создании прототипа заключается в том, чтобы обеспечить проверку адекватности идей, методов и способов представления знаний решаемым задачам. Создание первого прототипа должно подтвердить, что выбранные методы решений и способы представления пригодны для успешного решения, по крайней мере, ряда задач из актуальной предметной области, а также продемонстрировать тенденцию к получению высококачественных и эффективных решений для всего круга задач по мере увеличения объема знаний.

После разработки первого прототипа ЭС-1 круг предлагаемых для решения задач расширяется, и собираются пожелания и замечания, которые должны быть учтены в очередной версии системы ЭС-2. Осуществляется развитие ЭС-1 путем добавления дружественного интерфейса, средств для исследования базы знаний и цепочек выводов, генерируемых системой, а также для сбора замечаний пользователей и средств библиотеки решенных задач.

Выполнение экспериментов с расширенной версией ЭС-1, анализ пожеланий и замечаний служат отправной точкой для создания второго прототипа ЭС-2. Процесс разработки ЭС-2 — итеративный. Он может продолжаться от нескольких месяцев до нескольких лет в зависимости от сложности предметной области, гибкости выбранного представления знаний и степени соответствия управляющего механизма решаемым задачам (возможно, потребуется разработка ЭС-3 и т. д.). При разработке ЭС-2, кроме перечисленных задач, решаются следующие:

- анализ функционирования системы при значительном расширении базы знаний;
- исследование возможностей системы в решении более широкого круга задач и принятие мер для обеспечения таких возможностей;
- анализ мнений пользователей о функционировании ЭС;
- разработка системы ввода-вывода, осуществляющей анализ или синтез предложений ограниченного естественного языка и позволяющей взаимодействовать с ЭС-2 в форме, близкой к стандартным учебникам для данной области.

Если ЭС-2 успешно прошла этап тестирования, то она может быть классифицирована как промышленная экспертная система.

Этап тестирования. В ходе данного этапа производится оценка выбранного способа представления знаний в ЭС в целом. Для этого инженер по знаниям подбирает примеры, обеспечивающие проверку всех возможностей разработанной ЭС.

Различают следующие источники неудач в работе системы: тестовые примеры, ввод-вывод, правила вывода, управляющие стратегии.

Показательные тестовые примеры являются наиболее очевидной причиной неудачной работы ЭС. В худшем случае они могут оказаться вообще вне предметной области, на которую рассчитана ЭС, однако чаще их множество оказывается слишком однородным и не охватывает всю предметную область, поэтому при подготовке тестовых примеров следует классифицировать их по подпроблемам, выделяя стандартные случаи, определяя границы трудных ситуаций и т. п.

Ввод-вывод характеризуется данными, приобретенными в ходе диалога с экспертом, и заключениями, предъявленными ЭС в ходе объяснений. Методы приобретения данных могут не давать требуемых результатов (например, если задавались неправильные вопросы или собрана не вся необходимая информация). Кроме того, вопросы системы могут быть трудными для понимания, многозначными и не соответствующими знаниям пользователя. Ошибки при вводе могут возникать также из-за неудобного для пользователя входного языка. В ряде приложений целесообразно реализовать ввод информации не только в печатной, но и в графической или звуковой форме.

Выходные сообщения (заключения) системы могут оказаться непонятными для пользователя (эксперта) по разным причинам. Например, их может быть слишком много или, наоборот, мало. Также причиной ошибок может являться неудачная организация, упорядоченность заключений или не подходящий для пользователя уровень абстракций с непонятной лексикой.

Наиболее распространенный источник ошибок в рассуждениях касается правил вывода. Причинами могут являться отсутствие учета взаимозависимости сформированных правил, либо их ошибочность, противоречивость и неполнота. Если неверна посылка правила, то это может привести к употреблению его в неподходящем контексте. Если ошибочно действие правила, то трудно предсказать конечный результат. Правило может быть ошибочно, если при корректности его условия и действия нарушено соответствие между ними.

Нередко к ошибкам в работе ЭС приводят применяемые управляющие стратегии. Изменение стратегии бывает необходимо, например, если ЭС анализирует сущности в порядке, отличном от «естественного» для эксперта. Последовательность, в которой данные рассматриваются ЭС, не только влияет на эффективность работы системы, но и может приводить к изменению конечного результата. Так, рассмотрение *правила А* до *правила В* способно привести к тому, что последнее всегда будет игнорироваться системой.

Этап опытной эксплуатации. На этом этапе проверяется пригодность ЭС для конечного пользователя. Она определяется в основном удобством работы и полезностью. Под полезностью ЭС понимается ее способность в ходе диалога определять и удовлетворять потребности пользователя (решать поставленные задачи), выявлять и устранять причины неудач в работе. В свою очередь, удобство работы с ЭС подразумевает:

- естественность взаимодействия с ней (общение в привычном, не утомляющем пользователя виде);
- гибкость (способность системы настраиваться на различных пользователей, а также учитывать изменения в квалификации одного и того же пользователя);
- устойчивость системы к ошибкам (способность не выходить из строя при ошибочных действиях неопытного пользователя).

В ходе разработки ЭС почти всегда осуществляется ее модификация. Выделяют следующие виды модификации системы: переформулирование понятий и требований, переконструирование представления знаний в системе и усовершенствование прототипа.

2.4. Гипертекстовые АИС

Слово *гипертекст* (*hypertext*) буквально переводится как нелинейный текст (*nonlinear text*). Элемент гипертекста — *узел*, дискретный объект. Узлы, между которыми возможен переход, считаются *смежными*, а сама возможность перехода называется *связью*. Для описания и анализа объектов систем, в которых основным отношением является смежность (непосредственная связь) элементов, используется *теория графов*.

В системе гипертекста ссылки вида «смотри также», подстрочные примечания, библиографические ссылки, внутритекстовые и пометы, моделируются как ассоциативные связи. Следуя последним, можно читать материал в любом порядке. Если речь идет о достаточно обширном материале с большим количеством связей, то возникает сложное гипертекстовое пространство (*сеть*). Формирование, поддержание, исправление, наращивание и просмотр такой сети практически возможны только на компьютерной основе.

Гипертекстом часто называют как саму форму структурирования текстового материала (нелинейную, сетевую), так и технологию, без которой невозможна такая организация материала в широких масштабах.

История развития гипертекстовой технологии

Считается, что впервые идею гипертекста выдвинул В. Буш (V. Bush), советник президента Рузвельта по науке, в статье «Как мы могли бы мыслить» (As we may think) для журнала «Atlantic Monthly». В ней он представлял проект «memex» («MEMOry EXtention») как автоматизированное бюро, в котором человек хранил бы свои книги, записи, любую получаемую им информацию таким образом, чтобы в любой момент можно было воспользоваться ее с максимальной быстротой и удобством. Фактически это должно было быть сложное устройство, снабженное клавиатурой и прозрачными экранами, на которые бы проецировались тексты и изображения, хранящиеся на микрофильмах. В «memex» следовало установить логические и ассоциативные связи между любыми двумя блоками информации. В идеале речь шла о громадной библиотеке, универсальной информационной базе данных, в которой можно было бы не только получать информацию из любой точки земного шара, но и иметь удобный способ связи информационных сегментов друг с другом, так чтобы наиболее важные данные быстро смогли быть найдены.

В определенной степени эти возможности реализованы в настоящее время в сети Интернет в рамках «всемирной паутины» WWW (World Wide Web), базирующейся на протоколе HTTP (Hypertext Transfer Protocol), формате документов HTML (Hypertext Markup Language), программах-браузерах и других средствах [24].

Проект Буша был первым проектом технического устройства, специальной функцией которого была фиксация связей между смысловыми элементами информации и доступ к этим элементам по системе связей.

В 1963 г. Д. Энгельбарт (D. Engelbart) написал работу «Концептуальная интегрированная система для расширения человеческого интеллекта» (A conceptual framework for the augmentation of man's intellect). Он предложил проект системы, в которой пользователь и компьютер динамически обмениваются информацией, что создает эффект расширения естественного интеллекта пользователя. Спустя пять лет в Стэндфордском исследовательском институте была создана система NLS (oN-Line System), реализовавшая в более конкретной форме идеи Энгельбарта. Это была мощная система обработки текста и электронной коммуникации, которая позволяла нескольким пользователям обмениваться планами, записями, программами, устанавливать взаимосвязи между смежными по содержанию идеями, соображениями, справочными данными, документами, состав-

лять указатели и т. п. Данная система до сих пор еще используется в ВВС США для хранения и обработки технической документации.

Крупного успеха добился в 1967 г. Т. Нельсон (Т. Nelson) в университете Брауна (шт. Род-Айленд, США): в его лаборатории была разработана гипертекстовая система (Hypertext Editing System), которая применялась при ведении документации по проекту космического корабля «Apollo». Она являлась прообразом распределенной информационной системы, электронных конференций, систем автоматизированного проектирования (САПР) и других аналогичных реализаций.

Однако наибольшую известность получила работа Нельсона над проектом «Xanadu», близким по направленности к «memex», но ориентированным уже на компьютерную технику. Основная цель этого проекта состояла в том, чтобы способствовать революционному процессу перевода всей мировой литературы в такую форму, при которой возможен непосредственный доступ к любому произведению, извлечение из него любого фрагмента, включение его в новый смысловой ряд.

«...Мы создаем систему хранения и актуализации текстов со связями и окнами, — писал Т. Нельсон. — Документ, наша фундаментальная единица, может иметь окна к любым другим документам. Состав документов может постоянно расширяться, не требуя фундаментальных изменений структуры. Благодаря новым связям и окнам постоянно добавляются новые возможности доступа к старым материалам». Совокупность текстов, связанных по содержанию, внутри которой читатель может легко переходить от каждого из них к другим, была названа *гипертекстом* (термин ввел Т. Нельсон). К основным признакам гипертекста Нельсон относил возможность пользователя «делать заметки, отмечать направления своих поисков, а также возможность внесения в текст тех или иных изменений».

Проекты Энгельбарта и Нельсона носили весьма глобальный характер. Но они дали импульс практическим работам по созданию технических систем, делающих смежные тексты, доступными для читателя.

Отличительные черты гипертекста

В гипертекстовых системах база данных не строится в соответствии с какой-либо заранее (до накопления информации) установленной схемой связей, а организуется в виде открытой, свободно наращиваемой и изменяемой самим пользователем сети, узлы кото-

рой соединяются с уже имеющимися в базе узлами. Следовательно, гипертекст не требует предварительной формализации знания, но предполагает лишь его дискретность и возможность явного указания имеющихся связей между смысловыми единицами. К достоинствам гипертекста относят широкие возможности автоматизированного обучения. Гипертекст позволяет не только просмотреть большую группу релевантных документов, но и изучить механизм образования ассоциативных связей.

Классические системы и средства информационного поиска ориентированы, в первую очередь, на пользователя, желающего восполнить некий пробел в своих знаниях, который он может точно сформулировать. Однако многие пользователи не в состоянии четко сформулировать свои информационные потребности или просто желают подробнее ознакомиться с малоизвестной предметной областью. В этом случае поиск по дескрипторам или ключевым словам малоэффективен, часто они просто неизвестны пользователю.

В принципе, пользователь может работать с гипертекстовой системой без знания специальных языков поиска и запроса. Вместо них используется «браузинг» («browsing») — операция просмотра узлов гипертекстовой сети по связи, движение от известной информации к связанной с нею неизвестной, являющейся предметом поиска. Браузинг дает пользователю относительно легкий способ отыскания релевантной информации без изучения запросного языка независимо от объемов БД.

Другой отличительной чертой гипертекста является способность интегрировать разнотипные виды информации (текст, таблицы, векторную и растровую графику, мультимедиа) в единую гиперсреду (hypermedia).

Основные компоненты гипертекстовой системы

Системы гипертекста можно определить как системы, обеспечивающие создание нелинейных документов и взаимодействие с ними, или как системы ассоциативной организации и поиска информации. Структурно гипертекстовая система включает в себя:

- графический интерфейс, обеспечивающий пользователю навигацию через широкие массивы информации, активацию связей и чтение содержания узлов с помощью окон просмотра и перекрывающихся диаграмм;
- систему автора гипертекста, т. е. средства создания и управления узлами и связями;

- традиционный информационно-поисковый (IR) механизм: поиск по ключевым словам (КС), авторский поиск и т. д.;
- гипермедиа-машину управления информацией по узлам и связям;
- систему хранения: файловую систему или базу знаний, реляционную или объектно-ориентированную СУБД.

Программные средства гипертекста подразумевают многооконные интерфейсы, многоуровневые меню, возможность просмотра структуры текста, ведение версий, встроенные средства программирования, обработки транзакций, восстановления после сбоев и т. д.

Гипертекстовая система предназначена для интерактивного взаимодействия авторов, администраторов и читателей (пользователей).

Автор выбирает ключевые слова в тексте обрабатываемого документа и выделяет их для пользователя, задает переходы между текстами, систему этих переходов. Он обязан знать не только сам базовый материал, но и все возможные пути, которые предполагаемый оператор (читатель), возможно, выберет для использования.

Системный администратор-программист определяет методы, с помощью которых эти перемещения выбираются. *Пользователю* дана свобода выбора маршрута чтения. Ему предоставлено хотя и большое, но конечное число передвижений, определенных автором.

Для предотвращения таких негативных явлений, как дезориентация пользователя в большом количестве связей, интеллектуальная перегрузка, пользовательский интерфейс должен иметь средства руководства, которые могли бы обеспечить выбор надежных критериев и верных альтернатив в ходе принятия определенных решений в рамках избранных поисковых стратегий; запоминание «путей», обеспечение возможностей ориентации пользователя.

Выделим следующие функции гипертекстовых информационных систем.

1. Поиск текста по атрибутам, символьным строкам, создание, хранение и поиск различного рода подборок текста («папок»).
2. Поддержка ссылочных связей.
3. Поиск информации путем браузинга (быстрый просмотр).
4. Ведение наращиваемой системы документов вне априорно заданной структуры.
5. Навигация по связям с целью изучения и освоения знания о соответствующей предметной области.
6. Выделение в ходе навигации «виртуальных структур».
7. Автоматическое построение из гипертекстовой сети связанных текстов с использованием эвристических алгоритмов.

Функциональные характеристики систем гиперзаписи информации:

1. Интеграция данных.
2. Свобода доступа к информации (свобода визуального просмотра).
3. Функциональная гибкость системы — обеспечение возможности комплексного, разностороннего использования информации, формирование новых продуктивных поисковых стратегий.
4. Открытость, потенциальная возможность развития системы.
5. Многопользовательский разделенный доступ к ГБД.

Гипертекстовые базы данных

Обычно в гипертекстовой БД (ГБД) используются два типа данных: узел и связь. Для узла задаются *поля*:

- самостоятельные элементы в узлах (заголовок, основное содержание);
- версии (тогда узел в целом инвариантен);
- атрибуты (ключевые слова — атрибуты семантики, атрибуты логической роли — аргумент, факт, время ввода информации, личные пометки пользователя, число связей).

Для связей указываются *атрибуты* (тип связей — прямая ссылка, тематическая близость, смысловая смежность, родовидовая, причинно-следственная связь и т. п.). Идентификатором может служить, например, последовательный номер при вводе. В случае реляционной модели используют два варианта организации: отдельно для узлов и связей, или все в одной БД.

Широкие возможности при проектировании ГБД обеспечивает использование объектно-ориентированного программирования [5, 6]. Объектно-ориентированные СУБД являются хорошим инструментом для выполнения таких требований к гипермедиа системам, как то:

- открытость и распределенность. Открытость предполагает способность объединять разные информационные системы, стандартную аппаратуру и программные продукты различных производителей, что делает возможным обмен между ними данными; распределенность — доступ к информационным ресурсам независимо от местоположения пользователя;
- поддержка коллективной работы или соучастие, т. е. обеспечение возможности формирования и исследования гипертекста в интерактивном режиме несколькими пользователями; группового решения сложных неструктурированных проблем;

- целостность данных/корректность;
- поисковый и запросный механизм. В дополнение к навигационному доступу к информации, системы гипермедиа должны обеспечивать и запросный механизм для эффективного поиска. Возможны два вида запросов: структурный — для нахождения части сети, семантический — для поиска специального узла;
- вычисление, включает свойство наследования, поиск основных правил и вывод заключений;
- поддержка мультимедиа;
- расширяемость и приспособляемость (с помощью интерфейса программиста).

В гипербазе (HyperBase) основной класс — это объект гипербазы (HB_Object), который имеет три подкласса: узел (HB_Node), связь (HB_Link), составной объект (HB_Composite_Object). Операции создания, модификации, копирования, уничтожения, восстановления объекта (части или целого) поддерживаются для узлов, связей, составных объектов и атрибутов.

Информационный поиск в гипертекстовых массивах

Навигация. Движение в гипертекстовой сети, совершаемое в процессе чтения гипертекста, получило название *навигации*. В процессе навигации различают *медленное чтение* (сопровождается заметками) и *быстрое* (browsing — пролистывание). Браузинг эффективен в случае, когда читатель не знает точно, какими признаками характеризуются необходимые ему сведения. Различают навигацию, ориентированную *глобально* (читателю демонстрируется графический образ сети и положение текущего узла) и *локально* (демонстрируется только информация об окрестностях данного узла, выделение слов, оглавлений или начальных частей смежных фрагментов).

Повышает эффективность навигации справочный аппарат, который располагает возможностью выдавать рекомендации, отмечать пройденные и последующие узлы, организовывать экскурсии по системе, собирать информацию по ее использованию, строить типовые маршруты, предлагаемые впоследствии.

В ходе навигации создается и запоминается след — trail, который фактически является результатом поиска в гипертекстовой

сети. Возможна также автоматическая навигация, когда построение цепочек узлов идет по определенным критериям.

Навигация или браузеринг являются эффективными только для небольших гипертекстовых систем. Для обширных ГБД решающее значение играет поиск через запросы.

Запросы и поисковый механизм. Классические АИПС базируются на автоматическом поиске по ключевым словам, соединенным булевыми операторами. Однако лишь немногие запросы обеспечивают полный и точный поиск информации, выдача может быть как избыточной, так и пустой, информационный поиск по существу является неопределенным процессом. Комбинация поисковых методов может устранить или минимизировать эту неопределенность. В гипертекстовых системах комбинируется поиск по ключевым словам и по гипертекстовым связям — поиск проводится в подмножестве семантически близких узлов. Поисковая стратегия может быть разделена на поиск по содержанию и по структуре.

Поиск по содержанию — это классический способ, распространенный на гипертекстовые системы — все узлы и связи обрабатываются независимо и исследуются для сопоставления с поданным запросом.

Структурный поиск выделяет гипертекстовую подсеть, которая сопоставляется с поданным образцом. Комбинация запросов будет обеспечивать действие фильтров. Просмотр через фильтры используется в системах NoteCards и Neptune.

Запросы и индексы по содержанию. Обычно для гипертекстовых документов реализуется двухуровневая структура:

- гипериндекс (индексная информация);
- уровень гипербаза (собственно узлы и связи).

Гипериндекс состоит из множества индексов, соединенных вместе. Запрос описывается через термины индекса, затем идет обращение к гипербазе. Навигацию через гипериндекс и поиск информации в гипербазе называют *запрос через навигацию*.

Индекс задает множество индексных входов. Каждый индекс состоит из дескриптора (термина) или ключевого слова и локатора (подобие номера страницы). Поиск элементов гипертекста по ключевым словам можно рассматривать как нахождение неявных ссылок — фрагменты общими ключевыми словами можно полагать неявно семантически связанными.

Структурные запросы. В этом случае узлы трактуются не как фрагменты текста, а как некоторые логические единицы (мысли, высказывания, понятия, единицы знания). Тогда появляется возможность образования из двух высказываний, соответст-

вующих узлам, сложных высказываний с помощью связок. При этом важен факт корректного образования сложного высказывания, разрешением которого занимаются при разработке смысловых процессоров.

Примером может быть визуальный гипертекстовый язык запросов GraphLog, осуществленный в рамках внешнего интерфейса системы Neptune для *абстрактной машины гипертекста*. Он обращается к следующим понятиям: поисковый запросный механизм, приращение основной модели узлов и связей, виртуальные структуры, вычисления над гиперсетью, ведение версий. Используя GraphLog, запросы формулируются через рисованные графические образцы, которые появляются тогда, когда поиск выдает подграф гипертекстовой сети.

Язык обеспечивает спецификацию и манипуляцию произвольными подмножествами сети и поддерживает вычисление функций агрегации в подграфах гипертекстового документа. Он может поддерживать динамически заданные структуры и выражать структурные запросы, которые не доступны для традиционных языков реляционной алгебры. Так, он может выполнить упорядоченность произвольной длины операций соединения, невозможную в рамках реляционной алгебры.

Сети вывода. Известна модель Крофта и Тартла, основанная на правдоподобном выводе и использующая байесианскую сеть вывода.

Байесианская сеть есть направленный ориентированный (без петель) граф, где узлы представляют пропозициональные переменные и связи, или ребра представляют вероятностные отношения между пропозициями. Гипертекстовые системы могут быть сравнены с такой сетью — корни зависимого графа (графа зависимостей) есть гипертекстовые узлы.

Возможны и другие подходы.

Концептуальная сеть. В то время как содержательные узлы образуют документальную сеть, могут быть построены концептуальные узлы, также сеть. Эти понятийные (концептуальные) узлы могут служить индексами к документальной сети. Концептуальная сеть подобна *индексному пространству*, предложенному Фриссом и Кузинсом. Связи внутри концептуальной сети устанавливают соединения между понятиями. Тип связи может описывать природу семантического соединения, а значение, приписываемое ему, может отражать прочность связи. Результат — гипертекстовая БЗ, содержащая концептуальную сеть, вложенную внутрь документальной, может быть использован для анализа запроса.

Классификация гипертекстовых систем

Дж. Конклин (J. Conclin) (автор обзора о гипертекстах в журнале «Computer») классифицировал гипертекстовые системы по типовым применениям на:

- библиотечные макросистемы;
- средства исследования проблем;
- системы для просмотра базы данных (или для «браузинга»);
- системы широкого назначения (для экспериментирования на различных направлениях применения гипертекста).

Библиотечные макросистемы — не просто системы для библиотек или для поиска литературы. Речь идет, скорее, о системах, в которых постоянно добавляются узлы в гипертекстовую сеть; при этом узлы могут представлять собой относительно недолговечные тексты (например, материалы, записи, черновые заметки, используемые в процессе конкретной проектной работы). К таким системам относятся NLS Д. Энгельбарта, Xanadu Т. Нельсона и Textnet Р. Тригга (R. Trigg).

К этому типу можно отнести и систему управления текстами AAIS (Arizona Analyst Information System), которая поддерживает все стадии цикла исследования: выбор и ввод данных, анализ и снятие вопросов, подготовку статьи, составление библиографии. Основным конструктивным элементом в БД AAIS является текстовый «атом», который содержит информацию по какой-то одной или нескольким связанным темам на уровне, необходимом исследователю, определяя тем самым ее полезность, правдивость, своевременность и доказательность. Многочисленные идентификаторы, дифференцированные по определенным категориям (автор, организация, общие ключевые слова), определяют каждый текстовый «атом».

Средства исследования проблем. В гипертекстовой сети легко отображаются, соединяясь связями, идеи, соображения, аргументы разных людей, например участников какой-либо разработки. К гипертекстовым системам исследования проблем принадлежит, в частности, Item-Based Information System (IBIS), разработанная Х. Риттелем (H. Rittel) и предназначенная для подготовки коллективного решения.

В ней три типа узлов — *аспекты* (issue), *позиции* (position), *аргументы* (argument). Узлы создаются участниками обсуждения проблемы. IBIS — это объединение двух систем — телеконференции, которая делает возможным участие многих людей в одном обсуждении, и гипертекста, который позволяет легко переходить от одного ас-

пекта к другому, а также рассматривать разные аргументы относительно одного и того же аспекта.

Система SINVIEW похожа на IBIS, однако предусматривает взаимную оценку участниками обсуждения их позиций и аргументов по критериям значимости и релевантности. В конечном счете, она формирует структуру аргументов вместе с их средними оценками.

К данному типу также относится Writing Environment (WE), разработанная в университете шт. Северной Каролины (США). Дж. Конклин описал ее так: «Сеть собственных идей и внешних источников, имеющая произвольную структуру, представляется сначала в виде иерархической системы, а затем переводится в линейный поток слов, высказываний и т. п.».

Системы просмотра БД. Здесь основным требованием является легкость доступа к большим массивам информации, а добавление новой информации пользователям либо не разрешается, либо специально не поддерживается.

Сюда можно отнести *электронные учебники*, удобные тем, что в зависимости от качества выполнения учащимися упражнений, характера допускаемых ошибок осуществляется переход к соответствующему ранее пройденному материалу. Примером системы электронного учебника является E-book 3, созданная в институте автоматизации и исследования операций (IARO) Фрибургского университета (Швейцария). Сферы применения — математика, экономическая статистика, физика, химия.

Система Symbolics Document Examiner выводит на экран страницы многотомного справочника, позволяет делать закладки в разных местах и затем быстро перемещаться по ним.

Системы широкого назначения. Последний класс включает системы, основным назначением которых является исследование возможностей, открываемых гипертекстовой технологией. Особенность систем этого типа состоит в их гибкости, способности легко модифицироваться.

Например, система NoteCards, созданная в исследовательском центре «Хегох PARS», имеет интерфейс для программистов, позволяющий настраивать ее с учетом новых применений (создавать новые типы узлов, включать в узлы не только графику и изображения, но и «действия» — операторы, выполняемые при активизации данного узла).

Обучающая система Intermedia университета Брауна также ориентирована на экспериментирование, в частности, с типами узлов (например, узлам гипертекстовой сети могут соответствовать не только фрагменты текста, но и изображения, записанная и синте-

зированной речь, музыка). Система используется по двум курсам — биологии клеток и английской литературе. Студенты могут сами формировать различные ракурсы изучаемой проблематики, по-своему организуя учебный материал; добавлять к нему собственные заметки. Преподаватель или другой студент имеют возможность снабжать выполняемую работу комментариями, критическими замечаниями, которые соединяются гипертекстовой связью с соответствующим фрагментом текста.

Система Tektronix Neptune включает *гипертекстовую абстрактную машину*, обеспечивающую большое разнообразие операций с БД. Tektronix имеет различные типы *браузеров* (например, выбирающие узлы с определенными признаками, воспроизводящие случайное движение в гипертексте и т. д.), поддерживает разные версии одного узла и ответвления внутри других.

Guide (Owl Technologies) позволяет интегрировать текст, графику, видеодокументы; устанавливать между объектами до четырех типов связей и подключать к связям пользовательские программы.

На основе Guide в Кентском университете (Великобритания) разработан крупный проект Justus — большая интегральная система по законодательству. Узлы этой БД включают тексты и структурные элементы отчетов о судебных заседаниях. Обеспечивается прямой доступ ко всем материалам по перекрестным ссылкам. Именно в юридической сфере необходимо и важно сопоставление документов, что и является отличительной чертой гипертекстовой технологии.

В свое время фирмой Apple было разработано программное обеспечение HyperCard, предназначенное для создания гипертекста на связанных между собой дискретных карточках. Концепция HyperCard базируется на *гипертексте и объектно-ориентированном программировании*. Центральным объектом системы является «колода», состоящая из нефиксированного набора карточек, по которым пользователь может перемещаться и совершать различные операции (сортировку, поиск, «оживление» статических изображений и т. д.), а также связывать информацию на разных картах с помощью карты-указателя связей. Используемый в системе язык программирования HyperTalk имеет синтаксис, максимально приближенный к естественному английскому языку.

Гипертекстовая система MetaReference, разработанная в среде HyperCard, предназначена для автоматизации семинарской работы студентов. С помощью системы они могут ознакомиться с текущим состоянием обсуждения, имеющимися рефератами, замечаниями по ним, использованными первоисточниками.

В дальнейшем был разработан ряд подобных HyperCard систем: Hyperties или TIES (The Interactive Encyclopedia System, университет шт. Мэриленд, США), Hyperdoc (фирма GECI), HyperPAD (фирма Brightbill-Roberts).

Система Hyperdoc задумана как «ящик с инструментами», к которому можно подключать разнообразные внешние системы обработки текста, электронной графики, электронного издания.

Каждый экран в HyperPAD называется страницей, а массивы, которые их содержат, — блокнотами. Это объектно-ориентированная система; каждая кнопка, поле, фон, страница или блокнот являются объектами. Любой объект имеет атрибуты (цвет, размер, форму) и может иметь программный сценарий, который заставляет атрибут выполнять его задачи.

В заключение перечислим ряд отечественных разработок, среди которых наиболее распространены системы, построенные на так называемых «кнопках». Компьютерная поддержка связей реализуется в этих системах как поддержка переходов между узлами по выделенным в тексте словам. Эти системы могут быть классифицированы следующим образом:

1) кнопочные системы, ориентированные на чтение. Они предназначены для создания электронных книг, справочников, гипержурналов: GTVIEW, Flexis, ГИД, Dutl Help, User Help, HelpMake, Shelp;

2) системы, также построенные на кнопках, но представляющие конечному пользователю механизмы создания новых узлов и связей. Они могут использоваться и для обучающих систем, АИПС, для построения интеллектуального интерфейса пользователя: ГиперМетод, СПИГ, ТИГР, ЭЛБИ;

3) системы, ориентированные на подготовку документов в офисе: ГипСи, ГиперМетод;

4) системы, предназначенные для визуализации структуры гипертекстовой сети — пользователь в любой момент может представить структуру сети и текущее положение в ней: ГИПЕРНЕТ, Мета Дизайн, АСАД, HSS и Browser НМ;

5) гибридные системы, объединяющие гипертекстовую технологию и ЭС. Гипертекст играет подчиненную роль — сеть узлов определяется принятой моделью, и связи между узлами фиксированы: ДИАКОН, КОМОД;

6) системы, в которых реализуется автоматическая навигация. Они применяются для смыслового упорядочения неструктурированной информации и построения связанных текстов, для авторской работы. Это направление начало развиваться под названием «логи-

ко-смысловой метод». Были разработаны критерии смыслового упорядочения текстовых фрагментов, позволяющие формировать из них связный, последовательный и плавный текст: ГИПЕРЛОГ, БАГИС, СЕМПРО.

Каковы же причины успеха гипертекстовой технологии? По мнению Д. Р. Суонсона, существует знание, никем не высказанное и не осознанное, но содержащееся в опубликованных документах и открываемое при сопоставлении этих документов. Поскольку документы обычно относятся к разным темам и даже дисциплинам, открытие такого знания обычно занимает длительное время, происходит неожиданно, требует особых способностей от исследователя и совершается, когда появляется потребность в этом знании.

Обращаясь же к любому узлу гипертекстовой сети, пользователь всегда видит его в окружении каких-либо других узлов, т. е. любой вопрос (тема, проблема, аспект, идея, документ) всегда оказывается связанным с другими вопросами, аспектами и т. п.

Гипертекст не «освобождает» человека от напряжения творческой работы, а дает ему возможность осуществить эту работу с большей интенсивностью; это средство поддержки неформализованной интеллектуальной деятельности.

Контрольные вопросы

1. Назовите основные отличительные черты фактографических АИС.
2. В чем заключается сходство и различие моделей данных для фактографических АИС?
3. Перечислите преимущества и недостатки различных МД.
4. Перечислите основные понятия, связанные с табличными БД.
5. Каковы основные компоненты физической и логической структур БД ADABAS?
6. Назовите основные группы команд и операторов ЯЗ NATURAL и приведите несколько примеров.
7. Из каких компонент состоит поисковый критерий ЯЗ NATURAL?
8. Из каких типов файлов состоит БД в среде систем FoxBase/FoxPro?
9. Назовите основные группы команд и операторов входного языка FoxPro и приведите примеры.
10. Каковы основные отличительные характеристики документальных АИС?
11. Перечислите основные процессы и информационные массивы при навигации в случае «ручного поиска» документальной информации.
12. Что представляют собой «классические схемы» функционирования документальных АИС?

13. Опишите логическую и физическую структуры БД STAIRS.
14. Что собой представляет ЯЗ STAIRS?
15. Из каких компонент и таблиц состоит описание БД STAIRS?
16. В чем состоит процедура загрузки данных в БД STAIRS?
17. Каково взаимодействие физических файлов БД STAIRS при загрузке документов?
18. Опишите структуру навигации в БД STAIRS при простом текстовом поиске.
19. Опишите структуру навигации в БД STAIRS при текстовом поиске использованием контекстных связей.
20. Опишите структуру навигации в БД STAIRS при поиске по форматным полям.
21. Опишите структуру интерфейса ППП IRBIS.
22. Охарактеризуйте основные поисковые возможности IRBIS.
23. В чем состоят особенности информационного поиска в информационных ресурсах Интернета?
24. Опишите структуру поисковых систем Интернета и особенности реализации поисковых процедур.
25. В чем заключается отличие ЭС от других типов АИС? Сравните состав и структуру систем.
26. Какие классы ЭС Вам известны?
27. Какие способы применения ЭС Вы знаете?
28. В чем заключается сущность гипертекстовых систем?
29. Опишите основные этапы развития ГС.
30. Каковы типы навигации в базах данных ГС?
31. Перечислите и охарактеризуйте основные классы гипертекстовых АИС.

Глава 3

ТЕХНОЛОГИИ ПРОЕКТИРОВАНИЯ АИС

Современные информационные технологии предоставляют широкий набор способов реализации ИС, выбор которых осуществляется на основе требований со стороны предполагаемых пользователей, которые, как правило, изменяются в процессе разработки.

Под *проектом ИС* будем понимать проектно-конструкторскую и технологическую документацию, в которой представлено описание проектных решений по созданию и эксплуатации ИС в конкретной программно-технической среде [1, 38].

Под *проектированием ИС* понимается процесс преобразования входной информации об объекте, методах и опыте проектирования объектов аналогичного назначения в соответствии с ГОСТом в проект ИС. С этой точки зрения проектирование ИС сводится к последовательной формализации проектных решений на различных стадиях жизненного цикла ИС: планирования и анализа требований, технического и рабочего проектирования, внедрения и эксплуатации ИС [2].

Объектами проектирования ИС являются отдельные элементы или комплексы функциональных и обеспечивающих частей. Так, функциональными элементами в соответствии с традиционной декомпозицией выступают задачи, комплексы задач и функции управления. В составе обеспечивающей части ИС объектами проектирования служат элементы и комплексы информационного, программного, технического и других видов обеспечения системы [1, 38].

В качестве *субъекта проектирования ИС* выступают коллективы специалистов, которые осуществляют проектную деятельность, как правило, в составе *специализированной (проектной) организации*, и *организация-заказчик*, для которой необходимо разработать ИС. Масштабы разрабатываемых систем определяют состав и количество участников процесса проектирования. При большом объеме и жестких сроках выполнения проектных работ в разработке системы может принимать участие несколько проектных коллективов (организаций-разработчиков). В этом случае выделяется головная орга-

низация, которая координирует деятельность всех организаций-соисполнителей.

Форма участия соисполнителей в разработке проекта системы может быть различной. Наиболее распространенной является форма, при которой каждый соисполнитель выполняет проектные работы от начала до конца для какой-либо части разрабатываемой системы. Обычно это бывает функциональная подсистема или взаимосвязанный комплекс задач управления. Реже встречается форма, при которой некоторые соисполнители выполняют работы на отдельных этапах процесса проектирования. Возможен вариант, при котором функции заказчика и разработчика совмещаются, т. е. АИС проектируется собственными силами.

Осуществление проектирования ИС предполагает использование проектировщиками определенной технологии проектирования, соответствующей масштабу и особенностям разрабатываемого проекта.

Технология проектирования ИС — это совокупность методологии и средств проектирования ИС, а также методов и средств его организации (управление процессом создания и модернизации проекта ИС).

В основе технологии проектирования лежит технологический процесс, который определяет действия, их последовательность, требуемые состав исполнителей, средства и ресурсы.

Технологический процесс проектирования ИС в целом делится на совокупность последовательно-параллельных, связанных и соподчиненных цепочек действий, каждое из которых может иметь свой предмет. Действия, которые выполняются при проектировании ИС, могут быть определены как неделимые технологические операции или как подпроцессы технологических операций. Все действия делятся на собственно проектировочные, которые формируют или модифицируют результаты проектирования, и оценочные действия, которые вырабатывают по установленным критериям оценку результатов проектирования.

Таким образом, технология проектирования задается регламентированной последовательностью технологических операций, выполняемых на основе того или иного метода, в результате чего становится ясным, не только *что* должно быть сделано для создания проекта, но и *как, кем и в какой последовательности*.

Предметом любой выбираемой технологии проектирования должно служить отражение взаимосвязанных процессов проектирования на всех стадиях жизненного цикла ИС. К основным требова-

ниям, предъявляемым к выбираемой технологии проектирования, относятся следующие:

- созданный проект должен отвечать требованиям заказчика;
- максимальное отражение всех этапов жизненного цикла проекта;
- обеспечение минимальных трудовых и стоимостных затрат на проектирование и сопровождение проекта;
- технология должна быть основой связи между проектированием и сопровождением проекта;
- рост производительности труда проектировщика;
- надежность процесса проектирования и эксплуатации проекта;
- простое ведение проектной документации.

Основу технологии проектирования ИС составляет методология, которая определяет сущность, основные отличительные технологические особенности.

Методология проектирования предполагает наличие некоторой концепции, принципов проектирования, реализуемых набором методов, которые, в свою очередь, должны поддерживаться некоторыми средствами [1, 38].

Организация проектирования предполагает определение методов взаимодействия проектировщиков между собой и с заказчиком в процессе создания проекта ИС, которые могут также поддерживаться набором специфических средств.

3.1. Классификация методов проектирования систем

Методы проектирования ИС можно классифицировать по степени использования средств автоматизации, типовых проектных решений, адаптивности к предполагаемым изменениям.

Так, по *степени автоматизации* методы проектирования разделяются на:

- ручное, при котором проектирование компонентов ИС осуществляется без использования специальных инструментальных программных средств, а программирование — на алгоритмических языках;
- компьютерное, при котором производится генерация или конфигурирование (настройка) проектных решений на основе использования специальных инструментальных программных средств.

По степени использования типовых проектных решений различают следующие методы проектирования:

- оригинальное (индивидуальное), когда проектные решения разрабатываются «с нуля» в соответствии с требованиями к АИС. Характеризуется тем, что все виды проектных работ ориентированы на создание индивидуальных для каждого объекта проектов, которые в максимальной степени отражают все его особенности;
- типовое, предполагающее конфигурирование ИС из готовых типовых проектных решений (программных модулей). Выполняется на основе опыта, полученного при разработке индивидуальных проектов. Типовые проекты, как обобщение опыта для некоторых групп организационно-экономических систем или видов работ, в каждом конкретном случае связаны со множеством специфических особенностей и различаются по степени охвата функций управления, выполняемым работам и разрабатываемой проектной документации.

По степени адаптивности проектных решений выделяют методы:

- реконструкции, когда адаптация проектных решений выполняется путем переработки соответствующих компонентов (перепрограммирования программных модулей);
- параметризации, когда проектные решения настраиваются (генерируются) в соответствии с изменяемыми параметрами;
- реструктуризации модели, когда изменяется модель проблемной области, на основе которой автоматически заново генерируются проектные решения.

Сочетание различных признаков классификации методов обуславливает характер используемых технологий проектирования ИС, среди которых выделяют два основных класса: *каноническую* и *индустриальную технологии* (табл. 3.1) [1, 38]. Индустриальная технология проектирования, в свою очередь, разбивается на два подкласса: автоматизированное (использование CASE-технологий) и типовое (параметрически-ориентированное или модельно-ориентированное) проектирование. Использование индустриальных технологий не исключает использования в отдельных случаях канонических.

Для конкретных видов технологий проектирования свойственно применение определенных средств разработки ИС, которые поддерживают выполнение как отдельных проектных работ, этапов, так и их совокупностей. Поэтому перед разработчиками ИС, как правило, стоит задача выбора средств проектирования, которые по своим ха-

Таблица 3.1. Характеристики классов технологий проектирования

Класс технологии	Степень автоматизации	Степень типизации	Степень адаптивности
Каноническое	Ручное	Оригинальное	Реконструкция
Индустриальное автоматизированное	Компьютерное	Оригинальное	Реструктуризация модели (генерация ИС)
Индустриальное типовое	Компьютерное	Типовое сборочное	Параметризация и реструктуризация модели (конфигурация ИС)

характеристикам в наибольшей степени соответствуют требованиям конкретного предприятия.

Средства проектирования должны быть:

- инвариантны к объекту проектирования (в своем классе);
- охватывать в совокупности все этапы жизненного цикла ИС;
- технически, программно и информационно совместимыми;
- простыми в освоении и применении;
- экономически целесообразными.

Средства проектирования ИС можно разделить на два класса: *без использования ЭВМ и с использованием ЭВМ.*

Средства проектирования без использования ЭВМ применяются на всех стадиях и этапах. Как правило, это средства организационно-методического обеспечения операций и в первую очередь различные стандарты, регламентирующие процесс проектирования систем. Сюда же относятся единая система классификации и кодирования информации, унифицированная система документации, модели описания и анализа потоков информации и т. п.

Средства проектирования с использованием ЭВМ могут применяться как на отдельных, так и на всех стадиях и этапах процесса проектирования ИС и соответственно поддерживают разработку элементов, разделов, проекта системы в целом. Все множество средств проектирования с использованием ЭВМ делят на четыре подкласса.

1. Операционные средства, которые поддерживают проектирование операций обработки информации. К данному подклассу средств относятся алгоритмические языки, библиотеки стандартных подпрограмм и классов объектов, макрогенераторы, генераторы программ типовых операций обработки данных и т. п., а также средства расширения функций операционных систем (утилиты). В данный класс включаются также такие простейшие инструмен-

тальные средства проектирования, как средства для тестирования и отладки программ, поддержки процесса документирования проекта и т. п. Особенность последних программ заключается в том, что с их помощью повышается производительность труда проектировщиков, но не разрабатывается законченное проектное решение.

Таким образом, средства данного подкласса поддерживают отдельные операции проектирования ИС и могут применяться независимо друг от друга.

2. Средства, поддерживающие проектирование отдельных компонентов. К данному подклассу относятся средства общесистемного назначения:

- системы управления базами данных (СУБД);
- методо-ориентированные пакеты прикладных программ (решение задач дискретного программирования, математической статистики и т. п.);
- табличные процессоры;
- статистические ППП;
- оболочки экспертных систем;
- графические редакторы;
- текстовые редакторы;
- интегрированные ППП (интерактивная среда с встроенными диалоговыми возможностями, позволяющая интегрировать вышеперечисленные программные средства).

Для перечисленных средств характерно их использование для разработки технологических подсистем ИС: ввода информации, организации хранения и доступа к данным, вычислений, анализа и отображения данных, принятия решений.

3. Средства, поддерживающие проектирование разделов проекта. В этом подклассе выделяют функциональные средства проектирования.

Функциональные средства направлены на разработку автоматизированных систем, реализующих функции, комплексы задач и задачи управления. Разнообразие предметных областей порождает многообразие средств данного подкласса, ориентированных на тип организационной системы (промышленная, непромышленная сферы), уровень управления (например, предприятие, цех, отдел, участок, рабочее место), функцию управления (планирование, учет и т. п.).

К функциональным средствам проектирования систем обработки информации относятся типовые проектные решения, функциональные пакеты прикладных программ, типовые проекты.

4. Средства, поддерживающие разработку на стадиях и этапах процесса проектирования. К данному классу относятся средства автоматизации проектирования ИС (CASE-средства). Современные CASE-средства, в свою очередь, классифицируются в основном по двум признакам:

- 1) по охватываемым этапам процесса разработки ИС;
- 2) по степени интегрированности:
 - *отдельные локальные средства (tools)*;
 - *набор неинтегрированных средств*, охватывающих большинство этапов разработки ИС (toolkit);
 - *полностью интегрированные средства*, связанные общей базой проектных данных — репозиторием (workbench).

3.2. Жизненный цикл ИС

Потребность в создании ИС может обуславливаться либо необходимостью автоматизации или модернизации существующих информационных процессов, либо необходимостью коренной реорганизации в деятельности предприятия (проведении бизнес-реинжиниринга — РБП). Потребности создания ИС указывают, во-первых, для достижения каких именно целей необходимо разработать систему; во-вторых, к какому моменту времени целесообразно осуществить разработку; в-третьих, какие затраты необходимы для осуществления проектирования системы.

Проектирование ИС — трудоемкий, длительный и динамический процесс. Технологии проектирования, применяемые в настоящее время, предполагают поэтапную разработку системы. Этапы по общности целей могут объединяться в стадии. Совокупность стадий и этапов, которые проходит ИС в своем развитии от момента принятия решения о создании системы до момента прекращения ее функционирования, называется *жизненным циклом ИС*.

Суть жизненного цикла разработки ИС в различных подходах одинакова и сводится к выполнению следующих стадий [1, 38]:

1. *Планирование и анализ требований* (предпроектная стадия) — системный анализ. Исследование и анализ существующей информационной системы, определение требований к создаваемой ИС, оформление технико-экономического обоснования (ТЭО) и технического задания (ТЗ) на разработку ИС.

2. *Проектирование* (техническое проектирование, логическое проектирование). Разработка в соответствии со сформулированными требованиями состава автоматизируемых функций (функцио-

нальная архитектура) и состава обеспечивающих подсистем (системная архитектура), оформление технического проекта ИС.

Часто второй и третий этапы объединяют в одну стадию, называемую *техно-рабочим проектированием* или *системным синтезом*.

3. *Реализация* (рабочее проектирование, физическое проектирование, программирование). Разработка и настройка программ, наполнение баз данных, создание рабочих инструкций для персонала, оформление рабочего проекта.

4. *Внедрение* (тестирование, опытная эксплуатация). Комплексная отладка подсистем ИС, обучение персонала, поэтапное внедрение ИС в эксплуатацию по подразделениям экономического объекта, оформление акта о приемо-сдаточных испытаниях ИС.

5. *Эксплуатация ИС* (сопровождение, модернизация). Сбор рекламаций и статистики о функционировании ИС, исправление ошибок и недоработок, оформление требований к модернизации ИС и ее выполнение (повторение стадий 2—5).

Системный анализ

К основным целям процесса относятся следующие:

- формулировка потребности в новой ИС (идентифицировать все недостатки существующей ИС);
- выбор направления и определение экономической целесообразности проектирования ИС.

Системный анализ ИС начинается с описания и анализа функционирования рассматриваемого экономического объекта (системы) в соответствии с требованиями (целями), которые предъявляются к нему. В результате этого этапа выявляются основные недостатки существующей ИС, на основе которых формулируется потребность в совершенствовании системы управления этим объектом, и ставится задача определения экономически обоснованной необходимости автоматизации определенных функций управления, т. е. создается технико-экономическое обоснование проекта. После определения этой потребности возникает проблема выбора направлений совершенствования объекта на основе выбора программно-технических средств. Результаты оформляются в виде технического задания на проект, в котором отражаются технические условия и требования к ИС, а также ограничения на ресурсы проектирования. Требования к ИС определяются в терминах функций, реализуемых системой, и предоставляемой ею информацией.

Системный синтез

Этот процесс включает:

- разработку функциональной архитектуры ИС, которая отражает структуру выполняемых функций;
- разработку системной архитектуры выбранного варианта ИС, т. е. состав обеспечивающих подсистем;
- выполнение реализации проекта.

Этап составления функциональной архитектуры, представляющей собой совокупность функциональных подсистем и связей между ними, является наиболее ответственным с точки зрения качества всей последующей разработки.

Построение системной архитектуры на основе функциональной предполагает выделение элементов и модулей информационного, технического, программного обеспечения и других обеспечивающих подсистем, определение связей по информации и управлению между выделенными элементами и разработку технологии обработки информации.

Этап конструирования (физического проектирования системы) включает разработку инструкций пользователям и программ, создание информационного обеспечения, включая наполнение баз данных.

Внедрение разработанного проекта

Процесс предполагает выполнение этапов *опытного внедрения и промышленного внедрения*.

Этап опытного внедрения заключается в проверке работоспособности элементов и модулей проекта, устранении ошибок на уровне элементов и связей между ними и развертывании элементов системы на рабочих местах пользователя.

Этап сдачи в промышленную эксплуатацию заключается в организации проверки проекта на уровне функций и контроля соответствия требованиям, сформулированным на стадии системного анализа.

Эксплуатация и сопровождение проекта

На этой стадии выполняются этапы *эксплуатации и модернизации* проекта ИС.

Рассмотренная схема жизненного цикла ИС условно включает в свой состав только основные процессы, однако их реальный набор

и разбиение на этапы и технологические операции в значительной степени зависят от выбираемой технологии проектирования.

Важной чертой жизненного цикла ИС является его повторяемость «системный анализ — разработка — сопровождение — системный анализ». Это соответствует представлению об ИС как о развивающейся, динамической системе. При первом выполнении стадии «Разработка» создается проект ИС, а при повторном выполнении осуществляется модификация проекта для поддержания в актуальном состоянии.

Другой характерной чертой жизненного цикла является наличие ряда циклов внутри схемы:

- *цикл первичного проектирования ИС;*
- *цикл опытного внедрения* — выясняются частные ошибки в элементах проекта;
- *после сдачи в промышленную эксплуатацию* выявляют ошибки в функциональной архитектуре системы, связанные с несоответствием проекта требованиям заказчика, по составу функциональных подсистем, составу задач и связям между ними;
- цикл возникает в случае, когда требуется *модификация системной архитектуры* в связи с необходимостью адаптации проекта к новым условиям функционирования системы;
- цикл появляется, если *проект системы совершенно не соответствует требованиям*, предъявляемым к организационно-экономической системе ввиду того, что осуществляется моральное старение и требуется полное перепроектирование системы.

Чтобы исключить пятый цикл и максимально уменьшить необходимость прохождения третьего и четвертого циклов, проектирование ИС на всех этапах первого, основного цикла разработки ИС необходимо выполнять в соответствии со следующими требованиями:

- разработка ИС должна быть выполнена в строгом соответствии со сформулированными требованиями к создаваемой системе;
- требования к ИС должны адекватно соответствовать целям и задачам эффективного функционирования экономического объекта;
- созданная ИС должна соответствовать сформулированным требованиям на момент окончания внедрения, а не начала разработки;
- внедренная ИС должна развиваться и адаптироваться в соответствии с постоянно изменяющимися требованиями к ИС.

С точки зрения реализации перечисленных аспектов в технологиях проектирования ИС модели жизненного цикла, определяющие порядок выполнения стадий и этапов, претерпевали существенные изменения.

Модели жизненного цикла

Среди известных моделей жизненного цикла можно выделить следующие:

- *каскадная модель* (до 70-х гг.) — последовательный переход на следующий этап после завершения предыдущего;
- *итерационная модель* (70—80-е гг.) — итерационные возвраты на предыдущие этапы после выполнения очередного;
- *спиральная модель* (80—90-е гг.) — прототипная модель, предполагающая постепенное расширение прототипа ИС.

Каскадная модель. Для этой модели жизненного цикла характерна автоматизация отдельных несвязанных задач, не требующая выполнения информационной интеграции и совместимости, программного, технического и организационного сопряжения. В рамках решения отдельных задач каскадная модель жизненного цикла по срокам разработки и надежности оправдывала себя. Применение каскадной модели жизненного цикла к большим и сложным проектам вследствие большой длительности процесса проектирования и изменчивости требований за это время может привести к практической нереализуемости их.

Итерационная модель. Создание комплексных ИС предполагает проведение увязки проектных решений, получаемых при реализации отдельных задач. Подход к проектированию *снизу-вверх* обуславливает необходимость таких итерационных возвратов, когда проектные решения по отдельным задачам комплектуются в общие системные решения и при этом возникает потребность в пересмотре ранее сформулированных требований. Как правило, вследствие большого числа итераций возникают рассогласования в выполненных проектных решениях и документации. Запутанность функциональной и системной архитектуры созданной ИС, трудность в использовании проектной документации вызывают на стадиях внедрения и эксплуатации сразу необходимость перепроектирования всей системы. Длительный жизненный цикл разработки ИС заканчивается этапом внедрения, за которым начинается жизненный цикл создания новой ИС.

Спиральная модель. Используется подход к организации проектирования ИС *сверху-вниз*, когда сначала определяется состав функциональных подсистем, а затем постановка отдельных задач. Соответственно сначала разрабатываются такие общесистемные вопросы, как организация интегрированной базы данных, технология сбора, передачи и накопления информации, а затем технология решения конкретных задач. В рамках комплексов задач программирование осуществляется по направлению от головных программных модулей к исполняющим отдельные функции. При этом на первый план выходят вопросы взаимодействия интерфейсов программных модулей между собой и с базой данных, а на второй — реализация алгоритмов.

В основе спиральной модели жизненного цикла лежит применение прототипной технологии или RAD-технологии (Rapid Application Development — технологии быстрой разработки приложений — см. J. Martin. Rapid Application Development. New York: Macmillan, 1991). Согласно этой технологии ИС разрабатывается путем расширения программных прототипов, повторяя путь от детализации требований к детализации программного кода. Естественно, что при прототипной технологии сокращается число итераций и возникает меньше ошибок и несоответствий, которые необходимо исправлять на последующих итерациях, при этом проектирование ИС осуществляется более быстрыми темпами, упрощается создание проектной документации. Для более точного соответствия проектной документации разработанной ИС все большее значение придается ведению общесистемного репозитория и использованию CASE-технологий.

Жизненный цикл при использовании RAD-технологии предполагает активное участие конечных пользователей будущей системы на всех этапах разработки и включает четыре основные стадии информационного инжиниринга:

- *анализ и планирование информационной стратегии.* Пользователи вместе со специалистами-разработчиками участвуют в идентификации проблемной области;
- *проектирование.* Пользователи принимают участие в техническом проектировании под руководством специалистов-разработчиков;
- *конструирование.* Специалисты-разработчики проектируют рабочую версию ИС с использованием языков четвертого поколения;
- *внедрение.* Специалисты-разработчики обучают пользователей работе в среде новой ИС.

Формализация технологии проектирования ИС

Сложность, высокие затраты и трудоемкость процесса проектирования ИС на протяжении всего жизненного цикла вызывают необходимость, с одной стороны, выбора адекватной экономическому объекту технологии проектирования, с другой — наличия эффективного инструмента управления процессом ее применения. С этой точки зрения возникает потребность в построении такой формализованной модели технологии проектирования, на основе которой можно было бы оценить необходимость и возможность применения определенной технологии проектирования с учетом сформулированных требований к ИС и выделенных на экономическом объекте ресурсов, а в последующем контролировать ход и результаты проектирования.

Известные методы сетевого планирования и управления проектами решают только одну часть поставленной проблемы: отражают последовательность технологических операций с временными и трудовыми характеристиками. При этом не раскрывается в полной мере содержательная сторона процесса проектирования, необходимая сначала для понимания сущности и оценки эффективности технологии проектирования, а затем для использования в качестве инструкционного материала в непосредственной работе проектировщиков.

В наибольшей степени задаче формализации технологии проектирования ИС соответствует аппарат технологических сетей проектирования.

Технологическая сеть проектирования (ТСП) строится на основе отдельных технологических операций. Под ТСП понимается взаимосвязанная по входам и выходам последовательность технологических операций проектирования, выполнение которых приводит к достижению требуемого результата — созданию проекта ИС.

Технологические сети проектирования могут строиться с различной степенью детализации. Наиболее детализированная ТСП, в которой каждая технологическая операция является ручной, называется *канонической*. Каноническая ТСП наиболее пригодна для проектировщиков-исполнителей, так как является руководством по проектированию ИС. Вместе с тем каноническая ТСП всего проекта редко используется в полном объеме, скорее различные категории проектировщиков-исполнителей пользуются относящимися к их компетенции фрагментами канонической сети.

Для укрупнения ТСП применяются *технологические операции-агрегаты*, которым соответствуют фрагменты канонической

ТСП. Например, ТО «Проектирование схемы базы данных» декомпозируется на ряд взаимосвязанных ТО: «Нормализация таблиц», «Установление связей», «Отображение в схеме DDL СУБД» и т. д.

Для различных категорий разработчиков проекта ИС требуется различная степень агрегации-детализации ТСП. Наименее детализованная ТСП нужна заказчикам, для которых она представляет набор взаимосвязанных технологических этапов со входами, соответствующими предоставляемой разработчикам информации, и выходами, соответствующими получаемым проектным документам. Для руководителей проектов технологические операции, как правило, соответствуют календарным работам с четкими сроками сдачи и документальными результатами. В принципе для таких категорий пользователей ТСП может быть преобразована в традиционный сетевой график. На этом уровне представления ТСП могут не указываться отдельные ресурсы или средства проектирования.

Для взаимодействующих проектировщиков-исполнителей очень важно отражение в ТСП связей по входу-выходу, поскольку для качественного выполнения любой технологической операции необходимо точное выполнение требований по входу, соответствующему выходу другой ТО. Для конкретного проектировщика-исполнителя относящаяся к его компетенции технологическая операция-агрегат всегда может быть раскрыта в виде фрагмента канонической сети.

При использовании средства автоматизированного проектирования проектировщик-исполнитель может пользоваться технологическими операциями-агрегатами, объединяющими фрагменты канонической ТСП. Для таких ТО обязательно задается ссылка на используемое средство проектирования. Причем если средство проектирования является комплексным, то указываются конкретный компонент (функция, модуль, опция и т. д.) или компоненты этого средства.

Технологические сети проектирования могут иметь вариантный характер построения. Например, ТСП проектирования выходных форм отчетов зависит от средства проектирования, выбор которого, в свою очередь, определяется их сложностью. Для правильного выбора средства проектирования вводится специальная технологическая операция, которая сопоставляет параметры требований (например, итоги отчетов, число таблиц формы, число файлов базы данных и др.) с аналогичными параметрами средства проектирования. В зависимости от выбранного средства проектирования далее определяется конкретная ветка ТСП. Например, если в совокупности средств проектирования есть только генератор отчетов, рабо-

тающий с одним файлом, то в технологическую сеть потребуется ввести операцию проектирования выходного файла. Если ни одно из средств не подходит, то проектирование осуществляется в соответствии с канонической сетью проектирования.

3.3. Содержание и методы канонического проектирования ИС

Каноническое проектирование ИС отражает особенности ручной технологии индивидуального (оригинального) проектирования, осуществляемого на уровне исполнителей без использования каких-либо инструментальных средств, позволяющих интегрировать выполнение элементарных операций. Как правило, каноническое проектирование применяется для небольших локальных ИС [1, 38].

Стадии и этапы канонического проектирования ИС

В основе канонического проектирования лежит *каскадная модель* жизненного цикла ИС.

Процесс каскадного проектирования в жизненном цикле ИС в соответствии с применяемым в нашей стране ГОСТ 34601—90 «Автоматизированные системы. Стадии создания» делится на следующие семь стадий:

- 1) исследование и обоснование создания системы;
- 2) разработка технического задания;
- 3) создание эскизного проекта;
- 4) техническое проектирование;
- 5) рабочее проектирование;
- 6) ввод в действие;
- 7) функционирование, сопровождение, модернизация.

В целях изучения взаимосвязанных приемов и методов канонического проектирования ИС перечисленные семь стадий можно сгруппировать в часто используемые на практике четыре стадии процесса разработки ИС (табл. 3.2).

Состав и содержание работ на предпроектной стадии создания ИС

При исследовании существующего экономического объекта (системы) разработчики должны уточнить границы изучения, определить круг пользователей будущей ИС различных уровней и выде-

Таблица 3.2. Содержание и результаты основных стадий канонического проектирования АИС

№	Стадия	Этап стадии	Результаты	Документы
1	Предпроектная стадия (предпроектное обследование)	Сбор материалов обследования	Материалы обследования, которые должны содержать полную и достоверную информацию, описывающую изучаемую предметную область	—
		Анализ материалов обследования и разработка технико-экономического обоснования (ТЭО) и технического задания (ТЗ)	Количественные и качественные характеристики информационных потоков, описание их структуры и мест обработки, объемов выполняемых операций и трудоемкости их обработки	«Технико-экономическое обоснование проектных решений» (ТЭО) «Техническое задание» (ТЗ); для сложных ИС иногда включают разработку «Эскизного проекта» [1]
2	Технорабочее проектирование	Техническое проектирование	Работы по логической разработке и выбору наилучших вариантов проектных решений	«Технический проект» [1]
		Рабочее проектирование	Физическая реализация выбранного варианта проекта	«Рабочий проект» (иногда – «Технорабочий проект» (ТРП)) [1]
3	Внедрение проекта	Подготовка объекта к внедрению проекта	Работы по подготовке предприятия к внедрению разработанного проекта ИС	
		Опытное внедрение проекта	Проверка правильности работы некоторых частей проекта и исправление проектной документации	«Акт о проведении опытного внедрения» [38]
		Сдача проекта в промышленную эксплуатацию	Комплексная системная проверка всех частей проекта	Доработанный «Технорабочий проект» и «Акт приемки проекта в промышленную эксплуатацию» [38]
4	Эксплуатация и сопровождение проекта	Эксплуатация проекта	Сбор информации о работе системы в целом и отдельных компонент и статистики о сбоях системы в виде рекламаций и замечаний	
		Сопровождение и модернизация проекта	Ликвидация последствий сбоев в работе системы, исправление ошибок, невыявленных при внедрении проекта, также модернизация проекта	Модернизированный «Технорабочий проект» [38]

лить классы и типы объектов, подлежащих последующей автоматизации.

Важнейшими объектами обследования могут являться:

- *структурно-организационные звенья* предприятия (например, отделы управления, цехи, участки, рабочие места);
- *функциональная структура*, состав хозяйственных процессов и процедур;
- *стадии* (техническая подготовка, снабжение, производство, сбыт) и *элементы хозяйственного процесса* (средства и предметы труда, ресурсы, продукция, финансы).

При каноническом проектировании основной единицей обработки данных является *задача*. Поэтому функциональная структура проблемной области на стадии предпроектного обследования изучается в разрезе решаемых задач и их комплексов. При этом задача в содержательном аспекте рассматривается как совокупность операций преобразования некоторого набора исходных данных для получения информации, необходимой для выполнения функции управления или принятия управленческого решения. В большинстве случаев исходные данные и результаты их преобразований представляются в форме экономических документов. Поэтому к числу объектов обследования относятся компоненты потоков информации (документы, показатели, файлы, сообщения);

- *технологии, методы* и технические средства преобразования информации;
- *материальные потоки* и процессы их обработки.

Основной целью выполнения первого этапа предпроектного обследования «Сбор материалов» является:

- выявление основных параметров предметной области (например, предприятия или его части);
- установление условий, в которых будет функционировать проект ИС;
- выявление стоимостных и временных ограничений на процесс проектирования.

На этом этапе проектировщиками выполняется ряд технологических операций и решаются следующие задачи:

- предварительное изучение предметной области;
- выбор технологии проектирования; выбор метода проведения обследования;
- выбор метода сбора материалов;
- разработка программы обследования;
- разработка плана-графика сбора материалов обследования;
- сбор и формализация материалов.

Выполнение операции «Предварительное изучение предметной области» имеет своей целью на основе общих сведений об объекте выявить предварительные размеры объемов работ по проектированию и состав стоимостных и временных ограничений на процессы проектирования, а также найти примеры разработок проектов ИС для аналогичных систем.

Важной операцией, определяющей все последующие работы по обследованию объекта и проектированию ИС, является «Выбор технологий проектирования».

Методы проведения обследования

Перед началом работ по проведению обследования необходимо *выбрать метод проведения обследования*. Все методы можно объединить в группы по следующим признакам:

- по цели обследования выделяют метод организации локального проведения обследования, используемый для разработки проекта отдельной задачи или для комплекса задач, и метод системного обследования объекта, применяемый для изучения всего объекта с целью разработки для него проекта ИС в целом;
- по числу исполнителей, проводящих обследование, применяется индивидуальное обследование, осуществляемое одним проектировщиком, и бригадное с выделением ряда бригад — исполнителей, изучающих все подразделения предприятия, и одной координирующей бригады;
- по степени охвата предметной области применяют метод сплошного обследования, охватывающего все подразделения экономической системы, и выборочное, применяемое при наличии типовых по структуре подразделений (например, цехов или складов);
- по степени одновременности выполнения работ первого и второго этапов предпроектной стадии выделяют метод последовательного проведения работ, при котором проектировщики сначала собирают данные о предметной области, а затем их изучают (часто применяют при отсутствии опыта в выполнении такого рода работ), и метод параллельного выполнения работ, когда одновременно со сбором происходит изучение полученных материалов обследования, что значительно сокращает время на проведение предпроектной стадии и повышает качество получаемых результатов.

Выполнение работ по обследованию предметной области в каком-либо подразделении и сбору материалов можно проводить на основе предварительного *выбора методов*, совокупность которых можно разделить на две группы:

- *методы сбора, выполняемого силами проектировщиков-исполнителей*, включающие методы проведения бесед и опросов, анализа материалов обследования, личных наблюдений, фотографии рабочего дня и хронометража рабочего времени специалиста при выполнении им той или иной работы;
- *методы сбора, выполняемого силами специалистов* предметной области, которым предлагается либо заполнять тетрадь-дневник на осуществляемые работы, либо проводить документную инвентаризацию рабочего места, либо использовать метод самофотографии рабочего дня, позволяющий выявить состав операций и получаемые при этом документы;
- *метод бесед и консультаций с руководителями*, который чаще всего проводится в форме обычной беседы с руководителями предприятий и подразделений или в форме деловой консультации со специалистами по вопросам, имеющим глобальный характер и относящимся к определению проблем и стратегий развития и управления предприятием;
- *метод опроса исполнителей на рабочих местах*, который используется в процессе сбора сведений непосредственно у специалистов. Заранее составляют список сотрудников, с которыми намереваются беседовать, разрабатывают перечень вопросов о роли и назначении работ в деятельности объекта, порядке их выполнения;
- *метод анализа операций*, который заключается в расчленении рассматриваемого делового процесса, работы на ее составные части, задачи, расчеты, операции и даже их элементы. После этого анализируется каждая часть в отдельности, выявляются повторяемость отдельных операций, многократное обращение к одной и той же операции, их степень зависимости друг от друга;
- *расчетный метод*, который применяется для определения трудоемкости и стоимости работ, подлежащих переводу на выполнение с помощью ЭВМ, а также для установления объемов работ по отдельным операциям.

При выборе метода следует учитывать следующие критерии:

- степень личного участия проектировщика в сборе материала;
- временные, трудовые и стоимостные затраты на получение сведений в подразделениях.

Проектировщику необходимо знать и в каждом конкретном случае применять наиболее экономичный, обеспечивающий нужную полноту сведений метод сбора материалов обследования.

Обследование проводится по заранее разработанной *программе*, составленной во время выполнения операции по форме, содержащей перечень вопросов, ответы на которые дадут полное представление о деятельности изучаемого объекта и будут учтены при создании проекта ИС. Вопросы можно систематизировать по трем основным направлениям исследования объекта.

1. Получение представления об объекте изучения (например, предприятии) в целом, включая выяснение целей функционирования, значений основных параметров деятельности и т. д.

2. Изучение и описание организационно-функциональной структуры объекта (как правило, относится к аппарату управления). При этом исследуются функции, выполняемые в структурных подразделениях, хозяйственные процессы и процедуры, выявляются комплексы задач, обусловленные функциями, процессами и процедурами, определяется состав входной и выходной информации по каждой задаче.

3. Изучение и описание структуры информационных и/или материальных потоков: состава и структуры компонентов потоков, частоты их возникновения, объемов за определенный период, направления движения, процедур обработки, в которых участвуют эти компоненты. Источником сведений являются интервью со специалистами предметной области, экономическая документация и расчеты. Описание информационной структуры выполняется на уровне экономических документов и показателей.

Для организации труда проектировщиков во время сбора материалов обследования и его последующего анализа необходимо выполнение разработки «Плана-графика выполнения работ на предпроектной стадии».

«План-график» служит инструментом для планирования и оперативного управления предпроектной стадией.

Последней операцией, выполняемой проектировщиками на этом этапе, является «Проведение сбора и формализации материалов обследования», в процессе которой члены бригад должны:

- проинтервьюировать специалистов подразделений изучаемой предметной области; собрать сведения обо всех объектах обследования, в том числе о предприятиях в целом, функциях управления, методах и алгоритмах реализации функций, составе обрабатываемых и рассчитываемых показателей;

- выявить формы документов, отражающих хозяйственные процессы и используемые классификаторы, макеты файлов, сведения об используемых технических средствах и технологиях обработки данных; проконтролировать вместе с пользователем их правильность;
- сформировать «Отчет об обследовании» и выполнить другие работы.

Отчет об обследовании объекта

Сбор материалов обследования следует проводить с помощью стандартных форм и таблиц, которые удобно читать и обрабатывать.

Все получаемые документы охватывают три группы:

1. Содержащие описание общих параметров объекта.
2. Формализующих материалы обследования по каждому структурному подразделению.
3. Содержащие описание компонентов каждого информационного потока, включая документы, информационные файлы, процедуры обработки и характеристики этих компонентов.

Общие параметры описывают организационные структуры, распределение функций, реализуемых каждым структурным подразделением. В частности, общие параметры должны содержать:

- наименование объекта и его принадлежность (министерству, объединению, корпорации и т. п.);
- класс объекта (тип предприятия, вид производства, режимы работы);
- виды и номенклатуру продукции или услуг;
- виды и количество оборудования и материальных ресурсов;
- категории и численность работающих и т. д.

В эту группу входит также форма описания общих характеристик функций управления экономической системой, хозяйственных процессов и процедур, их реализующих. Эта форма предполагает отражение следующих параметров:

- наименование каждой функции, процесса и процедуры, описание экономической сущности задач, решаемых при выполнении процедуры, обработки информации;
- состав процедур обработки информации по каждой задаче;
- взаимосвязь задач, стоимостные затраты, связанные с их реализацией.

Описание организационной структуры должно включать состав и взаимосвязь подразделений и лиц, реализующих функции и зада-

чи управления. Описание производственной структуры объекта должно отражать состав и взаимосвязь подразделений, реализующих производство товаров или услуг. Описание функциональной структуры призвано отображать распределение функций, хозяйственных процессов и процедур управления между составляющими организационной структуры и должно предполагать проведение классификации процедур, связанных с обработкой данных, коммуникацией между сотрудниками или принятием управленческих решений.

Описание материальных потоков предполагает отображение маршрутов движения средств, предметов и продуктов труда, рабочей силы между подразделениями производственной структуры и будет включать: описание видов продукции или услуг, ресурсов; описание технологических операций, их частоту и длительность выполнения; объемы перемещаемых ресурсов, продукции или услуг, используемые средства транспортировки.

Характеристики структурных подразделений имеют в своем составе, помимо форм, аналогичных тем, которые входят в первую группу, формы описания информационных потоков, которые осуществляют связь задач внутри каждого подразделения, а также связи подразделений.

Форма описания документопотоков включает следующие характеристики: наименование входных документов, количество их экземпляров; объемные данные по каждому документопотоку; перечень информационных файлов, где используются эти документы; носитель, на котором хранятся данные; время создания; время использования; перечень полей файлов; выходные документы, получаемые на основе информации файлов.

Описания информационных потоков. Формы характеристик документов включают: наименование подразделения, тип документа (первичный, промежуточный или результатный), назначение документа, наименование документа, периодичность создания или время использования. Форма описания документов содержит: перечень показателей; описание структуры документов; перечень реквизитов; распределение реквизитов по разделам документа; типы реквизитов.

Форма характеристик процедур обработки данных включает: наименование подразделения, где используется процедура, задачу, в которую входит данная процедура; входную информацию, ее объемы; используемые файлы и их объемы; частоту обращения процедуры к файлу; блэк-схему процедуры; выходные данные процедуры. Форма описания процедур обработки содержит: наименование задачи; операции процедуры; количество операций; используемую технику; стоимостные и временные затраты.

Полученное в результате проведенной формализации описание объекта содержит исходные данные для проектирования ИС и определяет параметры будущей системы. Так, материальные потоки обуславливают объемы обрабатываемой информации, состав первичных данных, периодичность и сроки сбора, их источники, необходимые для разработки информационной базы. Функциональная структура объекта определяет комплексы автоматизируемых задач управления, для каждого из которых указывают: состав входных и выходных показателей; периодичность и сроки их формирования; процедуры использования данных показателей; распределение функций и процедур между персоналом и техническими средствами. Организационная структура объекта служит основанием для выделения лиц, определяющих условия решения задач обработки информации, а также получателей выходных показателей и документов.

Анализ материалов обследования

На основе формализованного описания предметной области выполняется этап «Анализ материалов обследования», целями которого являются:

- сопоставление всей собранной об объекте информации с теми требованиями, которые предъявляются к объекту, определение недостатков функционирования объекта обследования;
- выработка основных направлений совершенствования работы объекта обследования на базе внедрения проекта ИС, выбор направлений проектирования (выбор инструментария) и оценка эффективности применения выбранного инструментария;
- обоснование выбора решений по основным компонентам проекта ИС и определение общесистемных, функциональных и локальных требований к будущему проекту и его частям.

Рассмотрим технологическую сеть анализа материалов обследования, в которой в каждой из технологических операций используются документы обследования.

Анализ материалов обследования позволяет проектировщикам выделить и составить список автоматизируемых подразделений. На выбор объектов автоматизации оказывает влияние ряд факторов, например, таких, как:

- количество формализуемых функций в каждом конкретном подразделении;
- количество связей этого подразделения с другими подразделениями;

- важность этого подразделения в процессах управления объектом;
- степень подготовленности подразделения для внедрения ЭВМ и др.

Согласно этим факторам выделяют список наиболее важных подразделений. Например, для предприятия такими подразделениями являются отделы технико-экономического планирования, оперативного управления основным производством, технической подготовки производства, материально-технического снабжения, реализации и сбыта готовой продукции, бухгалтерия.

При *выявлении списка автоматизируемых задач* на операции, для которых необходимо разработать проекты, следует принимать к сведению следующие факторы:

- важность решения задачи для выполнения основных функций управления, деловых процессов и процедур в данном подразделении;
- трудоемкость и стоимость расчета основных показателей данной задачи за год;
- сильная информационная связь рассматриваемой задачи с другими задачами;
- недостаточная оперативность расчета показателей;
- низкая достоверность получаемых данных;
- недостаточное количество аналитических показателей, получаемых на базе первичных документов;
- неэквивалентный метод расчета показателей и др.

Кроме того, в период этой операции осуществляется выявление очередей проектирования решаемых задач. К задачам первой очереди относят наиболее трудоемкие задачи и обеспечивающие информацией все остальные задачи комплексов и подсистем (например, задачи планирования и бухгалтерского учета). Общим требованием к первоочередным задачам является получение нормативного коэффициента окупаемости капитальных затрат.

Далее выполняется операция, связанная с анализом всех полученных ранее результатов, исходных универсумов и *предварительным выбором комплекса технических средств* на операции ПЗ. На выбор типа ЭВМ оказывает влияние большое число факторов, которые принято объединять в следующие группы:

1. Факторы, связанные с параметрами входных информационных потоков, поступающих на обработку ЭВМ: объем информации, тип носителя информации, характер представления информации.
2. Факторы, зависящие от характера задач, которые должны решаться на ЭВМ, и их алгоритмов: срочность решения, возможность

разделения задачи на подзадачи, выполняемые на другой ЭВМ, количество файлов с условно-постоянной информацией.

3. Факторы, определяемые техническими характеристиками ЭВМ: производительность процессора, емкость оперативной памяти, поддерживаемая операционная система, возможность подключения различных устройств ввода-вывода.

4. Факторы, относящиеся к эксплуатационным характеристикам ЭВМ: требуемые условия эксплуатации.

5. Факторы, учитывающие стоимостные оценки затрат на приобретение, на содержание обслуживающего персонала, на проведение ремонтных работ.

Далее следует выполнение операции «Выбор типа операционных систем». Операционные системы осуществляют управление работой ЭВМ, ее ресурсами, запускают на выполнение различные прикладные программы, выполняют всевозможные вспомогательные действия по запросу пользователя.

К факторам, определяющим выбор конкретного класса ОС и его версии, относятся:

- необходимое множество поддерживаемых программных продуктов;
- требования к аппаратным средствам;
- возможность использования различных устройств ввода-вывода;
- требование поддержки сетевой технологии;
- наличие справочной службы для пользователя;
- наличие дружественного интерфейса и простота использования;
- возможность конфигурирования и быстрой настройки на новые аппаратные средства;
- быстроедействие;
- совместимость с другими ОС;
- поддержка новых информационных технологий и др.

Следующей операцией является операция «Выбор способа организации информационной базы (ИБ) и программного средства ведения ИБ». Информационная база имеет несколько способов организации как совокупность локальных файлов и интегрированную организацию в виде баз данных.

Локальная (файловая) организация подразумевает под собой хранение данных в виде совокупности не зависящих между собой локальных файлов, создаваемых для документа, задачи или комплекса задач.

Интегрированная база данных представляет собой совокупность взаимосвязанных, хранящихся вместе данных, используемых для одного или нескольких приложений. Данные, организованные в виде БД, могут быть организованы как централизованно (размещены на одной ЭВМ), так и в виде распределенных БД (размещенных на нескольких ЭВМ).

Программные средства ведения ИБ выбираются, исходя из класса систем хранения данных: системы управления файлами либо системы управления базами данных (СУБД). К основным факторам, определяющим выбор типа СУБД, относятся следующие:

- масштаб применения СУБД. По этому признаку выделяют персональные — настольные СУБД (например, FoxPro или Access) или промышленные — сетевые СУБД (например, Oracle, Sybase, Informix, MS SQL, ADABAS, InterBase и др.);
- язык общения. Разделяют СУБД с открытыми языками, замкнутыми или смешанными;
- число уровней в архитектуре. Существуют одноуровневые, двухуровневые, трехуровневые СУБД;
- выполняемые СУБД функции: информационные — организация хранения информации и доступа к ней и операционные функции, связанные с обработкой информации;
- сфера возможного применения СУБД: универсальное использование и специализированное.

При выполнении следующей операции осуществляется «Выбор методов и средств проектирования программного обеспечения системы», который напрямую зависит от выбранной технологии проектирования. В совокупность методов проектирования, используемых при каноническом подходе, входят такие, как метод структурного проектирования, модульного проектирования и др. Основными факторами, оказывающими влияние на выбор методов, являются их совместимость, сокращение времени и стоимостных затрат на проектирование, получение качественного продукта, который был бы удобен для последующей эксплуатации и сопровождения.

Выполнение всех этих операций завершается *составлением ТЭО и формированием ТЗ*. Целью разработки «Технико-экономического обоснования» проекта ИС являются оценка основных параметров, ограничивающих проект ИС, обоснование выбора и оценка основных проектных решений по отдельным компонентам проекта. При этом различают организационные параметры, характеризующие способы организации процессов преобразования информации в системе, информационные и экономические параметры, характеризующие затраты на создание и эксплуатацию системы, экономию от

ее эксплуатации. Основными объектами параметризации в системе являются задачи, комплексы задач, экономические показатели, процессы обработки информации.

Организационные параметры ИС дифференцируют по технологическим операциям процесса обработки информации: сбора, регистрации, передачи, хранения, обработки и выдачи информации.

Параметрами для *подготовительного этапа* технологии обработки информации могут быть: вид связи между источником информации и ЭВМ, территориальное размещение технических средств, наличие промежуточного носителя информации, способ обеспечения достоверности информации и т. п.

Для *основного этапа* технологии обработки информации в качестве параметров выступают: способ организации информационной базы, тип организации файлов, тип запоминающих устройств, режим обработки информации, тип ЭВМ, тип организации использования ЭВМ и т. п.

Для *заключительного этапа* — способ организации связи пользователя с ЭВМ, наличие промежуточного носителя, организация размножения результатной информации и т. п.

К *информационным параметрам* относятся такие, как достоверность, периодичность сбора, форма представления, периодичность обработки информации и т. д.

К *экономическим параметрам* ИС относятся: показатели годового экономического эффекта, коэффициента эффективности затрат и т. п.

Параметризация позволяет определить требования к разрабатываемой системе, оценить существующую ИС, пригодность типовых решений, выбрать проектные решения в соответствии с требованиями, предъявляемыми к ИС. К основным компонентам ТЭО относятся:

- характеристика исходных данных о предметной области;
- обоснование цели создания ИС;
- обоснование автоматизируемых подразделений, комплекса автоматизируемых задач, выбора комплекса технических средств, программного и информационного обеспечения;
- разработка перечня организационно-технических мероприятий по проектированию системы;
- расчет и обоснование эффективности выбранного проекта;
- выводы о техническом уровне проекта и возможности дальнейших разработок.

Состав технического задания на проектирование

На основе ТЭО разрабатываются основные требования к будущему проекту ИС и составляется «Техническое задание» согласно ГОСТ 34.602—89 «Техническое задание на создание автоматизированной системы», в состав которого входят следующие основные разделы.

1. В разделе «*Общие сведения о проекте*» указывают: полное наименование системы, код системы, код договора, наименование предприятия-разработчика и предприятия-заказчика, перечень документов, на основе которых создается система, плановые сроки начала и окончания работ по созданию системы, сведения об источниках финансирования, порядок оформления и предъявления заказчику результатов работ по созданию системы (частей).

2. Раздел описания «*Назначение, цели создания системы*» состоит из двух подразделов.

В подразделе «*Назначение системы*» указывают вид автоматизируемой деятельности и перечень объектов автоматизации, на которых предполагается ее использовать;

В подразделе «*Цели создания системы*» указываются наименования и требуемые значения технических, технологических, производственно-экономических и других показателей объекта автоматизации, которые будут достигнуты в результате внедрения ИС.

3. В разделе «*Характеристика объекта автоматизации*» приводятся: краткие сведения об объекте автоматизации; сведения об условиях эксплуатации объекта и характеристиках окружающей среды.

4. Раздел «*Требования к системе*» состоит из следующих подразделов: требования к системе в целом; требования к функциям (задачам), выполняемым системой; требования к видам обеспечения.

В подразделе «*Требования к системе в целом*» указывают требования к структуре и функционированию системы; к численности квалифицированных работников; к надежности и безопасности работы системы; к эргономике и технической эстетике, эксплуатации, техническому обслуживанию, ремонту системы; к защите информации от несанкционированного доступа; требования по сохранности информации при авариях; к защите от внешней среды; к патентной чистоте проектных решений; требования по унификации и стандартизации.

В подразделе «*Требования к функциям (задачам), выполняемым системой*», комплексам задач и отдельным задачам приводят по каждой подсистеме перечень функций, задач или их комплексов, подлежащих автоматизации; распределение их по очередям создания;

временной регламент реализации каждой функции, задачи или комплекса; требования к качеству реализации каждой функции, задачи, комплекса, к форме представления выходной информации; характеристики необходимой точности и времени выполнения, достоверности выдачи результата.

В подразделе «Требования к видам обеспечения» содержатся требования к математическому, программному, техническому, лингвистическому, информационному и методическому обеспечению ИС.

5. Раздел «Состав и содержание работ по созданию системы» должен содержать: перечень стадий и этапов работ по созданию системы в соответствии с ГОСТ 34.601—90; сроки выполнения; перечень организаций-исполнителей; перечень документов по ГОСТ 34.201—89 «Виды, комплектность и обозначение документов при создании автоматизированных систем», предъявляемых по окончании работ; вид и порядок проведения экспертизы технической документации и др.

6. В разделе «Порядок контроля приемки системы» указывают: виды, состав, методы испытания системы и ее частей; общие требования к приемке работ по стадиям; порядок утверждения приемных документов; статус приемочной комиссии.

7. В разделе «Требования к составу и содержанию работ по подготовке объекта автоматизации к вводу системы в действие» необходимо привести перечень необходимых мероприятий, которые следует выполнять при подготовке объекта к вводу ИС в действие и их исполнителей:

- приведение информации, поступающей в систему, к виду, пригодному для ввода в ЭВМ;
- создание условий функционирования объекта, при которых гарантируется соответствие создаваемой системы требованиям, содержащимся в ТЗ;
- создание необходимых для функционирования системы подразделений и служб;
- сроки и порядок комплектования штатов и обучения персонала.

8. В разделе «Требования к документированию» приводят перечень подлежащих разработке комплектов и видов документов, соответствующих требованиям ГОСТ 34.201—89 и научно-технической документации отрасли заказчика.

9. В разделе «Источники разработки» должны быть перечислены документы и информационные материалы (ТЭО, отчеты о законченных научно-исследовательских разработках, информационные материалы на отечественные, зарубежные системы-аналоги и др.).

10. В состав ТЗ при наличии утвержденных методик включают приложения, содержащие расчеты экономической эффективности системы; оценку научно-технического уровня системы.

Состав и содержание работ на стадии технорабочего проектирования

Работы на стадии «*Технорабочего проектирования*» выполняются на основе утвержденного «Технического задания». Разрабатываются основные положения проектируемой системы, принципы ее функционирования и взаимодействия с другими системами; определяется структура системы; разрабатываются проектные решения по обеспечивающим частям системы.

На стадии «Технорабочего проектирования» выполняются два этапа работ: *техническое и рабочее проектирование*. На первой стадии — «Техническое проектирование» — осуществляется логическая проработка функциональной и системной архитектуры ИС, в процессе которой строится несколько вариантов всех компонентов системы; проводится оценка вариантов по показателям стоимости, трудоемкости, достоверности получаемых результатов, и составляется «Технический проект» системы.

Все работы первого этапа можно разбить на две группы.

К первой относится разработка общесистемных проектных решений, в том числе:

- разработка общесистемных положений по ИС;
- изменение организационной структуры;
- определение функциональной структуры;
- разработка проектно-сметной документации и расчет экономической эффективности системы;
- разработка плана мероприятий по внедрению ИС.

При *разработке основных положений по системе* уточняются цели создания ИС и выполняемые ею функции, устанавливается ее взаимосвязь с другими системами и формируется документ «Основные положения». Далее *уточняется и изменяется организационная структура* и получается описание организационной структуры.

Наиболее принципиальной в данном комплексе работ является *разработка функциональной архитектуры ИС* на базе принципов выделения функциональных подсистем (модулей, контуров): предметного, функционального, смешанного (предметно-функционального) и проблемного.

Ко второй группе работ, выполняемых на этапе технического проектирования, относится разработка локальных проектных решений или следующие операции:

- разработка «*Постановки задачи*» для задач, входящих в состав каждой функциональной подсистемы;
- проектирование форм входных и выходных документов, системы ведения документов и макетов экранных форм;
- проектирование классификаторов экономической информации и системы ведения классификаторов;
- разработка структуры входных и выходных сообщений;
- проектирование состава и структур файлов информационной базы;
- проектирование немашиной и внутримашинной технологии решения каждой задачи;
- уточнение состава технических средств.

Основным компонентом локальных проектных решений, являющимся базой для разработки информационного, программного и технологического обеспечения для каждой задачи, является «*Постановка задачи*». Этот документ содержит три составные части:

- характеристику задачи;
- описание выходной информации;
- описание входной информации.

В состав раздела «*Характеристика задачи*» входят следующие компоненты: описание цели; назначение решения конкретной задачи; перечень функций и процессов, реализуемых решаемой задачей; характеристика организационной и технико-экономической сущности задачи; обоснование целесообразности автоматизации решения задачи; указание перечня объектов, для которых решается задача; описание процедур решения задачи; указание периодичности решения задачи и требований к организации сбора первичных данных; описание связей с другими задачами.

Под *целью* автоматизации решения задачи подразумевается получение определенных значений экономического эффекта в сфере управления какими-либо процессами системы или снижение стоимостных и трудовых затрат на обработку информации, улучшение качества и достоверности получаемой информации, повышение оперативности ее обработки и т. д., т. е. получение косвенного и прямого эффекта от внедрения данной задачи.

Под *экономической сущностью решаемой задачи* понимаются состав экономических показателей, рассчитываемых при ее решении, документы, в которые заносятся эти показатели, перечень исходных показателей, необходимых для получения результатных

документов и наименования тех первичных документов, в которых они содержатся.

Организационная сущность задачи — это описание порядка решения задачи; организационной формы, применяемой для ее решения; режима решения; состава файлов с постоянной и переменной информацией; способа получения и ввода первичной информации в ЭВМ; формы выдачи результатной информации: на печать, на экран, на магнитный носитель или передача по каналам связи.

Описание алгоритма решения задачи включает формализованное описание входных и результатных показателей и перечень формул расчета результатных показателей в случае решения задачи прямым методом счета или описание математической модели, экономико-математического метода, применяемого для ее реализации, и перечня последовательных шагов выполнения расчетов.

Далее указываются *периодичность* решения задачи и регламент выдачи результатных документов, *требования к организации сбора исходных данных*, т. е. к способу и техническим средствам съема, регистрации, сбора и передачи данных для обработки. Большое значение имеет описание *связей задач* внутри функциональной подсистемы, а также с задачами других подсистем или с внешней средой.

Описание выходной информации включает в себя:

- перечень и описание выходных сообщений, документов;
- перечень структурных единиц информации;
- периодичность возникновения и сроки получения информации;
- наименование, идентификатор каждой формы документа.

Описание входной информации включает:

- перечень входных сообщений;
- перечень структурных единиц информации;
- описание периодичности возникновения и сроков получения информации;
- наименование и идентификатор каждой формы документа.

Далее для каждой задачи разрабатываются все компоненты информационного, технического и лингвистического обеспечения, а также некоторые компоненты программного обеспечения.

Результатом работ на данной стадии является утвержденный «Технический проект», состав и содержание которого регламентируются стандартом (ГОСТ 34.201—89).

На втором этапе — «*Рабочем проектировании*» — осуществляется техническая реализация выбранных наилучших вариантов и разрабатывается документация «Рабочий проект». Наиболее ответственными работами, выполняемыми на этом этапе, являются «*Кодирование и составление программной документации*», содержание которых

отражено, например, в [8, 9]. В их состав входят следующие компоненты:

- описание программ;
- спецификация программ;
- тексты программ;
- контрольные примеры;
- инструкции для системного программиста, оператора и пользователя.

Большую роль в деле эффективного использования разработанного проекта ИС играет качественная *технологическая документация*, входящая в состав «Рабочего проекта». Эта часть проекта разрабатывается на операции «Изменение организационной структуры» и предназначена для использования специалистами в своей деятельности на каждом автоматизированном рабочем месте.

В состав технологической документации входят: технологические карты, разрабатываемые на процессы обработки информации при решении задач каждого класса, и инструкционные карты, составляемые на каждую технологическую операцию.

Технологическая документация разрабатывается в соответствии с требованиями ГОСТ 3.11.09—82 «Система технологической документации. Термины и определения основных понятий» и составляет содержание технологического обеспечения ИС, которое можно разделить на несколько типов в соответствии с выделением следующих классов задач, решаемых в ИС:

- системы обработки данных (СОД);
- системы поддержки принятия решений (СППР);
- системы автоматизированного проектирования новой продукции (САПР) и т. д.

К числу работ, выполняемых на этом этапе, относится «*Разработка правовых инструкций*», определяющих права и обязанности специалистов, которые работают в условиях функционирования на предприятии компонентов ИС.

Заключительной операцией служит «Оформление документации Рабочего проекта» согласно ГОСТам.

Состав и содержание работ на стадиях внедрения, эксплуатации и сопровождения проекта

На стадии «*Внедрение проекта*» проводятся подготовка и постепенное освоение разработанной проектной документации ИС заказчиками системы. В процессе выполнения работ на этой стадии осу-

ществляется выявление частных и системных принципиальных недоработок в предлагаемом для внедрения проектом решении.

Внедрение может осуществляться с использованием следующих методов:

- последовательного метода, когда постепенно внедряется одна подсистема за другой и задачи следуют одна за другой;
- параллельного метода, при котором все задачи внедряются во всех подсистемах одновременно;
- смешанного подхода, согласно которому проектировщики, внедрив несколько подсистем первым методом и накопив опыт, приступают к параллельному внедрению остальных.

Недостатком первого подхода является увеличение длительности внедрения, что приводит к росту стоимости проекта. При использовании второго подхода сокращается время внедрения, но возникает возможность пропуска ошибок в проектной документации, поэтому чаще всего используют смешанный метод внедрения проекта ИС. Внедрение проекта осуществляется в течение трех этапов, которыми являются:

- подготовка объекта к внедрению;
- опытное внедрение;
- сдача проекта в промышленную эксплуатацию.

Первый этап — *«Подготовка объекта к внедрению»*. На этом этапе осуществляются следующие операции:

- изменяется организационная структура объекта (предприятия);
- набираются кадры соответствующей квалификации в областях обработки информации, эксплуатации системы и сопровождения проектной документации;
- оборудуется здание под установку вычислительной техники;
- выполняются закупка и установка вычислительной техники с периферией;
- в цехах, отделах устанавливаются средства сбора, регистрации первичной информации и передачи по каналам связи;
- осуществляется установка каналов связи; проводится разработка новых документов и классификаторов;
- осуществляется создание файлов информационной базы с нормативно-справочной информацией.

На вход этого этапа поступают компоненты «Технического проекта» в части «Плана мероприятий по внедрению», решения по техническому и информационному обеспечению, технологические и инструкционные материалы «Рабочего проекта». В результате выполнения этапа составляется «Акт готовности объекта к внедрению» проекта ИС. Затем формируется состав приемной комиссии, разра-

бывается «Программа проведения опытного внедрения» и издается «Приказ о начале опытного внедрения».

Второй этап — «*Опытное внедрение*». На этом этапе внедряются проекты нескольких задач в нескольких подсистемах. В процессе опытного внедрения выполняются следующие работы:

- подготовка исходных оперативных данных для задач, которые проходят опытную эксплуатацию;
- ввод исходных данных в ЭВМ и выполнение запланированного числа реализации;
- анализ выходных данных на предмет наличия ошибок.

В случае обнаружения ошибок осуществляются поиск их причин и источников, внесение коррективов в программы, в технологию обработки информации, в работу технических средств, в исходные оперативные данные и в файлы с условно-постоянной информацией. Кроме того, выявляется неквалифицированная работа операторов, что служит основанием для проведения комплекса мер по улучшению подготовки кадров.

После устранения ошибок получают «Акт о проведении опытного внедрения», который служит сигналом для начала выполнения следующего этапа.

На третьем этапе — «*Сдача проекта в промышленную эксплуатацию*» — используют следующую совокупность документов:

- договорную документацию;
- приказ на разработку ИС;
- ТЭО и ТЗ;
- исправленный техно-рабочий проект;
- приказ о начале промышленного внедрения;
- программу проведения испытаний;
- требования к научно-техническому уровню проекта системы.

В процессе сдачи проекта в промышленную эксплуатацию осуществляются следующие работы:

- проверка соответствия выполненной работы договорной документации по времени выполнения, объему проделанной работы и затратам денежных средств;
- проверка соответствия проектных решений по ИС требованиям ТЗ;
- проверка соответствия проектной документации ГОСТам и ОСТам;
- проверка технологических процессов обработки данных по всем задачам и подсистемам;
- проверка качества функционирования информационной базы, оперативности и полноты ответов на запросы;
- выявление локальных и системных ошибок и их исправление.

Кроме того, приемная комиссия определяет научно-технический уровень проекта и возможности расширения проектных решений за счет включения новых компонентов. В результате выполнения работ на данном этапе осуществляется доработка «Технорабочего проекта» за счет выявления системных и локальных ошибок и составляется «Акт сдачи проекта в промышленную эксплуатацию».

На четвертой стадии — «Эксплуатация и сопровождение проекта» — выполняются следующие процессы:

- эксплуатация проекта;
- сопровождение и модернизация проекта.

На этой стадии решается вопрос о том, какими силами (персоналом объекта-заказчика или организации-разработчика) будут осуществляться эксплуатация и сопровождение проекта, и в случае выбора второго варианта заключается «Договор о сопровождении проекта».

В процессе выполнения этапа «Эксплуатация проекта» осуществляются исправления в работе всех частей системы при возникновении сбоев, регистрация этих случаев в журналах, отслеживание технико-экономических характеристик работы системы и накопление статистики о качестве работы всех компонентов системы.

На этапе «Сопровождение и модернизация проекта» выполняется анализ собранного статистического материала, а также соответствия параметров работы системы требованиям окружающей среды посредством создаваемой для этих целей комиссии. Полученные результаты позволяют:

- сделать заключение о необходимости модернизации всего проекта или его частей;
- определить объемы доработок, сроки и стоимость выполнения этих работ с целью получения «Технорабочего проекта», прошедшего модернизацию.

В случае выявления факта морального старения проекта комиссией принимается решение о целесообразности проведения его утилизации или разработки нового проекта.

3.4. Проектирование ИС и РБП — реинжиниринг бизнес-процессов (BPR — business process reengineering)

Рассматриваемые здесь подходы касаются в основном *корпоративных АИС* и заключаются во встраивании *информационных технологий (ИТ)* в *деловые процессы организации (бизнес-процессы)* в качестве организационной и неотъемлемой части последних.

Это встраивание является основанием для переосмысления и переструктуризации самих бизнес-процессов — *реинжиниринг бизнес-процессов (РБП) или, в оригинале, — business process reengineering (BPR)*.

В совокупности АИС важное место занимают *автоматизированные системы управления предприятиями*, которые называют *корпоративными ИС*, подчеркивая их направленность на решение широкого круга информационно-управленческих задач, причем во многих, часто удаленных подразделениях, на самых разных уровнях управления предприятием, объединением или отраслью.

Разработка корпоративных информационных систем в настоящее время переживает период перехода к концепции нового системного проектирования. Влияние новых информационных технологий и систем на рост рыночной динамики (наряду с другими, в первую очередь, экономическими причинами) объективно привело к возникновению так называемых *киберкорпораций* (рис. 3.1), постоянно и гибко реконструирующих свой бизнес [13].

Проектирование таких ИС (ранее определяемых в отечественной практике как АСУП или ОАСУ) всегда содержало декларации о включении человека в эти системы. Если для некоторой информационно-справочной системы общего назначения ее пользователь мог (пусть с натяжкой) выступать как элемент, внешний по отношению к системе, то рассматриваемые ИС по своей сути — *человеко-машинные информационно-управляющие системы*. Этот факт часто упускается еще на этапах анализа и построения общей архитектуры ИС. (Выражение «пользователь системы» дополнительно может

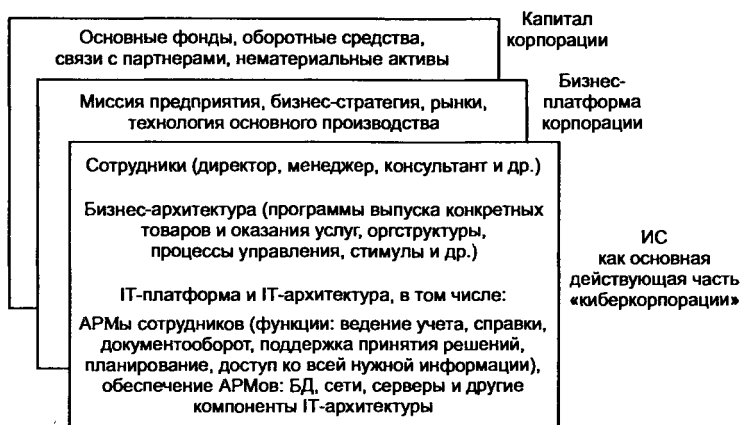


Рис. 3.1. Трёхслойная схема современного предприятия — «киберкорпорации»

подталкивать к концептуальной ошибке.) Теперь, когда в центр бизнес-реинжиниринга ставится всемерная поддержка, усиление информационных и аналитических возможностей деятельности каждого работника, какое-либо отделение ИС от функционирования предприятия в целом становится неприемлемым. В силу этого, в процессах проектирования целесообразно считать, что корпоративная ИС составляет информационно-управляющую систему, включающую бизнес-архитектуру предприятия, его персонал, используемую ИТ-архитектуру, и является действующей частью *киберкорпорации*.

Это положение позволяет точнее определить расширяющиеся границы корпоративной ИС. Следует исходить из того, что *в виде ИС проектируется часть предприятия*, которая непосредственно осуществляет «бизнес», т. е. организационно-производственную деятельность.

С учетом этого, в более широком смысле, новое системное проектирование (НСП) является методологией соединения постоянного бизнес-реинжиниринга и новых информационных технологий (ИТ).

Приведем две иллюстрации, заимствованные из [27], отражающие принципиальные недостатки каскадной схемы и разницу между идеализированными и фактическими процессами проектирования АИС. На рис. 3.2, *а* показан идеальный вариант каскадной схемы, по которой полагается проектировать ИС. Рис. 3.2, *б* иллюстрирует реальные итерации, заставляющие возвращаться к этапам проектирования и определения требований даже в ходе работ по комплексному тестированию ИС.

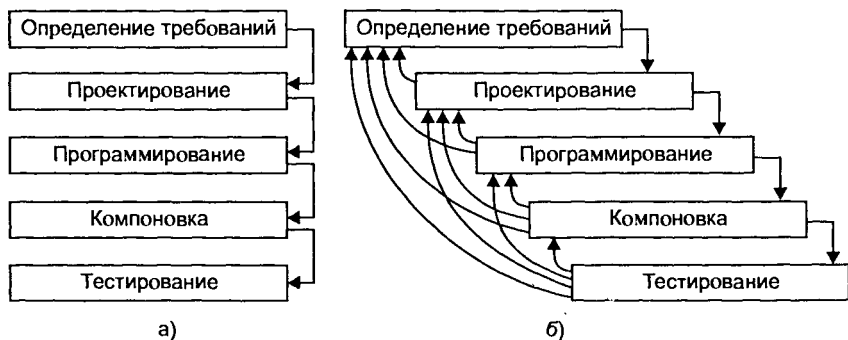


Рис. 3.2. Идеальный (*а*) и реальный (*б*) процессы разработки по каскадной схеме (на основе [39])

На рис. 3.3, а показано плановое распределение специалистов, которые должны были бы работать (при последовательном, конвейерном стиле) на разных этапах каскадного проектирования.

На рис. 3.3, б приведена соответствующая схема Э. Ферентино, из которой видно, что группа, определяющая требования пользователей и разрабатывающая внешние спецификации системы, работает постоянно на всем цикле жизни системы, выполняя и корректирующие, и контролирующие функции. С тех пор требования к параллельности и спиральности проектирования, комплексности групп разработчиков возросли. Тем не менее до сих пор большое число управляющих проектами ИС (в том числе больших ИС) считает верной схему рис. 3.3, а.

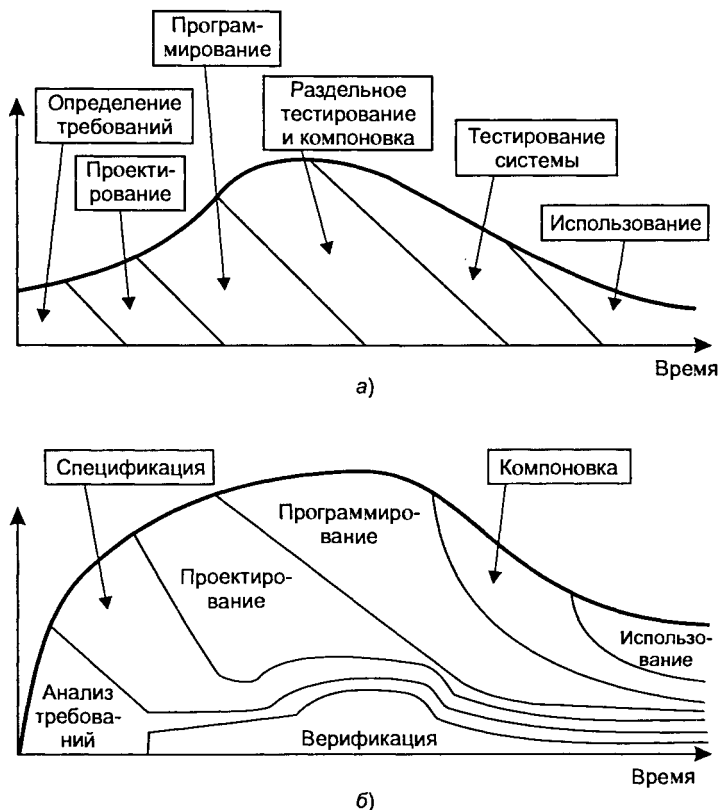


Рис. 3.3. Распределения людских ресурсов при разработке программной системы: а — конвейерное (1970 г.); б — реальное (схема Э. Ферентино, 1982 г.)

Основные недостатки каскадных схем

1. *Отставание* — существенное запаздывание с получением результатов.

2. *Бесполезность или вред* — и в зарубежной, и в отечественной литературе практики и аналитики оценивали проектирование ИС как действие, очень часто ведущее к примитивной автоматизации (по сути — «механизации») существующих производственных функций работников. В [26] отмечается, что «...легче идти по проторенной дорожке документирования сложившегося бумажного потока, чем определять насущные потребности бизнеса».

Таким образом, фиксировались неправильные способы работы, возможно, приносящие значительный вред предприятию (в отечественной практике АСУ популярным было выражение «автоматизация беспорядка»).

3. *Жесткость* — из-за определяющего влияния иерархических структур на процессы и результаты проектирования ИС для представления в ИС функций и данных применявшиеся подходы получили общее условное название «структурное проектирование». Однако жесткость иерархических структур ограничивает их полезность, и чем дальше, тем менее допустимы эти ограничения.

Один из итогов заключается в том, что в значительной степени именно новейшие достижения в ИТ давали потребителям новые возможности предъявлять более высокие требования к производителям и стимулировать конкуренцию.

С другой стороны, новые ИТ служат технологической платформой реального РБП на предприятии и создания новых отношений и возможностей для людей в компьютеризованных коллективах. В этих фактах — корни взаимосвязей ИТ и РБП. Они состоят в том, что и новые рыночные требования, и ответ предприятий на их появление в большой степени стали возможными только на основе последних достижений в ИТ. Более того, ИТ как инструмент для достижения целей РБП и ИТ как компонент самого РБП перекрываются или соединяются.

В [27] содержится описание влияний реконструкций бизнес-процессов на новые ИТ-архитектуры, в первую очередь — на архитектуры систем с базами данных. Там В. Меллинг описал модель Дж. Хендерсона для понимания взаимодействия бизнес-структур и ИТ (рис. 3.4). В этой модели определены:

- основная бизнес-платформа — это набор стратегий, рынков, предписаний, технологий производства продуктов и ресурсов,

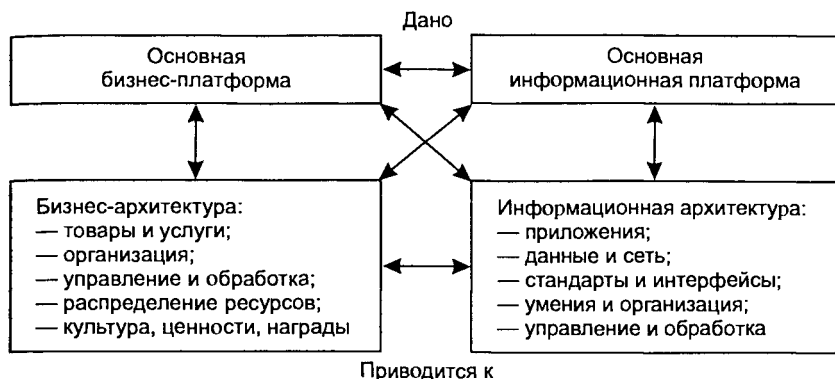


Рис. 3.4. Модель Дж. Хендерсона [27]

выбранных предприятием как соответствующий поставленным целям;

- бизнес-архитектура — набор товаров и услуг, организационных структур, процессов управления, распределения ресурсов, ценностей и стимулов, который является необходимым для внедрения основной бизнес-платформы;
- основная ИТ-платформа — ряд адекватных компьютерных технологий, которые могут быть доступны предприятию, и способы, которыми эти технологии могут быть использованы для повышения конкурентоспособности;
- ИТ-архитектура — набор определенных архитектур и продуктов, выбранный для реализации основной ИТ-платформы, а также инфраструктуры поддержки, уровни квалификации, процессы принятия решений и административные механизмы, используемые для развертывания этих архитектур.

Основываясь на этой модели, можно сделать следующие выводы.

1. Существует двунаправленное воздействие основных бизнес- и ИТ-платформ.

2. Если основная бизнес- или ИТ-платформа меняется, то маловероятно, что соответствующая наследуемая ИТ-архитектура сохранится.

3. Соответствие между бизнес- и ИТ-архитектурами является решающим фактором успеха, но на достижение этого успеха может уйти значительное время.

В табл. 3.3 показана взаимосвязь бизнес-архитектуры и ИТ-архитектуры в современных условиях.

Таблица 3.3. Взаимосвязь бизнес- и ИТ-архитектур

Бизнес-архитектура	ИТ-архитектура
Автоматизация бизнес-подразделения	Различные поставщики оборудования, сети, платформы, операционные системы. Покупай, а не производи
Меньшее количество уровней управления	Повсеместные почта, заметки, управление образами, телеконференции
Реорганизация работы ориентированности на задачи к ориентированности на процессы	Переход от OLTP-мониторов к менеджерам процессов
Интеграция цепочки поставщиков	Приложения клиент/сервер от нескольких поставщиков. Многопротокольная маршрутизация. Надежная передача сообщений
Глобализация	Переносимость приложений различных производителей. Глобальные сети. Бесперебойная (24×365) работа
Интенсивная фокусировка на обслуживание клиента	Быстрое развитие приложений. Приложения клиент/сервер от нескольких поставщиков. Надежная передача сообщений. Непрерывная (24×365) работа
Возросшая мобильность рабочих Рост телекоммуникаций	Беспроволочные коммуникации. Асинхронные сообщения. Тиражирование баз данных. Непрерывная (24×365) работа
Интенсивная фокусировка на стоимости	Использование новейших технологий

ИТ-архитектуры и общий бизнес-реинжиниринг

Если динамика рынка заставляет корпорацию делать реинжиниринг основных видов деятельности, т. е. менять бизнес-платформу, то должна измениться не только архитектура, но и ИТ-платформа.

Отсюда следуют выводы:

- методы проектирования ИС в НСП должны обладать повышенной гибкостью для обеспечения живучести предприятия в условиях общего, тотального бизнес-реинжиниринга;
- стоимость такого проектирования ИС возрастает, так как в общем случае при смене бизнеса должна обеспечиваться смена не только ИТ-архитектуры, но и ИТ-платформы;
- для виртуальных киберкорпораций есть надежда обеспечить большую стабильность в применяемых ИТ, так как с изменением чисто управленческих задач можно ограничиваться корректировкой ИТ-архитектуры в части отдельных прикладных функциональных и информационных компонент;

- для обеспечения такого уровня изменчивости ИТ-архитектуры необходимо применять комплекс инструментов и специальных проектных приемов, обеспечивающих соответствующую динамику компонентного реинжиниринга ИС предприятия;
- методологии и отдельные методы НСП должны радикально отличаться от классических.

В результате всей суммы произошедших изменений стало возможным говорить о возникновении нового направления разработки АИС. Это направление — новое системное проектирование, или НСП, является интеграцией подходов бизнес-реинжиниринга, новых информационных технологий и социопсихологических методов, позволяющих учесть то, что в производственных процессах и в ИС должны работать конкретные люди.

Так же, как объявление РБП в качестве нового течения оправдано новыми рыночными условиями и взаимосвязями с ИТ, так и объявление НСП объясняется, в первую очередь, новыми требованиями к создаваемым корпоративным ИС, а также методами проектирования, развиваемыми в самих ИТ.

Высокая динамика изменений в требованиях к корпоративной ИС требует планирования, постоянной модификации как минимум ИТ-архитектуры, использованной в ИС. В этих условиях старые методики структурного проектирования чаще всего перестают работать. Проектирование ИС по классическим правилам полноты и целостности, мало жизненное ранее, часто становится практически бессмысленным.

Основные принципиальные положения НСП следующие:

- определяется объект НСП — проектируемая как человеко-машинная система ИС, которая непосредственно осуществляет организационно-производственную деятельность предприятия, а не является всего лишь инфраструктурной или сервисной прослойкой;
- описывается перечень работ, составляющих основу НСП, формирующих требования к ИТ для НСП, дающих базис для выстраивания организационных схем процессов проектирования ИС;
- предварительно формулируются принципиальные положения, необходимые для конкретизации этих работ и методов именно в контексте НСП;
- данный перечень работ сопровождается указанием методов и средств их выполнения, необходимыми, в первую очередь, при проектировании ИС;

- рассматриваются адаптивные подходы к организации процессов проектирования ИС, учитывающие современную динамику требований и три составные части НСП;
- приводятся рекомендации руководителям проектов ИС по формированию ИТ и проектных бригад;
- отмечается тенденция в дальнейшем развитии проектирования ИС на перспективу ближайших лет.

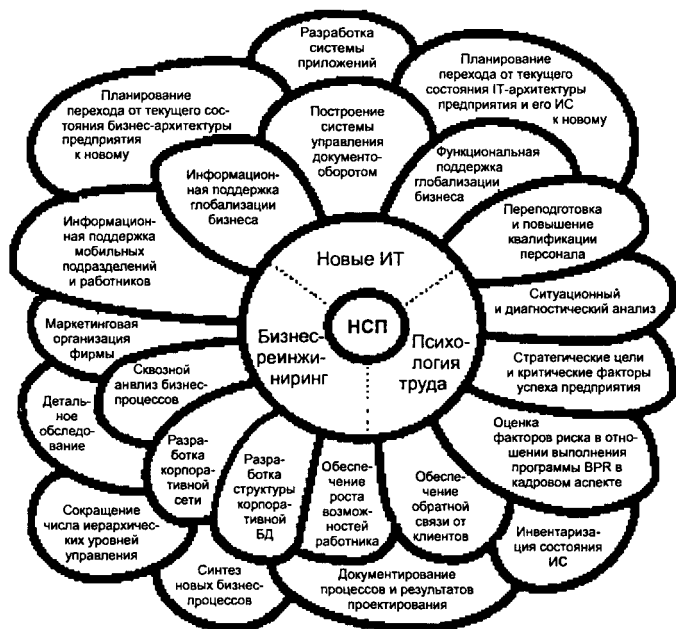


Рис. 3.5. Процессы нового системного проектирования

Работы в НСП проводятся в той последовательности, которая обеспечивает адаптируемость к условиям конкретного предприятия и проекта ИС. На рис. 3.5 иллюстрируются приводимые ниже процессы НСП в виде модели-«ромашки».

Направления НСП и используемые в них методы

Ситуационный и диагностический анализ положения предприятия. Применяются методы и программные инструменты:

- финансового анализа положения предприятия (финансовой устойчивости, ликвидности баланса, коэффициентов деловой активности и др.);

- степени и динамики прибыльности отдельных товаров и процессов (продуктов, услуг, технологий, работ);
- маркетингового анализа (товаров и услуг, имиджа предприятия и конкурентов и др.) на различных секторах рынка, маркетингового прогноза;
- социопсихологического анализа (установок руководства предприятия, других групп работников, кадровой ситуации в целом), его информационной поддержки и автоматизации.

Анализ стратегических целей предприятия и критических факторов его успеха. Делается заключение о технологических, рыночных и общественных тенденциях и возможностях предприятия, формулируются положения новой бизнес-архитектуры или (в случае более радикального реинжиниринга) положения новой бизнес-платформы (см. рис. 3.4).

Используются функции прогноза в аналитических маркетинговых системах, базы данных прецедентов, линии открытой конъюнктурной информации, сведения о наиболее успешно работающих конкурентах и др.

Анализ факторов риска предприятия в отношении выполнения программ бизнес-реинжиниринга в кадровом аспекте (для жесткого РБП, тотального реинжиниринга, структурной реорганизации или др.) и возможности управления этими факторами.

Применяются методы социопсихологических экспертиз, оценивается возможность перестройки установок персонала, планируются тренинги персонала начиная с руководства предприятия, моделируются последовательности других шагов по подготовке персонала к реинжинирингу.

Инвентаризация и оценка состояния ИС предприятия: по применяемым прикладным системам, системам классификации и кодирования информации, информационному составу баз данных, методам поддержки принятия решений, использованию локальных и глобальных сетевых технологий, составу компьютерного парка, открытости архитектуры и другим показателям качества применяемых ИТ. Кроме того, оценивается полезный результат, который вносит каждая подсистема (автоматизированная задача, функция) в деятельность предприятия.

Применяются средства информационного и функционального моделирования систем (отдельные инструменты для описания ИТ-моделей, CASE-системы, автоматизированные тезаурусные системы, системы моделирования локальных компьютерных сетей и др.), логические правила классификации понятий, известные системы классификации и кодирования, используются сведения о стан-

дартах в области ИТ, промышленных технологиях, служащих типичными и перспективными представителями ИТ в своих классах. Определяются стоимостные оценки эффективности использования каждой подсистемы (при невозможности их получения — оценки в натуральных единицах или в качественном выражении).

Детальное обследование предприятия (или его частей) и построение моделей существующей структуры организации, процедур и показателей деятельности (текущее состояние оргструктуры, нормативные документы предприятия, показатели результатов деятельности подразделений и предприятия в целом), анализ документов и регламентов, используемых в производственных процессах. Оценивается полезный результат, который вносит каждая автоматизированная задача, комплекс функций в деятельность предприятия.

Применяются CASE-системы и отдельные специальные инструменты моделирования: средства укрупненного формального описания объекта (например, описание иерархии функций и подразделений), декларативные детальные функциональные модели бизнес-процедур, имитационные модели в терминах массового обслуживания, динамические модели на сетях Петри, декларативные описания информационных элементов и структур данных, составляющих потоки данных; строится (или дополняется) тезаурус понятий, составляющих специфическую для предприятия понятийную модель и определяющих профессиональный жаргон, строятся активные понятийные модели на основе фреймовых представлений и др. Применяются стоимостные оценки эффективности автоматизации задач (комплексов функций), при невозможности их получения используются оценки в натуральных единицах или качественном измерении.

Сквозной анализ и синтез новых бизнес-процессов: определяется и оптимизируется их вклад в производственную деятельность, в первую очередь — в виде конечных результатов и показателей эффективности.

Применяются методы функционального и организационного проектирования:

- вычленение главных или определение новых ключевых функциональных ролей работников с их ориентацией на результат бизнес-процессов в целом, проектирование объемов полномочий и ресурсов, необходимых для выполнения всех функций в процессе;
- проектирование новых и планирование преобразований существующих процессов и структуры для усиления функциональ-

ных ролей работников в бизнес-процессах и минимизации числа лиц, принимающих решения;

- ввод в бизнес-процессы системы измерений, позволяющей в каждый момент времени оценивать состояние дел в денежных единицах, процентах роста, прогнозе времени выполнения или отклонения от плановых показателей и т. д.

Строятся (впоследствии — реконструируются) целевые модели предприятия: понятийная, организационная, информационная, функциональная, территориальная и др., при этом применяются:

- программные инструменты (компоненты CASE-систем, отдельные программы) моделирования и оценки бизнес-процессов, использующие методы формализованного статического описания, функционально-стоимостного бизнес-анализа (ABC, «activity-based costing»), динамического моделирования (CP-модели, модели по типу языка JPSS и др.);
- CASE-системы для фиксации принимаемых решений в виде новых функциональных, информационных, объектно-ориентированных и других моделей.

Введение необходимых элементов маркетинговой организации фирмы как производителя рыночных товаров (услуг).

Разрабатываются или приобретаются информационно-аналитические системы для поддержки выполнения маркетинговых экспертиз в жизненном цикле товара, применяются системы поддержки хранилищ данных (Data WareHouse — DWH) и оперативной аналитической обработки (Online Analyse Processing — OLAP).

Сокращение числа иерархических уровней управления с использованием:

- социопсихологических методов компоновки новых структур и отношений (специальные тренинги, мониторинг отношений, корректировка видов и форм мотиваций);
- средств автоматизированной поддержки групповой работы в новых условиях: средства workflow, системы групповой разработки, параллельного проектирования и др.;
- БД шаблонов-заготовок рабочих документов, нормативов, постоянного отслеживания реальной текущей ситуации с доступными работнику ресурсами;
- корпоративной почты, телеконференций и видеоконференций.

Создание и информационная поддержка автономных и мобильных бизнес-подразделений и работников, обеспечение «полевых» инженеров и ремонтников, бригад спасателей или скорой помощи постоянной связью с корпоративной ИС.

Применяются различные технические средства ИТ, например:

- ноутбуки с модемной (в том числе — радио) связью и коммуникационными программами, имеющими простой дружественный интерфейс;
- тиражирование (репликации) документов и баз данных, асинхронные режимы работы с ИС в трехзвенных архитектурах «клиент — сервер приложений — сервер баз данных» и др.

Обеспечение роста возможностей каждого работника, выполнение максимума функций в бизнес-процессах лицом, получающим конечный результат.

Здесь также применяются технические методы и средства новых ИТ:

- средства доступа ко всем необходимым данным в режимах использования распределенных баз данных, средства репликаций данных, управления событиями в данных и процессах обработки транзакций;
- концепция и программные средства DWH, средства OLAP, быстрой разработки приложений (RAD) для создания «ИС руководителя» (EIS), средства поддержки принятия решений (DSS) на основе DWH, OLAP и EIS;
- средства DSS на основе методов логического вывода, нейронных сетей и нейрокомпьютеров, анализа прецедентов, и др.;
- единый интерфейс (APM) пользователя для работы с разными компонентами данных и приложений, использование в этом интерфейсе средств, повышающих простоту поиска информации и обращения к конкретным прикладным функциям, например интерфейсы географических систем, естественного языка, речевого ввода.

Разработка концепции и структуры корпоративной БД для новой ИС, реализация структуры БД и управление ее развитием.

Применяются:

- методы компонентного проектирования предметных баз данных как для операционных, так и для исторических БД хранилищ данных, архивов документов, геоинформационных данных и др.;
- разработка процедур компонентной коррекции корпоративной БД при изменении бизнес-процедур, видов деятельности, применяемых приложений и географического размещения предприятия;
- постоянная актуализация понятийной модели предприятия для учета новых понятий, возникающих как при замене при-

кладных компонентов на функционально сходные, так и при изменении видов деятельности предприятия;

- подключение корпоративной БД к каналам глобальной информационной магистрали, предоставление прав на включение информации из нее в БД работникам всех иерархических уровней;
- динамическое администрирование фрагментами распределенной корпоративной БД при изменении их логической структуры, частоты их использования и места размещения.

Разработка концепции и структуры внутренней корпоративной сети. Применяются технические стандарты открытых систем, (например, технологии Интернет и WWW для построения корпоративной сети по типу Интранет).

Закладывается минимум оперативного резервирования ресурсов сети для снятия ограничений на ее развитие и реконфигурирование.

Разработка системы приложений как набора компонентов, опирающихся на общую понятийную модель и доступных для переуконфигурирования включением новых, в первую очередь, покупных компонентов.

Применяются:

- СУБД и модели БД с использованием языков (моделей данных), отвечающих промышленным юридическим стандартам представления и обработки данных;
- опробованные юридические стандарты открытых систем в части обмена запросами, данными, документами, объектами;
- разработка приложений на основе переносимых RAD-систем (в том числе с элементами объектно-ориентированного программирования).

В перспективе возможно использование новых стандартов в области объектно-ориентированных сред.

Информационная и функциональная поддержка глобализации бизнеса. Применяется подключение предприятия к глобальным коммуникациям. Используются:

- глобальные цифровые (компьютерные) сети и их услуги, например, построение выходов из корпоративных сетей в Интернет;
- инструменты и средства работы в глобальных сетях: средства гипертекстового просмотра БД серверов WWW, приложения для удаленных финансовых расчетов и др.;
- режимы и стандарты информационной супермагистрали для повсеместного доступа к информации любых видов — от прейскурантов и типовых условий возможных бизнес-партне-

ров до динамических потоков конъюнктурной и справочной информации общего характера;

- отказ от встраивания ограничений на возможности компьютерного общения в аппаратную архитектуру, архитектуру каналов связи, в программное обеспечение или в выделенный центр удаленного администрирования распределенной корпоративной сетью;
- средства защиты конфиденциальных данных, не ограничивающих возможности свободного обращения абонентов по нужному адресу (кроме особых случаев, в которых оправдано создание *компьютерных островов*);
- режимы работы коммуникаций и ИС в режиме *24 (часа) × 365 (дней)*.

Построение системы поддержки и управления документооборотом как части системы реализации актуального набора бизнес-процедур.

Такая система применяется в качестве средства планирования организации работ, измерений показателей их выполнения, контроля и самоконтроля исполнения.

Для этого используются средства корпоративной и глобальной электронной почты, электронных архивов документов, инструментальных и инфраструктурных систем классов *groupware* и *workflow*, написание и администрирование конкретных регламентов (бизнес-процедур), охватывающих сотрудников предприятия, предоставление каждому сотруднику динамических отчетов о ситуации с выполнением регламентированных работ, достигнутыми значениями оценочных показателей и др.

Переподготовка и повышение квалификации персонала. Работникам предоставляется максимум базовой информации в качестве основы для принятия самостоятельных решений, что способствует формированию у них знаний и навыков, сводящих последующие накладные расходы сотрудников на осуществление бизнес-процедур к минимуму, например:

- мультимедийные обучающие компьютерные программы с динамическими сценариями имитации различных ситуаций;
- контекстные подсказки, гипертекстовые справочные руководства, контекстные обучающие программы.

Планирование последовательности шагов перехода от текущего состояния бизнес-архитектуры предприятия к новому (с оценкой стоимости перехода).

Планирование таких шагов в части подготовки персонала, управления ресурсами и проектами, финансового учета и анализа и др., в том числе с использованием программных систем управления

проектами (построение и динамический пересчет линейных и сетевых план-графиков, планирование ресурсов, оценка стоимости проекта).

Планирование и осуществление перехода от текущего состояния ИТ-архитектуры предприятия и его функционирующей ИС к новому.

Например, в части реконструкции корпоративной БД и комплексов приложений применяются:

- программные системы управления проектами разработки ИС;
- средства разработки и реализации схем переноса и реинжиниринга БД;
- программы интерфейсного использования имеющихся (наследованных) или вновь комплексируемых компонентов: приложений, предметных БД и подсистем в новой ИС, реализация технического и смыслового аспектов совместного функционирования компонентов;
- известные методы и программные инструменты для встраивания имеющихся прикладных программ в новую среду (изменение языка программирования, интерфейсов с базами данных и др.).

Документирование процессов и результатов проектирования и перепроектирования как бизнес-процессов, так и компьютерных компонентов ИС.

Применяются:

- средства выдачи отчетов и справок CASE-систем и других специальных программ моделирования;
- развитые средства редакторов текстов и графики (может быть, с элементами анимации или мультимедиа) для создания качественной документации на бизнес-условия, процедуры и процессы;
- включение актуальных документов в контуры корпоративной сети, программы обучения, контекстную помощь и т. п.

Создание внешней документации программ производства и поставок товаров и услуг основной деятельности предприятия на конкурентно высоком уровне.

Формируются выходные потоки информации, направленные на клиентов, бизнес-партнеров, правительственные круги, широкую публику, для формирования которых используются:

- редакторы, системы компьютерной верстки, анимации и мультимедиа для создания интерактивных справочных приложений, видеодисков, каталогов, прейскурантов и др.;

- системы программирования объектов, обеспечивающих для получателя «удаленную интерпретацию» содержания интерактивных справочных приложений, видеодисков, каталогов, прейскурантов и др.;
- программирование WWW-серверов, другие возможности информационной супермагистрали для размещений своей внешней документации основной деятельности.

Обеспечение оперативной обратной связи с возможными потребителями, коммерческими клиентами, бизнес-партнерами и др.

Применяются методы и системы маркетингового мониторинга и анализа с получением первичной и вторичной информации. Используются методы и средства ИТ для:

- создания приложений, обеспечивающих обратную связь с клиентами и потребителями через системы глобальной сети;
- обеспечения круглосуточного функционирования ИС предприятия с целью информирования, приема и выполнения заявок и претензий клиентов;
- администрирование оперативных БД с осуществлением непрерывного функционирования OLTP.

3.5. Автоматизация проектирования АИС (CASE-технологии)

Одним из центральных средств нового системного проектирования, как это указано выше, являются *средства автоматизации проектирования*.

Перечисленные факторы способствовали появлению программно-технологических средств специального класса — CASE-средств, реализующих CASE-технологию создания и сопровождения ИС. Термин CASE (Computer Aided Software Engineering) используется в настоящее время в весьма широком смысле. Первоначальное значение термина CASE, ограниченное вопросами автоматизации разработки только лишь программного обеспечения (ПО), в настоящее время приобрело новый смысл, охватывающий процесс разработки сложных ИС в целом. Теперь под термином CASE-средства понимаются программные средства, поддерживающие процессы создания и сопровождения ИС, включая анализ и формулировку требований, проектирование прикладного ПО (приложений) и баз данных, генерацию кода, тестирование, документирование, обеспечение качества, конфигурационное управление и управление проектом, а также

другие процессы. CASE-средства вместе с системным ПО и техническими средствами образуют полную среду разработки ИС.

Появлению CASE-технологии и CASE-средств предшествовали исследования в области методологии программирования. Программирование обрело черты системного подхода с разработкой и внедрением языков высокого уровня, методов структурного и модульного программирования, языков проектирования и средств их поддержки, формальных и неформальных языков описаний системных требований и спецификаций и т. д.

CASE-технология представляет собой методологию проектирования ИС, а также набор инструментальных средств, позволяющих в наглядной форме моделировать предметную область, анализировать эту модель на всех этапах разработки и сопровождения ИС и разрабатывать приложения в соответствии с информационными потребностями пользователей. Большинство существующих CASE-средств основано на методологиях структурного (в основном) или объектно-ориентированного анализа и проектирования, использующих спецификации в виде диаграмм или текстов для описания внешних требований, связей между моделями системы, динамики ее поведения и архитектуры программных средств [1, 38].

ER-моделирование структуры предметной области

Успех любого приложения зависит от того, насколько хорошо смоделирована и разработана база данных, что ставит эту разработку в центр внимания.

База данных создается в несколько этапов, на каждом из которых необходимо согласовывать структуру данных с заказчиком и, что самое важное, подвергнуть созданную структуру данных экспертизе внутри группы, которая создает систему. Поэтому представление данных должно быть простым и понятным всем заинтересованным лицам. Именно по этой причине наибольшее распространение получило представление базы данных под названием «сущность—связь» (entity—relationship), которое также известно как ER-диаграмма. Модели, представленные в виде ER-диаграмм, крайне просты и удобны для понимания. Фрагмент такой модели изображен на рис. 3.6.

ER-диаграммы были приняты в качестве основы для создания стандарта IDEFIX. Предварительный вариант этого стандарта был разработан в военно-воздушных силах США и предназначался для увеличения производительности при разработке компьютерных систем. В 1981 г. этот стандарт был формализован и опубликован орга-

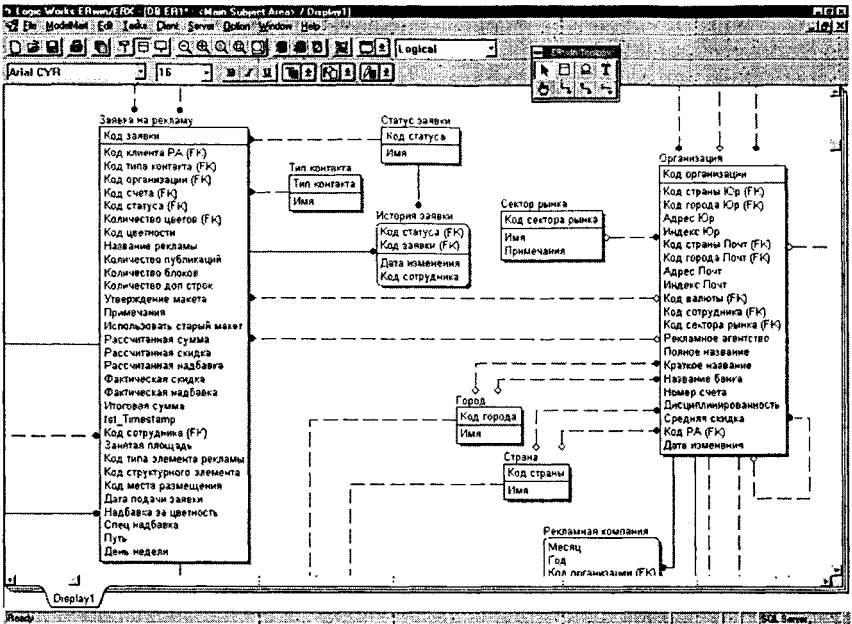


Рис. 3.6. Модель базы данных разрабатываемого приложения, представленная в виде ER-диаграммы, построенной в среде ERWIN

низацией ICAM (Integrated Computed Aided Manufacturing) и с тех пор является наиболее распространенным стандартом для создания моделей баз данных по всему миру.

С развитием компьютерных технологий и появлением CASE-моделирования возникла потребность в инструментах, которые бы поддерживали стандарты моделирования. Современный инструмент моделирования баз данных должен удовлетворять ряду требований.

1. Позволять разработчику сконцентрироваться на самом моделировании, а не на проблемах с графическим отображением диаграммы. Инструмент должен автоматически размещать сущности на диаграмме, иметь развитые и простые в управлении средства визуализации и создания представлений модели.

2. Инструмент должен проверять диаграмму на согласованность, автоматически определяя и разрешая несоответствия. Кроме того, он должен быть настраиваемым и при желании предоставлять разработчику некоторую свободу в действиях и право самому разрешать несоответствия или отступления от методологии.

3. Инструмент моделирования должен поддерживать как логическое, так и физическое моделирование.

4. Современный инструмент должен автоматически генерировать базу данных на СУБД назначения.

Рассмотрим вкратце средство моделирования баз данных ERWIN версии 3.5, продукт компании Logic Works [23], точнее, ERWIN 3.5/ERX, который предназначен для работы именно с системами управления базами данных. Этот инструмент моделирования полностью поддерживает стандарт IDEFIX и является лидером на рынке инструментов разработки баз данных.

Разработка в среде ERWIN

Обычно разработка модели базы данных состоит из двух этапов: составление логической и создание на ее основе физической модели. ERWIN полностью поддерживает такой процесс, он имеет два уровня представления модели: *логический* (logical) и *физический* (physical). Таким образом, разработчик может строить логическую модель базы данных, не задумываясь над деталями физической реализации, т. е. уделяя основное внимание требованиям к информации и бизнес-процессам, которые будет поддерживать будущая база данных. ERWIN имеет очень удобный пользовательский интерфейс, позволяющий представить базу данных в самых различных аспектах. Например, ERWIN содержит такие средства визуализации, как *храняемое отображение* (stored display) и *предметная область* (subject area). Хранимые отображения позволяют иметь несколько вариантов представления, в каждом из которых могут быть подчеркнуты определенные детали, которые вызвали бы перегрузку модели, будучи размещенными на одном представлении. Предметные области помогают вычлнить из сложной и трудной для восприятия модели отдельные фрагменты, которые относятся лишь к определенной области, из числа тех, которые ею охвачены. Интерфейс среды разработки ERWIN представлен на рис. 3.6.

ERWIN является не только инструментом для дизайнера баз данных, он также поддерживает автоматическую генерацию спроектированной и определенной на физическом уровне структуры данных. ERWIN 3.5 поддерживает ряд клиент-серверных и локальных СУБД. В этот список входят такие продукты, как Microsoft SQL Server, Oracle, Sybase, DB2, INFORMIX, Red Brick, Teradata, PROGRESS, Microsoft Access, FoxPro, Clipper и многие другие. Для каждой из перечисленных СУБД в ERWIN предусмотрено присо-

единение по протоколу этой СУБД и поддержка всех присущих ей средств управления данными.

ERWIN не поддерживает моделирования механизмов защиты базы данных, однако с помощью макроязыка можно автоматически выдать права на объект, пользуясь языком определения прав, который применяется в конкретной СУБД.

ERWIN имеет средство, выполняющее процесс, обратный генерации или *обратную разработку* (reverse engineering), т. е. можно осуществить присоединение к БД и, получив информацию о структуре, отобразить ее в графическом интерфейсе, сохранив все сущности, связи и атрибуты. Таким образом, можно переносить существующую структуру данных с одной платформы на другую, а также исследовать конкретные БД.

В состав ERWIN входит также средство интерактивной разработки Complete-Compare. ERWIN демонстрирует разногласия между моделью и базой данных, эти несоответствия можно переносить или оставлять без изменений. С помощью этого средства можно все изменения модели вносить в базу данных автоматически без необходимости контроля за соответствием модели и базы данных «вручную», при этом существующие данные не будут затронуты.

ERWIN тесно интегрирован с другими продуктами Logic Works. Словарь данных, созданный при анализе бизнес-процессов с помощью инструмента BPwin, может быть использован как основа для построения модели базы данных. Однако взаимосвязь между этими двумя инструментами двусторонняя, модели BPwin и ERWIN можно постоянно поддерживать в согласованном состоянии. Интеграция этих двух продуктов очень важна с точки зрения их совместного использования при разработке программного обеспечения, так как отпадает необходимость в повторном выполнении действий и процесс создания словаря данных становится практически автоматическим.

Проектирование модели в ERWIN наглядно представляет ограничения ссылочной целостности в независимом от СУБД виде. В то же время для выбранной целевой СУБД ERWIN автоматически генерирует нужные элементы ссылочной целостности — внешние и альтернативные ключи, триггеры, ограничения.

Моделирование в ERWIN

Процесс построения информационной модели состоит из следующих шагов:

- определение сущностей;
- определение зависимостей между сущностями;

- задание первичных и альтернативных ключей;
- определение атрибутов сущностей;
- приведение модели к требуемому уровню нормальной формы;
- переход к физическому описанию модели: назначение соответствий имя сущности — имя таблицы, атрибут сущности — атрибут таблицы;
- задание триггеров, процедур и ограничений;
- генерация базы данных.

ERWIN создает визуальное представление (модель данных) для решаемой задачи. Это представление может использоваться для детального анализа, уточнения и распространения как части документации, необходимой в цикле разработки. Однако ERWIN далеко не только инструмент для рисования. ERWIN автоматически создает базу данных (таблицы, индексы, хранимые процедуры, триггеры для обеспечения ссылочной целостности и другие объекты, необходимые для управления данными).

Как это отмечалось выше, в ERWIN поддерживаются два уровня представления и моделирования — логический и физический. Логический уровень означает прямое отображение фактов из реальной жизни. Например, люди, столы, отделы, собаки и компьютеры являются реальными объектами. Они именуется на естественном языке, с любыми разделителями слов (пробелы, запятое и т. д.). На логическом уровне не рассматривается использование конкретной СУБД, не определяются типы данных (например, целое или вещественное число) и не определяются индексы для таблиц.

Целевая СУБД, имена объектов и типы данных, индексы составляют второй (физический) уровень модели ERWIN. ERWIN предоставляет возможности создавать и управлять этими двумя различными уровнями представления одной диаграммы (модели), равно как и иметь много вариантов отображения на каждом уровне.

Диаграмма ERWIN строится из трех основных блоков — сущностей, атрибутов и связей. Если рассматривать диаграмму как графическое представление правил предметной области, то сущности являются существительными, а связи — глаголами.

Выбор между логическим и физическим уровнем отображения осуществляется через линейку инструментов или меню. Внутри каждого из этих уровней есть следующие режимы отображения:

- *сущности* — внутри прямоугольников отображается имя сущности (для логической модели) или имя таблицы (для физического представления модели); служит для удобства обзора большой диаграммы или размещения прямоугольников сущностей на диаграмме;

- *определение сущности* служит для подробного представления (презентации диаграммы);
- *атрибуты* — при переходе от предметной области к модели требуется вводить информацию о том, что составляет сущность. Эта информация вводится путем задания атрибутов (на физическом уровне — колонок таблиц). В этом режиме прямоугольник-сущность делится линией на две части — в верхней части отображаются атрибуты (колонки), составляющие первичный ключ, а в нижней — остальные атрибуты. Этот режим является основным при проектировании на логическом и физическом уровнях;
- *первичные ключи* — внутри прямоугольников (сущностей) показываются только атрибуты/колонки, составляющие первичный ключ;
- *пиктограммы* — каждой таблице может быть поставлена в соответствие пиктограмма (bitmap);
- *показ глагольной фразы* — на дугах связей показываются глагольные фразы, связывающие сущности (для логического уровня) или имена внешних ключей (для физического уровня).

Диаграмма может занимать более чем один экран и более чем один лист при печати. Для обзора модели предусмотрены, кроме прокруток экрана, режимы уменьшения/увеличения изображения, отображение всей модели или выделенной ее части.

Все объекты модели ERWIN могут редактироваться средствами, принятыми в Windows — группировка, копирование, удаление, перемещение, использование системного буфера. Установка цветов и шрифтов осуществляется в удобных диалогах.

Компоненты модели, представленные текстом (имена сущностей, атрибутов, текстовые элементы) могут редактироваться непосредственно на экране.

Идентификация и редактирование сущностей и связей в ERWIN

На диаграмме сущность изображается прямоугольником. В зависимости от режима представления диаграммы прямоугольник может содержать имя сущности, ее описание, список атрибутов и другие сведения.

Горизонтальная линия прямоугольника разделяет атрибуты сущности на два набора — атрибуты, составляющие первичный ключ (в верхней части), и прочие — в нижней части.

Сущность представляет собой тип реальных или абстрактных объектов и ей соответствует таблица в реальной СУБД. В ERWIN сущность визуально представляется тремя основными видами информации:

- атрибуты, составляющие первичный ключ;
- неключевые атрибуты;
- тип сущности (независимая/зависимая).

Первичный ключ — это атрибут или набор атрибутов, уникально идентифицирующий экземпляр сущности. Если несколько наборов атрибутов могут уникально идентифицировать сущность, то выбор одного из них осуществляется разработчиком на основании анализа предметной области.

Для каждого первичного ключа ERWIN создает при генерации структуры БД уникальный индекс.

Экземпляры *независимой сущности* могут быть однозначно идентифицированы без определения ее связей с другими сущностями. *Зависимая сущность*, наоборот, не может быть уникально идентифицирована без определения связей с другими сущностями. Зависимая сущность отображается в ERWIN прямоугольником с закругленными углами.

Альтернативный ключ — это атрибут (или группа атрибутов), не совпадающий с первичным ключом и уникально идентифицирующий экземпляр сущности. Например, для сущности служащий (идентификатор служащего, фамилия, имя, отчество) группа атрибутов «фамилия», «имя», «отчество» может являться альтернативным ключом (в предположении, что на предприятии не работают полные тезки).

Для альтернативного ключа, как и для первичного, ERWIN автоматически создает индексы при генерации БД.

Инвертированные индексы. Атрибуты, составляющие альтернативный ключ, однозначно (уникально) идентифицируют экземпляры сущности. В ERWIN также можно составлять группы атрибутов, которые не идентифицируют уникально экземпляры сущности, но часто используются для доступа к данным. Для каждой такой группы атрибутов ERWIN создает неуникальные индексы.

Одни и те же атрибуты сущности могут входить в несколько различных групп ключей.

Унификация атрибутов. Зависимая сущность может наследовать один и тот же внешний ключ от родительских сущностей, как непосредственно, так и через несколько связей. Если не введены различные роли для такого множественного наследования, ERWIN счита-

ет, что в зависимой сущности атрибуты внешнего ключа появляются только один раз.

Унификация — это объединение двух или более групп атрибутов внешних ключей в один внешний ключ (группу атрибутов), в предположении, что значения одноименных атрибутов в дочерней сущности всегда одинаковы.

Рассмотрим пример: сущность «сотрудник» имеет первичный ключ «код сотрудника» и связан идентифицирующей связью с сущностями «супруга» и «дети». При этом происходит миграция первичного ключа в зависимые сущности. В свою очередь, сущность «супруга» связана неидентифицирующей связью с сущностью «дети». Имеются два пути миграции ключа, однако в сущности «дети» атрибут «код сотрудника» появляется один раз в качестве элемента первичного ключа.

Существуют случаи, когда унификация атрибутов дает неверный с точки зрения предметной области результат. Для отмены унификации для атрибутов вводятся имена ролей.

Связь — функциональная зависимость между двумя сущностями (в частности, возможна связь сущности с собой самой). Например, важно знать фамилию сотрудника, также, в каком отделе он работает. Таким образом, между сущностями «отдел» и «сотрудник» существует связь «состоит из» (отдел состоит из сотрудников). Связь — это понятие логического уровня, которому соответствует внешний ключ на физическом уровне. В ERWIN связи представлены пятью основными элементами информации:

- тип связи (идентифицирующая, неидентифицирующая, полная/неполная категория, неспецифическая связь);
- родительская сущность;
- дочерняя (зависимая) сущность;
- мощность связи (cardinality);
- допустимость пустых (null) значений.

Связь называется *идентифицирующей*, если экземпляр *дочерней сущности* идентифицируется через связь с *родительской*.

Атрибуты, составляющие первичный ключ родительской сущности, при этом входят в первичный ключ дочерней сущности. Дочерняя сущность при идентифицирующей связи всегда является зависимой.

Связь называется *неидентифицирующей*, если экземпляр дочерней сущности идентифицируется иначе, чем через связь с родительской сущностью. Атрибуты, составляющие первичный ключ родительской сущности, при этом входят в состав неключевых атрибутов дочерней сущности.

Для определения связей в ERWIN выбирается ее тип, затем отмечаются родительская и дочерняя сущность. Идентифицирующая связь изображается сплошной, а неидентифицирующая — пунктирной линией. Линии заканчиваются точкой со стороны дочерней сущности.

При определении связи происходит миграция атрибутов первичного ключа родительской сущности в соответствующую область атрибутов дочерней сущности. Поэтому такие атрибуты не вводятся вручную.

Атрибуты первичного ключа родительской сущности по умолчанию мигрируют со своими именами. ERWIN позволяет ввести для них роли, т. е. новые имена, под которыми мигрирующие атрибуты будут представлены в дочерней сущности. В случае неоднократной миграции атрибута такое переименование необходимо. Например, сущность «посредническая сделка» имеет атрибут «код предприятия-продавца» и «код предприятия-покупателя». В данном случае первичный ключ сущности «предприятие» («код предприятия») имеет две роли в дочерней сущности.

На физическом уровне имя роли — это имя колонки внешнего ключа в дочерней таблице.

Мощность связи представляет собой отношение количества экземпляров родительской сущности к соответствующему количеству экземпляров дочерней сущности. Для любой связи, кроме неспецифической, эта связь записывается как $1 : n$.

ERWIN в соответствии с методологией IDEFIX предоставляет четыре варианта для n , которые изображаются дополнительным символом у дочерней сущности: ноль, один или больше (по умолчанию); ноль или один; ровно N , где N — конкретное число.

Допустимость пустых значений в неидентифицирующих связях ERWIN изображает пустым ромбиком на дуге связи со стороны родительской сущности.

Имя связи на логическом уровне представляет собой «глагол», связывающий сущности. Физическое имя связи (которое может отличаться от логического) для ERWIN означает имя ограничения (constraint) или индекса.

Связи категоризации. Некоторые сущности определяют целую категорию объектов одного типа. В ERWIN в таком случае создается сущность для определения категории и для каждого элемента категории, а затем вводится для них связь категоризации. Родительская сущность категории называется *супертипом*, а дочерние — *подтипом*.

Например, сущность «сотрудник» может содержать данные как о штатных работниках, так и о временно нанятых. Первые и вторые имеют различающиеся, частично пересекающиеся наборы атрибутов (минимальное пересечение подтипов составляет первичный ключ).

Общая часть этих атрибутов, включая первичный ключ, помещается в сущность-супертип «сотрудник».

Различающаяся часть (например, данные почасовой оплаты для временных работников и данные о зарплате и отпуске для штатных работников) помещается в сущности-подтипы.

В сущности-супертипе вводится атрибут-дискриминатор, позволяющий различать конкретные экземпляры сущности-подтипа.

В зависимости от того, все ли возможные сущности-подтипы включены в модель, категорийная связь является полной или неполной.

Продолжая пример, отметим, что если супертип может содержать данные об уволенных сотрудниках, то эта связь — неполной категоризации, так как для него не существует записи в сущностях-подтипах.

В ERWIN полная категория изображается окружностью с двумя подчеркиваниями, а неполная — окружностью с одним подчеркиванием.

Реализация ссылочной целостности с помощью ERWIN. *Ссылочная целостность* — это обеспечение требования, чтобы значения внешнего ключа экземпляра дочерней сущности соответствовали значениям первичного ключа в родительской сущности. Ссылочная целостность может контролироваться при всех операциях, изменяющих данные (*insert, update, delete*). Средства контроля ссылочной целостности в ERWIN включают автоматическую генерацию триггеров и использование механизмов декларативной ссылочной целостности (для тех СУБД, которые поддерживают данные механизмы).

Для каждой связи на логическом уровне могут быть заданы требования по обработке операций *insert/update/delete* для родительской и дочерней сущности. ERWIN представляет следующие варианты обработки этих событий:

- отсутствие проверки;
- проверка допустимости;
- запрет операции;
- каскадное выполнение операции (*delete/update*);
- установка пустого (*null*) или заданного значения по умолчанию.

Обратное проектирование (Reverse engineering). Обратное проектирование, т. е. восстановление информационной модели по существующей базе данных, используется при выборе оптимальной платформы (rightsizing) для существующей настольной (desktop) базы данных или базы данных на mainframe, а также при расширении (или модификации) существующей структуры, которая была построена без необходимой сопроводительной документации. После завершения процесса восстановления модели ERWIN автоматически «раскладывает» таблицы на диаграмме. Теперь можно выполнять модификации уже с использованием логической схемы — добавлять сущности, атрибуты, комментарии, связи и т. д. По завершении изменений одна команда — синхронизировать модель с базой данных — актуализирует все проведенные изменения.

Построение модели может быть выполнено как на основании данных каталога базы данных, так и на основании пакета операторов SQL, с помощью которого была создана база данных.

Моделирование играет большую роль в разработке успешных информационных систем. Использование ERWIN поможет правильно оценить стоящие задачи, предложить адекватное решение (анализ бизнес-процессов, BPwin) и разработать центральную часть любой информационной системы — базы данных — с использованием информации, полученной во время обследования предприятия (моделирование базы данных, ERWIN). Эти инструменты сами по себе не являются решением проблемы, но их грамотное и своевременное использование поможет свести рутинный труд разработчика к минимуму, позволит ему сконцентрироваться на собственно разработке системы и снизит потери времени, которые обычно происходят при согласовании моделей со специалистами предметной области. Кроме того, использование этих инструментов дает возможность получить набор полностью документированных и согласованных моделей, что в значительной степени облегчит поддержку созданных систем в будущем, а также может быть повторно использовано при разработке других систем.

3.6. Методы и модели оценки и измерения эффективности АИС

Обобщенной целью оценки систем информационного поиска является выявление полезности системы на практике. Оцениваемую систему можно рассматривать как инновацию (т. е. нововведение),

степень общественного признания которой определяется рядом признаков:

- *относительное преимущество* — степень, в которой инновация воспринимается лучше предшествующей идеи с точки зрения экономической прибыльности, социального престижа, физического удобства и психологического удовлетворения;
- *совместимость* — степень, в которой инновация воспринимается как наследующая существующие социально-культурные ценности, убеждения и нормы, прошлый опыт, зарекомендовавшие себя идеи, приемы и оборудование, а также осознанные потребности потенциальных пользователей;
- *сложность* — степень, в которой инновация воспринимается как относительно трудная для понимания и применения с точки зрения существующих знаний и навыков;
- *возможность тестирования* — степень, в которой инновацию можно предварительно протестировать в ограниченном масштабе, таким образом снижая риск и неопределенность, связанные с внедрением;
- *возможность наблюдения* — степень, в которой инновация и эффект от ее применения наглядно видны для других.

Стандарты оценки качества информационных систем

Стандарт РФ ГОСТ Р ИСО/МЭК 9126—93. В отечественной практике разработки и сопровождения автоматизированных программных систем основным регламентирующим документом является ГОСТ Р ИСО/МЭК 9126—93 [11].

Стандарт определяет шесть характеристик, которые с минимальным дублированием описывают качество программного обеспечения. Данные характеристики образуют основу для дальнейшего уточнения и описания качества программного обеспечения и базируются на требованиях ISO 8402 (см. ниже).

Определения характеристик и соответствующая модель процесса оценки качества, приведенные в указанном стандарте, используются тогда, когда известны требования к программной продукции и оценивается ее качество в процессе жизненного цикла. Эти характеристики могут применяться к любому виду программного обеспечения, включая программы для ЭВМ и данные, входящие в программно-технические средства (встроенные программы).

Рассмотрим вкратце указанные характеристики качества программного обеспечения.

1. *Функциональные возможности (Functionality)* — набор атрибутов, относящийся к набору функций и их конкретным свойствам. Функциями являются те, которые реализуют установленные или предполагаемые потребности. Данные признаки задают то, что программное обеспечение выполняет для удовлетворения потребностей, тогда как другие описывают в основном, когда и как это осуществляется.

2. *Надежность (Reliability)* — набор атрибутов, относящийся к способности ПО сохранять свой уровень качества функционирования при установленных условиях за установленный период времени. Износ или старение программного обеспечения не происходит. Ограничения надежности возникают из-за ошибок в требованиях, проекте и реализации.

3. *Практичность (Usability)* — набор признаков, относящихся к объему работ, требуемых для использования и индивидуальной оценки такого использования определенным или предполагаемым кругом пользователей. «Пользователи» интерпретируются как большинство непосредственных пользователей интерактивного программного обеспечения. Круг пользователей может включать операторов, конечных пользователей и косвенных пользователей, на которых влияет данное ПО или которые зависят от его использования. Практичность должна рассматриваться во всем разнообразии условий эксплуатации пользователем, которые могут влиять на программное обеспечение, включая подготовку к использованию и оценку результатов.

4. *Эффективность (Efficiency)* — атрибуты, характеризующие соотношение между уровнем качества функционирования программного обеспечения и объемом используемых ресурсов при установленных условиях. Ресурсы могут включать другие программные продукты, технические средства, материалы (бумага, гибкие диски и пр.), услуги эксплуатирующего, сопровождающего или обслуживающего персонала, а также время, расходуемое на решение задач.

5. *Сопровождаемость (Maintainability)* — признаки, относящиеся к объему работ, требуемых для проведения конкретных изменений (модификаций). Изменения могут включать исправления, усовершенствование или адаптацию к окружающей обстановке, требованиям и условиям функционирования.

6. *Мобильность (Portability)* — атрибуты, характеризующие способность ПО быть перенесенным из одного окружения в другое. Окружение может включать организационное, техническое, программное, информационное окружения.

Процесс (модель) оценивания качества ПО состоит из трех стадий (табл. 3.4) и может применяться в любой подходящей фазе жизненного цикла и для каждой компоненты программного продукта.

Таблица 3.4. Структура модели процесса оценивания ПО

№	Стадия	Этапы	Результаты
1	Установление требований к качеству	—	Установление требований в терминах характеристик качества и возможных комплексных показателей. Требования к продукции в целом могут отличаться от требований к компонентам
2	Подготовка к оцениванию	Выбор метрик (показателей качества)	Каждый количественный признак и каждое количественно оцениваемое взаимодействие ПО с его окружением могут быть приняты в качестве метрики
		Определение уровней ранжирования	Для каждого диапазона количественных значений показателя устанавливается уровень ранжирования (например – «отличный», «хороший», «средний», «низкий» и пр.)
		Определение критериев оценки	Приведение результатов оценивания различных характеристик к некоему общему итогу. Оценщик должен подготовить процедуры, используя, например, таблицы решений
3	Процедура оценивания	Измерение	Выбранные метрики применяются к ПО. Результатом являются значения в масштабе метрики
		Ранжирование	Для всех измеренных значений устанавливается уровень
		Оценка	Заключение о качестве ПО. Обобщенное качество сравнивается с такими факторами, как время или стоимость. Результатом является решение руководства по приемке, отбраковке, выпуску или невыпуску продукта

Зарубежные стандарты. В настоящее время для оценки эффективности, качества и производительности информационных технологий существуют зарубежные системы стандартов, такие, как ESA, ISO 9000 или Baldrige Award, IEC TC (International Electrotechnical Commission Technical Committee 65C) и некоторые другие.

Стандарты контроля и качества IEC TC включают следующие составляющие:

- EIAMUG — оценка параметров функционирования и управления системами;
- CCS — стандарты обработки данных в сетях;
- SMM — стандарты контроля и управления;
- FCS — стандарты оценки функциональных возможностей.

В рамках стандартов IEC TC выполняются следующие работы:

- разрабатываются новые концепции контроля и управления автоматизированными системами;

- новые концепции и методы внедряются в промышленность;
- новые концепции и методы широко используются в сфере образования.

Направления стандартизации ИЕС ТС:

- направления, посредством которых осуществляется процесс эффективного управления;
- методы разработки требований качества, управления и контроля;
- механизмы принятия эффективного решения в области управления автоматизированными промышленными, компьютерными системами или иными системами автоматизации документооборота;
- механизмы создания требований по эффективному управлению автоматизированных систем в рамках их жизненного цикла.

В рамках указанных стандартов создаются стандартные тесты для оценки эффективности специальных информационных систем, например, стандартный тест PEMS — оценка качества и эффективности информационных технологий в медицине по системе ИЕС ТС 65С.

Методы осуществления стандартизации в рамках ИЕС ТС:

- разработка моделей управления и контроля качества;
- использование автоматизации в процессе управления и принятия решений;
- осуществление контроля и управления по задачам;
- создание инструментов для выполнения требований.

Методика разработки стандартов ИЕС ТС основана на большом опыте и анализе широкого спектра требований и методов работы различных организаций и фирм в разных странах Европы.

Стандарты Качества ISO. Организация международных стандартов ISO была создана в 1947 г., в настоящее время ее членами являются около 100 стран. Выполнение технической работы в ISO возложено на более чем 2700 технических комитетов, в состав которых входят представители правительственных, промышленных и научно-исследовательских кругов (около 500 организаций).

Стандарт ISO очень популярен в Европе. Сегодня стандартами охвачены многие технологические отрасли — от программирования и телекоммуникаций до банковской и финансовой сфер.

Создание стандартов проводится в соответствии с тремя принципами.

Во-первых, они являются результатом консенсуса всех заинтересованных сторон-производителей: поставщиков, потребителей, про-

фессиональных разработчиков, правительственных и исследовательских организаций.

Во-вторых, стандарты имеют действительно мировое распространение и удовлетворяют как производителей, так и потребителей.

В-третьих, появление новых стандартов диктуется исключительно требованиями свободного рынка.

Обычно инициаторами разработки стандарта являются производители, далее соответствующая рабочая группа определяет техническую область, на которую предполагаемый стандарт будет распространяться. На следующем этапе идет выработка технических спецификаций, первая версия стандарта утверждается (за стандарт должно проголосовать 75 % кворума) и публикуется. С этого момента стандарт становится официальным (ISO International Standard).

По мере совершенствования технологий, появления новых материалов, методов обработки, повышения требований к качеству и надежности изделий возникает необходимость в пересмотре стандартов. В ISO существует правило: все стандарты должны пересматриваться не реже одного раза в пять лет. Сегодня ISO принадлежит более 9300 различных стандартов, описание которых занимает около 180 тыс. страниц текста на английском языке. ISOplus выполняет роль экспертной системы по поддержке стандартов.

Некоторые стандарты ISO:

- ISO 8402. Управление качеством и гарантии качества;
- ISO 9001. Системная модель качества для процессов проектирования, разработки, производства, установки и обслуживания;
- ISO 9002. Системная модель качества для процессов проверки качества проектирования, установки и обслуживания;
- ISO 9003. Системная модель качества для процессов проверки качества при окончательном тестировании;
- ISO 10011-1, ISO 10011-2, ISO 10011-3. Руководство по аудиту качества систем;
- ISO 10012-1. Требования по качеству, предъявляемые к измерительной аппаратуре;
- ISO 10013. Руководство по созданию качественной документации.

Однако качество конечного продукта, требования рынка и технологический уровень самих производственных процессов оказываются вне интересов ISO.

Стандарт качества *Baldrige Award*. Объединение международных рынков, повышение требований к качеству и жесткая конкуренция привели к появлению параллельного стандарта качества *Malcolm Baldrige National Quality Award* (кратко *Baldrige*

Award), весьма популярного в США. Главная цель Baldrige Award — способствовать организациям в создании конкурентоспособных продуктов.

Baldrige Award ориентирована на виды деятельности, способствующие повышению конкурентоспособности компаний и предусматривающие для достижения этого различные способы: за счет обращения лицом к рынку и клиентам, нацеленности на конечный результат, постоянного совершенствования деловых процессов, тесной привязки к общей стратегии бизнеса, интеграции процессов на основе анализа, повышения квалификации персонала, расширения информационных связей.

Каждый год Baldrige Award проводит награждение компаний, добившихся выдающихся успехов в области качества, оценка строится по 28 критериям, разбитым на семь категорий: лидерство на рынке, информация и анализ, стратегия планирования качества, управление персоналом, управление качеством, достигнутые качественные результаты, мнения клиентов.

Стандартные тесты измерения производительности. Оценка производительности и средства ее измерения (контрольно-оценочные тесты) привлекают особое внимание как изготовителей, так и пользователей вычислительных систем. Более существенной представляется проблема адекватной интерпретации многочисленных показателей производительности, полученных на основе разнообразных методик, число которых приближается к сотне.

Как правило, на практике приходится решать три проблемы, связанные с анализом результатов контрольного тестирования производительности:

1) отделение показателей, которым можно доверять безоговорочно, от тех, что должны восприниматься с известной долей настороженности (проблема достоверности оценок);

2) выбор контрольно-оценочных тестов, наиболее точно характеризующих производительность при обработке типовых задач пользователя (проблема адекватности оценок);

3) правильное истолкование результатов тестирования производительности (проблема интерпретации).

Контрольно-оценочные тесты классифицируются, как:

- *тесты производителей*, которые разрабатываются фирмами-изготовителями компьютеров для «внутреннего» применения — оценки качества собственных продуктов;
- *стандартные тесты*, разработанные для сравнения широкого спектра компьютеров и часто претендующие на роль полностью универсальных средств измерения производительности;

- *пользовательские тесты*, подготовленные крупными фирмами, специализирующимися на внедрении компьютерных технологий, или совместными усилиями групп пользователей, объединенных сходством решаемых задач. Эти средства предназначены специально для выбора компьютеров и ПО, наиболее подходящих под определенные прикладные задачи. Такой подход позволяет получить наиболее точные оценки производительности для конкретных информационных систем.

Основными характеристиками производительности вычислительного комплекса являются:

- *пиковая производительность* — теоретический максимум быстрой реакции при самых идеальных условиях;
- *натуральная производительность*, определяемая отношением числа операций на выполнение задачи к длительности выполнения соответствующей программы. Эти факторы позволяют с помощью контрольно-оценочных тестов определить минимум ресурсов системы для выполнения задачи.

Для оценки производительности вычислительного комплекса используются широко известные стандарты:

- **Wenstone** (1976 г.) и **Dhrystone** (1984 г.) — оценка пропускной способности системы. Комплексы тестов состоят из нескольких модулей, имитирующих нагрузку компьютерной системы в наиболее типичных режимах выполнения задач. Каждый модуль выполняется многократно, и в соответствии с исходной статистикой оценивается производительность;
- **Linpac** (1976 г.) — определение средней и предельной производительности системы. Тест имеет особую значимость при использовании компьютеров с векторной архитектурой и параллельной обработкой, т. е. дает характеристики глубины программного параллелизма вычислительной системы, определяет максимальные размеры обрабатываемых матриц и максимальную точность задач;
- **Perfect** — оценка производительности суперкомпьютеров, Unix-систем и рабочих станций. Осуществляет оптимизацию вычислительных и управляющих комплексов программ, обрабатывающих большие массивы информации, позволяет увеличить скорость обработки, выполняется дважды — до и после оптимизации программного комплекса;
- **TPC** (1988 г., основан Комитетом по тестам производительности в составе компаний IBM, Control Data, Hewlett-Packard) — оценка производительности систем, а также характеристик

стоимости приобретения и эксплуатации в течение 5 лет (цикл морального устаревания) компьютерного, серверного, периферийного оборудования и всего комплекса программного обеспечения. За основу взят стандарт Debit-Credit (1973 г., Bank of America).

Пакет ТРС-А позволяет дать оценку быстродействия вычислительной системы не только в локальной конфигурации, но и при работе в глобальной сети.

Пакет ТРС-В позволяет оценить возможности СУБД в условиях интенсивной нагрузки и при системном администрировании.

Пакет ТРС-С оценивает эффективность системы продажи и дистрибуции товаров, распределение продукции между торговыми домами и регионами;

- SPEC SFS (1988 г., основатели — компании Appollo Computers, Hewlett-Packard, Mips Computer Systems и Sun Microsystems) — определение характеристик работы системы при наращивании ее функциональных и производственных возможностей. Определяется производительность вычислительных систем с быстродействующими компьютерами в системе с мультипрограммной обработкой и сравнивается с производительностью системы VAX 11/780. Результат — число сценариев, выполненных за один час.

Стандарты оценки качества программного обеспечения. Фирма Software Engineering Institute (SEI) предложила концепцию «Улучшение процессов создания ПО» (Software Process Improvement — SPI), которая опиралась на статистические методы контроля технологических процессов, разработанные в Японии в конце 30-х гг. Позднее появились другие концепции:

- «Сквозной контроль качества» (Total Quality Management — TQM);
- «Реинжиниринг бизнес-процессов» (Business Process Reengineering — BPR), предполагающий модернизацию базовых бизнес-процессов организации;
- «Постепенное совершенствование деловых процессов» (Business Process Improvement — BPI).

В основе всех этих концепций лежит общее понимание жизненного цикла программного обеспечения как совокупности фаз, которые проходит программный продукт в процессе своего развития:

- выработка исходных требований к ПО со стороны пользователя;
- формулирование общих требований к ПО со стороны разработчика;

- проектирование архитектуры;
- детальная реализация ПО;
- инсталляция ПО в организации заказчика;
- эксплуатация.

Для поддержания жизненного цикла ПО фирмы-разработчики организуют свою деятельность по нескольким ключевым направлениям: управление проектом, тестирование и проверка функционирования, конфигурационный менеджмент, сопровождение.

Методология СММ. Университет Карнеги-Меллона под эгидой Министерства обороны США в 1987 г. разработал специальную систему оценки технологических процессов в организациях-разработчиках программного обеспечения. Предложенная ими *модель уровней зрелости* (Capability Maturity Model — СММ) основана на так называемых уровнях зрелости, среди которых можно выделить:

1) *начальный*, на котором находится большинство фирм-разработчиков. Процесс разработки носит неструктурированный и случайный характер, коммерческий успех определяется скорее выдающимися способностями какого-нибудь талантливого разработчика или менеджера, нежели организационной инфраструктурой фирмы, в которой отсутствует стабильная среда разработки и сопровождения;

2) *повторяемый*. Процесс создания программного обеспечения становится возможным благодаря жесткому управлению, планированию, контролю, выработке исходных требований и методам оценки в соответствии с определенными стандартами на разработку ПО;

3) *фиксированный*, на котором процессы управления и разработки полностью документированы, стандартизованы и интегрированы в единый технологический поток, контролируемый управляющим персоналом;

4) *управляемый*. Качество процессов и готового продукта можно оценить количественно. Для контроля над процессами используются количественные показатели (метрики). Все процессы предсказуемы и укладываются в заранее определенные рамки;

5) *оптимизируемый*, на котором фирмы стремятся улучшить свою работу, руководствуясь количественными критериями качества.

Каждый уровень характеризует определенную степень качества выпускаемых изделий. Таким образом, чем выше уровень зрелости компании, тем выше ее статус и авторитет в компьютерных кругах и в глазах пользователей.

Модель оценки качества SCOPE*PROGERT. В состав SCOPE*PROGERT (разработчик — немецкий исследователь-

ский центр GMD) входит несколько компонентов, соответствующих различным сторонам деятельности по созданию программ:

- специфицирование процессов тестирования (ProcePT);
- инжиниринг моделей качества (Model Y7), суть которого в определении модели качества для данного проекта с использованием методов оценки качества, что подразумевает встраивание этой модели в существующие технологические и бизнес-процессы;
- измерение характеристик (SIW), представляется в виде модели, на которой проводятся количественные исследования характеристик системы;
- моделирование процессов разработки (SPM) — данный компонент предназначен для оценки качества процессов разработки, причем на основе количественных показателей.

Модель оценки качества Trilliurn. Модель Trilliurn, созданная в 1994 г. фирмами Bell Canada, Nothern Telecom и Bell-Nothern Research, предлагает способ оценки процессов выпуска продуктов в телекоммуникационной и информационной областях, учитывающий все аспекты жизненного цикла ПО.

Моделью Trilliurn охвачены следующие виды деятельности:

- управление качеством и проектирование бизнес-процессов;
- оценка технологической зрелости;
- создание, разработка и системное проектирование;
- совместное и надежное проектирование.

В основе способа классификации уровней зрелости в модели Trilliurn лежит фактор риска.

Модель оценки качества Cleanroom. Разработана для создания сложных программных комплексов Cleanroom Software Engineering. Она охватывает такие стороны практики, как реализация модели СММ, планирование и управление проектами, выработка спецификаций, профилактика ошибок, тестирование и сопровождение. Cleanroom — это совокупность административно-технологических процессов, позволяющих коллективам разработчиков планировать, измерять, специфицировать, проектировать, кодировать, тестировать и сертифицировать программные продукты.

Методология Cleanroom построена на трех концепциях: модульном принципе специфицирования и проектирования, математическом доказательстве правильности применяемых алгоритмов и использовании статистики по результатам тестирования, как основы для оценки надежности программ (сертификации). Спецификации Cleanroom дают полное и точное описание функций системы. Выработка спецификаций способствует более глубокому пониманию

требований, предъявляемых к конечному продукту, и его функций, а сами спецификации служат основой для тестирования, сертификации и дальнейшего развития системы.

Инструментом автоматизированного тестирования и оценки надежности ПО в методологии Cleanroom служит среда Cleanroom Certification Assistant, в основе которой лежит идея использования статистических результатов тестирования для подсчета надежности ПО математическими методами. Компонент — Cleanroom Certification Model — фиксирует результаты тестирования в виде показателей среднего времени наработку на отказ, которые и используются для вычисления метрик надежности.

Другие модели оценки качества. Система Process-Weaver (компании Cap Gemini Innovation) позволяет автоматизировать процесс разработки с использованием терминов взаимозависимых заданий и совокупности входных/выходных данных, необходимых для реализации конкретной системы.

Также представляет интерес программа AMItool, созданная в Европейском центре ядерных исследований (CERN) и позволяющая с помощью метрики AMI (Application of Metrics In Industry) дать количественную оценку состояния фирмы, а также предлагающая план улучшения технологических процессов.

Важная проблема, которую приходится решать во время работы над проектами, — это обмен информацией. Для этого используется система WIT (WWW Interactive Talk), позволяющая организовать дискуссию участников в сети Интернет.

Основным элементом системы WIT является *Дискуссионное окно (Discussion Area)*, соответствующее предметной области, в рамках которой ведется обсуждение. Внутри дискуссионного окна имеется набор *Тем (Topics)*, связанный с определенными аспектами данной дискуссии. Участники дискуссии выражают свою точку зрения в виде *Предложений (Proposals)*, которые передаются в WIT в форме сообщений.

На всех стадиях работы над проектом рождаются многочисленные документы (исходные требования, системные спецификации, исходные тексты программ, инструкции по эксплуатации и пр.), созданные различными специалистами. В крупных распределенных проектах, в которых задействованы большие коллективы разработчиков, обостряется проблема эффективного использования информационных материалов. Для ее решения была создана система LIGHT (Life cycle Global HyperText — глобальная гипертекстовая система для поддержки жизненного цикла ПО), которая, как и WIT, опирается на технологию WWW.

Таким образом, на зарубежном рынке информационных технологий существует достаточно много систем и стандартных тестов оценки эффективности информационных технологий.

В то же время на отечественном рынке информационных технологий международные стандарты оценки качества не вполне адаптированы. Очевидно это связано с тем, что стандартные тесты выбора информационных технологий должны учитывать особенности современного состояния отечественного рынка информационных технологий — быстро меняющиеся техническая база, программное обеспечение, существующие подходы к эксплуатации технического парка и программного обеспечения, насыщенность рынка информационных технологий, быстрое моральное устаревание, смена информационных систем и современное экономическое состояние.

Контрольные вопросы

1. Каковы основные этапы канонического проектирования АИС?
2. Содержание и результаты предпроектного обследования.
3. Содержание и результаты технорабочего проектирования.
4. Что такое модель жизненного цикла и каковы ее разновидности?
5. Перечислите основные понятия и определения, связанные с формализованным подходом к проектированию АИС.
6. Что такое RAD- и CASE-технологии?
7. Укажите основные факторы, определяющие выбор программных средств реализации АИС.
8. Укажите основные факторы, определяющие выбор технического комплекса АИС.
9. Укажите основные факторы, определяющие выбор структуры информационной базы АИС.
10. Что такое киберкорпорация?
11. Перечислите недостатки моделей жизненного цикла.
12. В чем, по Вашему мнению, заключается сущность BPR и на какие типы АИС этот подход распространяется?
13. Перечислите основные компоненты подхода «новое системное проектирование».
14. В чем состоит сущность CASE-технологий?
15. Опишите основные характеристики программного продукта ERWIN.
16. Что представляют собой сущности и связи?
17. Что такое «обратное проектирование» в ERWIN?
18. Охарактеризуйте основные требования к оценке качества систем согласно отечественным стандартам.
19. В чем заключаются основные требования к качеству АИС согласно зарубежным стандартам?

Глава 4

АВТОМАТИЗИРОВАННЫЕ ИС ПО ОБЛАСТЯМ ПРИМЕНЕНИЯ

В настоящей главе будут рассмотрены некоторые основные типы АИС, ориентированные на различные области применения. Как это будет видно из содержания материала, эти системы, хотя и характеризуются в целом схемами рис. 1.2, 1.19, 1.20 и др., все же обладают определенными видовыми особенностями.

4.1. Автоматизированные информационные системы по законодательству

Законодательство Российской Федерации в настоящее время быстро обновляется — ежедневно высшими органами государственной власти принимается несколько десятков нормативных актов, что составляет около 10 000 документов в год, с учетом ведомственных документов (распоряжения, инструкции и т. п.) цифра увеличивается до 90—100 тыс. При этом в широкой печати публикуется лишь около 10 % этого объема, а половина — с задержкой от двух недель и более.

Все перечисленное служит причиной растущего спроса на информационные и консультационные услуги в этой области. Реально в настоящее время используются три основных источника нормативной информации: печатные издания (периодика и различные сборники); услуги консультационных фирм; использование автоматизированных информационных систем по законодательству (АИСЗ).

Основной трудностью применения АИСЗ является то, что всякий юридический (в частности, судебный) процесс должен быть основан на печатных источниках информации и при упоминании документа возможны ссылки только на официальные издания, а любые формы документов на магнитных носителях юридической силы не имеют. Поэтому специалист пользуется ими как справочными,

но если принимается решение, которое влечет за собой финансовую или иную ответственность, необходимо использование официального источника.

Краткая историческая справка по АИСЗ

За рубежом базы по законодательству начали появляться со второй половины 60-х гг., вместе с развитием современных технологий и систем телекоммуникации. Это было обусловлено тем, что обычным библиотекам было уже не под силу собирать, хранить и систематизировать массивы нормативных материалов. Первоначально были созданы системы *индексного типа, или электронные каталоги*, которые давали возможность найти полные сведения о выходных данных документов — названии, номере, дате издания, принявшем органе и т. д. В Европе первой индексной системой компьютерного поиска юридической информации была бельгийская система «Credos», которая появилась в 1967 г. благодаря совместным усилиям университетов Бельгии и бельгийского союза адвокатов и нотариусов. Однако пользователь «Credos» не мог напрямую общаться с информационным банком, и для получения интересующих его данных он должен был сначала обратиться в специальное информационное бюро (срок ожидания ответа 2—8 дней). Позднее некоторые индексные системы стали работать в диалоговом режиме, предполагающем прямое общение удаленного пользователя с информационным банком. Примером может служить система «Finlex», созданная в 1982 г. министерством юстиции Финляндии.

Поскольку *индексные системы* не дают возможности ознакомиться с полным текстом документов, они в дальнейшем были замещены *полнотекстовыми*, позволяющими находить документ и работать с его текстом.

В наше время практически все экономически развитые страны имеют справочные АИСЗ. В США это «WRU», «LEXIS», «WESTLAW», «JURIS», «FLITE»; в Великобритании — «INFOLEX», «PRESTEL», «POLIS», «LEXIS»; «ENLEX»; в Бельгии — «Credos»; в Германии — Система Бундестага, «JURIS», «LEXinform», «NOMOS DATA POOL»; в Австрии — «RDB»; в Канаде — «DATUM»; в Финляндии — «Finlex»; во Франции — «IRETIV», «JURIDIAL», «JURISDATA», «SINDONI». В большинстве случаев эти системы носят негосударственный характер. Так, с 1971 г. в США для нужд Конгресса на государственные средства создавалась правовая система «JURIS». Однако уже с 1982 г. государство вынуждено было договориться о совместном финансировании проекта с частной фир-

мой West Publishing Company. С октября 1993 г. система целиком передана в руки двух компаний — West Publishing Company и Mead Data Systems.

Отечественные АИСЗ

Первыми АИС по законодательству в СССР были следующие системы:

- «АИС Законодательство», разработанная ВНИИСЗ (Всесоюзный НИИ советского законодательства) в среде «ПЕГАС» (ЕС ЭВМ);
- «БД Свод Законов СССР», функционировавшая в Минюсте СССР, на базе операционной системы IN-PICK (ЭВМ IN-500/5000 Realite (Франция)).

Это были системы, реализованные на больших и средних ЭВМ и ориентированные на ограниченный круг пользователей. На информационной базе этих АИС Главный демонстрационно-испытательный вычислительный центр (ГДИВЦ) ВДНХ СССР разработал в 1988 г. первый в СССР банк данных по законодательству для персональных компьютеров ЮРИУС, который выпускался с 1988 по 1993 г. Базы данных ЮРИУС эксплуатировались на III—IV съездах народных депутатов РСФСР, в рамках службы Президиума Верховного Совета РСФСР [3, 31, 44].

Затем появились прототипы информационных системы «Гарант» и «КонсультантПлюс», которые были разработаны НПО «Вычислительная математика и информатика» («ВМИ») в 1990 г.

В 1992 г. отдел НПО «ВМИ», занимавшийся разработкой «Гаранта», выделился в отдельную фирму (НПП «Гарант-Сервис»), а НПО «ВМИ» приступило к разработке проекта «КонсультантПлюс». В 1992 г. была создана общероссийская сеть «КонсультантПлюс», которая охватила почти 200 городов России.

В настоящее время в число наиболее крупных разработчиков таких баз входят фирмы «Гарант-сервис», ВМИ, ЦКР, «Интралекс». Представим некоторые из них.

Компания «Гарант-сервис» оказывает информационно-правовую поддержку на основе системы «Гарант» с 1990 г. Основные поставляемые базы: «Законодательство России» (10 000 документов), «Московское законодательство» (4000 документов), «Банковское законодательство» (4000 документов), «Таможенное законодательство» (5000 документов), «Жилищное законодательство» (2100 документов), «Землепользование, недропользование, природоохрана» (3000 документов), «Международное право» (1000 документов),

«Международное право Россия-СНГ» (2400 документов), «Уголовное и административное право» (2000 документов), «Суд и арбитраж» (3000 документов), «Комментарии к законодательству» (50 документов), «Проекты законов» (более 100 документов), «Налоги, бухучет, предпринимательство» (4000 документов) с дополнительными «Вопросами и ответами» (3500 документов), «Формы правовых документов» (1300 документов), «Архивы Гаранта» (4500 документов). Количественные данные здесь указаны на 2002 г.

Каждая база имеет значительное число гипертекстовых ссылок (от 30 000 до 130 000 в зависимости от базы), комментарии юристов. «Гарант» — полнотекстовая база данных, имеющая разветвленный системный аппарат поиска. В системе содержатся все редакции документов с момента их первой публикации. Обновление баз производится на дискетах, CD ROM, по электронной почте, через Интернет.

В компьютерных системах «Консультант Плюс» фирмы ВМИ содержится более 200 тыс. документов. Это нормативные акты РФ, документы министерств и ведомств, консультации специалистов Минфина, ГНС и ЦБ РФ по финансовому анализу, свод precedентов судебной и арбитражной практики, законодательство более 60 регионов страны.

Как альтернатива этим основным базам появились ежемесячно обновляющиеся CD-выпуски различных фирм. Среди них лидирует компакт диск «Ваше право» фирмы ИСТ. Диск предоставляет информацию об основах законодательства России, региональном законодательстве, федеральных законах, указах, постановлениях, инструкциях и т. д. Каждый выпуск помимо последних редакций документов содержит специальное тематическое приложение, которое более глубоко раскрывает особенности законодательства в какой-либо определенной сфере жизни, например «Автомобиль и закон». Отличительной особенностью диска «Ваше право» является скоростной полнотекстовый поиск документов с учетом всех слов русского языка. Сохранен и привычный поиск по шаблону, для решения сложных задач предусмотрен расширенный поиск. В этом режиме можно использовать в запросе логические выражения. Гипертекстовая навигация по ссылкам между документами обеспечивается обозревателем Microsoft Internet Explorer. Появилась возможность встраивать найденный документ или его часть в другие приложения Windows.

Фирма «Референт-сервис» выпускает CD ROM «Референт», который содержит информацию по тематике «Налогообложение и бухучет», материалы «Финансовой газеты», ответы аудиторов на во-

просы пользователей, а также шаблоны документов для бухгалтера. Версия обновляется ежемесячно.

Издательство «Элекс» выпустило диск «Законодательство России». Ежемесячный выпуск содержит документы по федеральному законодательству (40 000 документов), документы органов власти и управления 37 субъектов РФ, а также справочно-библиографическую информацию, образцы и формы документов, толковый юридическо-экономический словарь с 6000 терминами (всего включено 105 000 документов). Система предлагает развитые средства поиска документов и работы с ними.

Фирма ЦКР выпускает диски информационно-консультационной системы «Юридический советник», включающей в себя следующие разделы: основные документы российского права (4000 документов), юридический и деловой практикум (150 документов) с примечанием юристов по оформлению, консультации юриста, юридический словарь на 4000 слов и выражений. Имеется облегченная версия под названием «Юридический справочник для всех». Есть возможность воспользоваться on-line услугами системы «Кодекс».

Типология юридических ИС

АИСЗ являются частью следующих типов юридических информационных систем [32, 44].

1. *Справочно-информационные системы общего назначения*, ориентированные на доступ пользователей любой профессиональной ориентации к нормативно-правовым актам. К этим системам относятся АИСЗ ЮРИУС, Консультант, Гарант, Кодекс и др. (Российская Федерация).

2. *Глобальные информационные службы (хост-системы)*, предоставляющие доступ удаленным пользователям к библиографической, полнотекстовой или другой информации. Крупнейшей в мире коммерческой службой, обеспечивающей доступ к юридической информации, является система LEXIS (США). Начало ее разработки относится к 1967 г., когда было заключено соответствующее соглашение между коллегией адвокатов шт. Огайо и фирмой Mead Data Corp.

Известны другие системы данного типа:

- JURIS GmbH (Саарбрюккен, ФРГ), которая предоставляет интерактивный доступ к правовой информации, включающей научную литературу, тексты законов, отчеты о судебной практике, а также пресс-релизы Федерального конституционного суда и других высших федеральных судов;
- ITALGIURE/FIND (Италия), WESTLAW (США) и пр.

3. *Системы информационной поддержки деятельности правотворческих органов* (законодательные и конституционные собрания, конгресс, госдума, сенат и пр.). Спецификой таких систем является необходимость хранения и поиска многих версий и редакций нормативно-правовых документов с учетом вносимых поправок и изменений.

4. *Системы автоматизации делопроизводства* судов, полиции, и других правоприменительных или правоохранительных органов. Необходимо заметить, что в странах с прецедентным правом (США) судопроизводство также может рассматриваться как деятельность во многом правотворческого характера, в связи с чем данные системы интегрируются с приведенными выше (п. 3).

5. *Прочие виды систем* — автоматизация юридических библиотек и вузов, консультационных пунктов, юридические экспертные системы и пр.

Основными особенностями юридических АИС являются:

- необходимость предоставления адресного *доступа к полным текстам*, причем в области автоматизированных систем правовой информации основное место занимает спор между сторонами контролируемого индексирования и поиска по ключевым словам (КС) из текста;
- в информационных языках для поиска в БД по законодательству необходим учет *контекстных связей*, возможность использования в запросе предлогов и частиц (И, ИЛИ, НЕ), регламентированных прилагательных (типа «обязательный», «произвольный» и пр.), что отличает их от обычных документальных АИПС;
- тексты нормативных актов должны подвергаться так называемой *юридической обработке* (вид аналитико-синтетической обработки), при котором тексту (или контексту) приписываются не только классификационные индексы, ключевые слова или дескрипторы (как при обычном индексировании), но и комментарии специалистов, ссылки на предшествующие версии, связанные документы, решения судов и пр.

Интегральный банк юридической информации ЮРИУС/JURIUS (тематическая, логическая, физическая структура)

ЮРИУС — Юридическая универсальная информационная система — представляет собой совокупность средств и методов поиска текстовой юридической информации и подробно рассматривается

здесь, поскольку является *первой отечественной полнотекстовой информационным системой* [3, 31, 44].

Тематическая и логическая структура БД. С тематической точки зрения ИБ состоит из следующих основных БД, которые приведены в табл. 4.1 (в хронологическом порядке разработки и выпуска в рамках проекта ЮРИУС).

Таблица 4.1. Тематическая структура баз данных ЮРИУС (по состоянию на 1994 г.)

База данных	Аббре-виатура	Объем, Мбайт	Число документов	Период охвата
Свод Законов СССР, действующие с 1927 г. законодательные акты Союза ССР	СЗ	60	3225	с 1927 по 1989 г.
Трудовое законодательство СССР, акты о труде РСФСР и Союза ССР	ТЗ	20	Около 1000	1920–1987
Комментарий Законодательства о труде, КЗОТ РСФСР и Основы законодательства с комментариями специалистов	КЗОТ	4	Компьютеризованный справочник	
Финансовое законодательство СССР	ФЗ	5	Около 500	1976–1989
Новейшее Законодательство СССР – акты, принятые в 1989–1991 гг. Съездами Народных депутатов, Верховным Советом СССР, Президиумом ВС СССР и Президентом СССР	НЗС	25	Около 2000	1989–1991
Новейшее Законодательство России – акты, принятые с 1990 г. Съездами, Верховным Советом, Президиумом ВС РФ, Президентом и Правительством РФ	НЗР	150	Около 12 000	1990–1994

С логической точки зрения входящие в ЮРИУС БД имеют относительно стандартную структуру и включают две компоненты: регистрационные карты (РК) и полные тексты. РК представляют собой форматированные записи, содержащие относительно стандартный набор библиографических данных, а также ссылку на соответствующий полный текст (рис. 4.1). Полные тексты документов состоят из страниц двух типов:

- логических, т. е. структурных единиц текста (пункт, параграф, статья);
- физических (принудительное разбиение длинного неструктурированного текста на фрагменты одинаковой длины).

Строго говоря, вышеизложенное описание (как и рис. 4.1) в точности соответствует только базам данных НЗС, НЗР, ФЗ, в которых поиск осуществляется по цепочкам: *Термин — РК — Доку-*

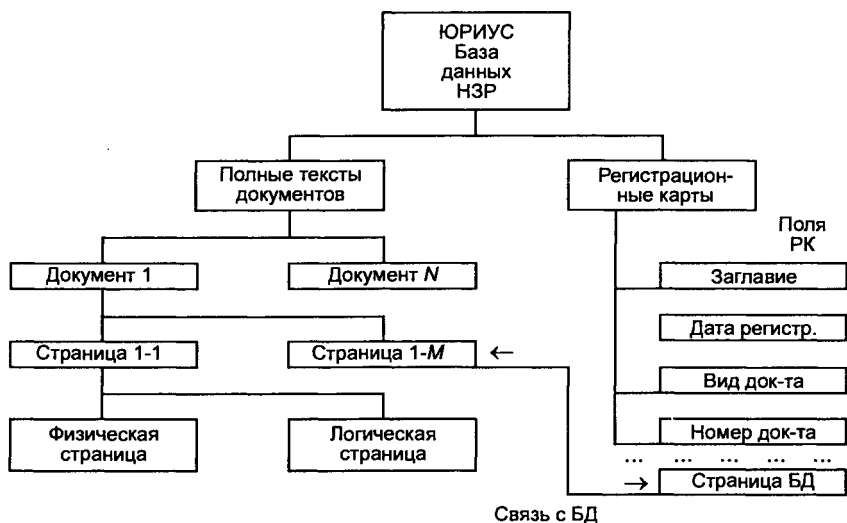


Рис. 4.1. Логическая структура БД ЮРИУС

мент — Страница (интерфейс IX), или Термин — Страница — Документ — РК (интерфейс DC), или Термин — Страница — Термин — Страница — ... (гипертекстовый документальный интерфейс DCH). В базе данных СЗ используется измененная последовательность: Термин — РК — Страницы — Документ, поскольку каждый документ задается начальной и конечной страницей, указанными в РК.

Кроме этого, в БД «СЗ» имеется возможность отнесения документа к тому или иному тому СЗ (с номерами 1—10). Это связано с традицией выпуска Свода законов в форме десятитомного печатного издания.

Физическая структура БД ЮРИУС (рис. 4.2) является примером реализации документальной ИС в среде системы программирования с элементами реляционной СУБД (см. также графу 4 табл. 2.2).

Файл текстовой части БД (szdoc.dbf) — один или несколько файлов, в которых содержатся полные тексты актов (см. также рис. 4.10). На логическом уровне образует иерархическую структуру: БД (том), документ, страница (см. рис. 4.1).

Словарный файл текстовой части (dcfrv.dbf) — представляет собой список слов и стандартных словосочетаний (например «статья 256», «п. 13», «N 1400-РП»), извлеченных из текста, сопровождаемых частотами появления в данной БД (см. рис. 4.7, 4.8). Практика выделения словосочетаний при индексировании с целью вклю-

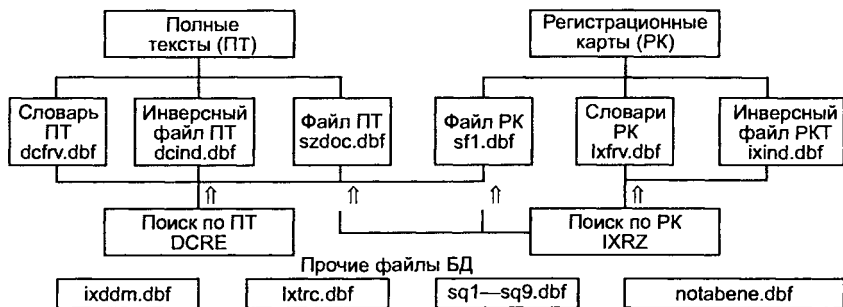


Рис. 4.2. Физическая структура БД и использование файлов модулями пользовательского интерфейса (БД НЗР)

чения их в словарь и инверсный список является достаточно известной.

Инверсный файл текстовой части (dcind.dbf) — список кодов слов и словосочетаний, сопровождаемых номерами страниц. Словарный и инверсный файл используются для сквозного полнотекстового поиска.

Справочно-поисковые файлы (СПФ) (до девяти различных файлов *sf1.dbf* — *sf9.dbf*). Стандартным является файл *регистрационных карт* нормативных актов (РК) — *sf1.dbf*, запись которого содержит наименование, дату, номер, вид, ссылки на страницы БД и другие поля, перечень которых может изменяться для конкретной БД (рис. 4.3).



Рис. 4.3. Взаимодействие программных компонент ЮРИУС в процессе создания и использования БД

Словарь справочно-поисковых файлов (ixfrv.dbf) содержит значения и коды полей (например РК) совместно с частотой появления (рис. 4.3) и ссылкой на номер файла СПФ.

Инверсный файл СПФ (ixind.dbf) содержит коды слов и словосочетаний. Словарный и инверсный файлы используются для поиска записей СПФ (РК, рубрики указателя и т. д.) с доступом к странице БД.

Файл синонимов (ixtrc.dbf) служит для расшифровки кодов или аббревиатур, а также для обеспечения двуязычного поиска в словарных файлах (см. рис. 4.2).

Файл описания СПФ (словарь данных ixddm.dbf — табл. 4.2) — содержит данные о полных, сокращенных и внутренних именах полей каждого СПФ, типах данных, разделителях слов, методах обработки числовых кодов и т. д. Используется при поиске через СПФ и при построении словарных и инверсных файлов.

Содержание колонок табл. 4.2 (см. также прил. 3—4):

- FLDNB — номер поля;
- FLDNM — имя поля в файле sf1.dbf;
- HDR01 — полное название поля (выводится в строке комментария, рис. 4.4);

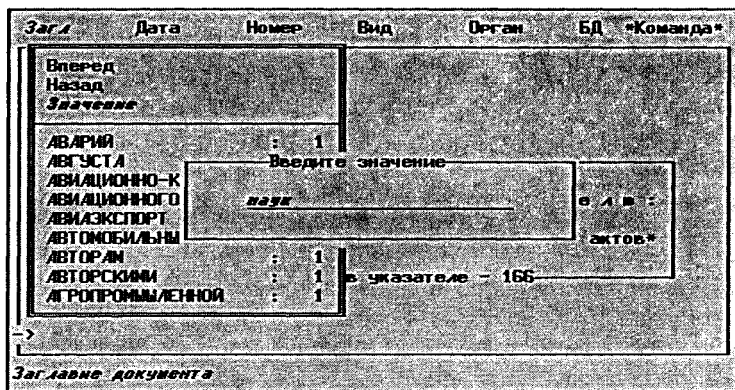


Рис. 4.4. Поиск в частотном словаре (по заголовкам)

- HDR02 — краткое название (выводится в строке меню, рис. 4.4.);
- MULTI — поле элементарное, форматное (1) или множественное, текстовое (число большее 1);
- DESCR — инвертируемое (пусто) или нет (непустое значение);

Таблица 4.2. Пример словаря данных для регистрационных карт БД по финансовому законодательству (ФЗ)

FLDNB	FLDNM	HDR01	HDR02	MULTI	DESCR	FLENG	LONGF	HINMB	INITV	TERMW	PURGW	RLENG	FIDEC	DATYP	TRCOD
00		Регистрационные карты нормативных актов	PK	10	+	0	0	10	02			0	0		0
01	N1	Заглавие документа	Загл	40		240	3	55	A	...;	::;"()/?!<>	20	0	A	0
02	D1	Дата утверждения	Дата	1		8	1	8				8	0	A	0
03	A1	Номер акта	Номер	3		23	1	9		, *	10	6	X	0	
04	V1	Вид документа	Вид	5		35	1	2		, *	20	0	A	0	
05	O1	Орган, утвердивший акт	Орган	5		44	1	6	A	, *	20	0	A	0	
06	R1	Классификационная рубрика ЭК	Рубрика	10		120	1	10		:/	:"[()]	20	0	A	0
07	K1	Ключевые слова	Ключ. сл.	10		160	1	7	A	,.	*	20	0	A	0
08	I1	Опубликовано (источник)	Источник	5		79	1	18		;'[]()	20	0	A	0	
09	S1	Статус документа	Статус	1		30	1	1				30	0	A	0
10	T1	База данных (том)	БД	1		2	1	1				2	0	A	0

- **FLENG** — длина поля в БД (в файле *sf1.dbf*);
- **LONGF** — короткое (1) или длинное (более 1) поле. Длинное поле логически содержится в нескольких полях *sf1.dbf*. В данном случае «Заглавие документа» помещается в поля N1, N2, N3 файла *sf1.dbf*;
- **HINMB** — «гистограмма» значений — количество различных значений поля, накопленных в частотном словаре;
- **INITV** — начальное значение, с которого частотный словарь выдается в меню интерфейса пользователя (см. рис. 4.4);
- **TERMW** — разделители слов (используются для извлечения слов из множественного поля);
- **PURGW** — удаляемые символы;
- **RLENG** — длина термина, выделяемого из поля и помещаемого в частотный словарь;

- FIDEC — число цифровых позиций в поле (слове), если поле имеет тип N или X;
- DATYP — тип поля (A — символьное, N — числовое, X — смешанное);
- TRCOD — использовать или нет расшифровку аббревиатур или перевод слова при его выдаче в составе частотного словаря.

Файлы хранимых запросов (sq1.dbf—sq9.dbf) содержат запросы к СПФ БД, отлаженные и сохраненные пользователем.

Файл заметок (notabene.dbf) позволяет пользователю дополнить СПФ собственными именованными прямыми ссылками на страницы БД.

Программные средства БД ЮРИУС. Рассмотрим структуру программных средств ИБД ЮРИУС (рис. 4.3).

Средства администратора БД и АРМ подготовки данных. Программные средства ЮРИУС позволяют осуществлять выделение тематических фрагментов БД на основе дескрипторного поиска и объединения нескольких фрагментов и БД (нескольких БД) при установке у пользователя.

Функции интерфейса оператора подготовки данных (ОПД) встроены в интерфейс АБД, однако в состав ЮРИУС входит также автоматизированное рабочее место (АРМ) ОПД, предназначенное для децентрализованного использования в пунктах подготовки данных.

АРМ оператора подготовки данных реализует следующие функции:

- ввод текстов актов и регистрационных карт;
- поиск РК по названию, дате, номеру;
- просмотр и корректировка РК и текста акта.

Различаются: *центральный интерфейс АБД*, предназначенный для создания и поддержания дистрибутивной версии БД, и *локальный интерфейс* администратора базы данных, предназначенный для выполнения на ПЭВМ конечного пользователя (единичного или в режиме сети).

Центральный интерфейс АБД обеспечивает функции:

- построение словарного и инверсного файлов для СПФ;
- корректировка (вставка, удаление, редактирование элементов) СПФ с соответствующим обновлением ассоциированных файлов;
- построение словарного и инверсного файлов для полнотекстового словарного поиска;

- просмотр словарей, визуальное обнаружение ошибок, исправление словарей и текста;
- загрузка данных в БД;
- слияние нескольких БД (соединение и обновление файлов, описанных выше). Построение частотных словарей сопровождается накоплением статистики словоупотреблений с целью создания критериев отбора терминов индексирования.

Основные операции при загрузке БД в рамках центрального интерфейса (случай поиска по РК) включают построение на основе файла SF1 файлов IXIND, IXFRV, управление процессом загрузки (типом обработки полей) осуществляется с помощью файла IXDDM, структура и содержание основных полей которого приведены в табл. 4.2.

Локальный интерфейс АБД реализует следующие функции при установке БД на машине пользователя: определение и выбор устанавливаемого фрагмента БД; инсталляция выбранного фрагмента БД; объединение нескольких фрагментов БД (нескольких БД из ЮРИУС).

Интерфейсы конечного пользователя.

Интерфейс IX обеспечивает поиск по значениям полей РК с последующим переходом к соответствующему полному тексту.

Первоначально пользователю предъявляются частотные словари полей РК в форме PullDown-меню (см. рис. 4.4). Выбранные из словаря значения комбинируются в логическое выражение запроса. Найденные РК могут просматриваться в сокращенном (рис. 4.5) или полном (рис. 4.6) форматах, затем пользователь может перейти к просмотру полного текста (рис. 4.7).

После запуска интерфейсной программы на экране терминала появляется заставка системы, а после инициализации — Главное меню.

Найдено: 4 записей		Номер первой записи: 1	
Вперед	Назад	Конец	Печать
Выберите →, ← и нажмите <Enter>			
Дата	№	Вид	Орган
93.07.12.	650.	ВРС.	ПРАВИТЕ
93.07.12.	655.	ПОС.	ПРАВИТЕ
93.07.13.	668.	ПОС.	ПРАВИТЕ
93.07.23.	1309-Р.	РАС.	ПРАВ
Наименование			
*ОБ УТВЕРЖДЕНИИ ПОЛОЖЕНИЯ О МИНИСТ			
*ВОПРОСЫ ОБЕСПЕЧЕНИЯ ПОТРЕБНОСТЕЙ			
*О ПОДПИСАНИИ СОГЛАШЕНИЯ МЕЖДУ ПРА			
*О ПРОВЕДЕНИИ ЮЖИ КОНГРЕССА ЕВРО			
Просмотр списка РК вперед			

Рис. 4.5. После поиска. Найдено 4 документа (сокращенный формат)

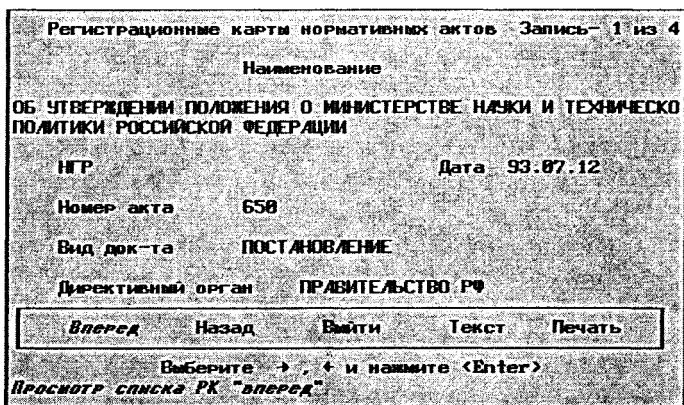
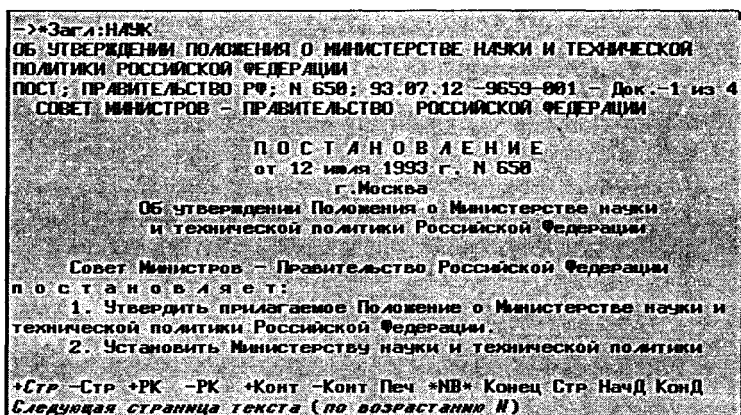


Рис. 4.6. Просмотр РК в полном формате

Рис. 4.7. Просмотр полного текста
(начало документа)

Структура Главного меню определяется структурой поискового файла, содержащего РК документов, и состоит из следующих строк и компонент (см. рис. 4.4):

- *строка горизонтального меню.* Перемещение по горизонтальному меню осуществляется с помощью клавиш <←> и <→>;
- *строки вертикального меню,* зависящие от выбора горизонтальной рубрики. Активизация вертикальной рубрики производится клавишами <↑>, <↓> и нажатием <Enter>;
- *строка запроса, строка предупреждений, строка комментария.*

Рубрики горизонтального меню разделяются на две группы:

- постоянные рубрики — пункты *Команда*, *Запрос*, *NotaBene*;
- переменные рубрики — список поисковых полей регистрационных карт. Этот список определяется составом РК (содержанием ixddm.dbf. табл. 4.2) и может меняться для различных БД из Банка ЮРИУС.

При поиске в словаре по Значению необходимо в экранном окне набрать несколько первых символов отыскиваемого в словаре термина (рис. 4.4). На экране отображается страница словаря начинающая с первого термина, содержащего введенные символы. После формирования запроса необходимо перейти к рубрике *Команда* и выполнить запрос. Результат поиска (в сокращенном формате одна запись — одна строка экрана) представлен на рис. 4.5, в полном (одна запись РК на экране) — на рис. 4.6. Для перехода к просмотру полного текста (рис. 4.7) необходимо подвести курсор к соответствующей строке (рис. 4.5) или к позиции Текст (рис. 4.6) и нажать <Enter>.

Поисковый критерий генерируется автоматически в процессе ввода пользователем (или выбора из словаря) списка значений. Основные принципы построения запроса следующие:

- критерии для разных полей соединяются связкой AND (И);
- критерий для поля составляется из нескольких значений через AND (И) (для множественного поля) либо OR (ИЛИ) (элементарное поле);
- отрицание (связка НЕ) задается знаком '-' («минус») перед соответствующим словом (значением).

Например, если пользователь выбрал значения АКАДЕМИЯ, НАУК для поля Загл и 93.04.01, 95.07.07 для поля Дата, эквивалентный поисковый критерий будет выглядеть следующим образом:

```
(Загл='АКАДЕМИЯ') AND (Загл='НАУК') AND (Дата='93.04.01' OR
Дата='95.07.07') .
```

Интерфейс полнотекстового поиска (ДС). Пример главного меню ДС-интерфейса приводится на рис. 4.8. Поиск в частотном словаре активизируется рубрикой Значение.

Для поиска страниц, содержащих термин запроса, необходимо установить курсор на термин и нажать <Enter> (рис. 4.9). Переход к полному тексту (рис. 4.10) осуществляется активизацией рубрики Текст Главного меню.

Выберите ← и нажмите <Enter>					
Вперед	Назад	Значение	Конец	Текст	Отмена
Термин		Число страниц		Примечания	
АБАКАН		1			
АБДУЛБАСИРОВ		1			
АБЗАЦ		8			
АБЗАЦЕ					
АБЗАЦЕМ					
АБЗАЦЫ					
АБОНЕМЕН					
АВАНСИРО					
АВАРИЙНО-СПАСАТЕЛЬНЫ		1			
АВГУСТА		16			
АВГУСТЕ		2			
					Док. в БД - 166

Рис. 4.8. Поиск в частотном словаре полных текстов

Выберите ← и нажмите <Enter>					
Вперед	Назад	Значение	Конец	Текст	Отмена
Термин		Число страниц		Примечания	
	Найдено -	12 страниц			
МИННАЗКИ		12			
МИНОБОРОНЫ		8			
МИНОБРАЗОВАНИЕ		18			
МИНОБРАЗОВАНИЯ		1			
МИНПРИРОДЫ		18			
МИНСВЯЗИ		2			
МИНСЕЛЬХОЗ		12			
МИНСЕЛЬХОЗА		3			
МИНСЕЛЬХОЗУ		3			
МИНСКЕ		3			
МИНТРАНСУ		6			
Запрос: МИННАЗКИ					Док. в БД - 166
Выбрать для поиска термин - *МИННАЗКИ*					

Рис. 4.9. После поиска. Найдено 12 страниц

Запрос: МИННАЗКИ

(О СОЗДАНИИ В Г. МОСКВЕ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОГО ИНСТИТУТА СПЕЦИАЛЬНЫХ ПОЛИМЕРНЫХ МАТЕРИАЛОВ С ОПЫТНЫМ ПРОИЗВОДСТВОМ НА БАЗЕ ПОДПРАСП; ПРАВИТЕЛЬСТВО РФ; N 1240-Р; 93.07.12-9626-001 -СТР. -1 ИЗ 12
 СОВЕТ МИНИСТРОВ - ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

РАСПОРЯЖЕНИЕ
 от 12 июля 1993 г. N 1240-Р
 г.Москва

Принять предложение Роскомоборонпрома, согласованное с *Минназками* России и правительством Москвы, о создании в г.Москве Научно-исследовательского института специальных полимерных материалов с опытным производством на базе подразделения Института коррозии и государственного предприятия "Спецмагнит".

Председатель Совета Министров -
 Правительства Российской Федерации В.Черномырдин

*Стр -Стр *Цит -Цит *Конт -Конт Печ *НВ* РК Конец Стр НачД КонД
 Следующая страница текста (по возрастанию N)

Рис. 4.10. Первая из найденных 12 страниц

4.2. Офисные информационные системы

Возможности современных программно-технических средств автоматизации и связи позволяют создавать высокоэффективные системы информационно-коммуникационного обслуживания широкого класса пользователей, деятельность которых напрямую связана с учетом, хранением и контролем исполнения документов (законов, указов, постановлений, распоряжений, протоколов, писем и т. п.).

Современная офисная система представляет собой совокупность информационных технологий, ориентированную на совместное и скоординированное использование «электронных» методов сбора, обработки, хранения и передачи информации, в том числе и управленческой, которая может быть представлена в виде текстов, рисунков, графиков, таблиц, аудио- и видеоинформации [30].

Основные функции современной офисной системы:

- интеграция технологий делопроизводства в единый процесс;
- переход на использование безбумажной технологии;
- автоматизация рутинных операций по обработке информации;
- архивизация информации;
- создание и поддержка приложений;
- обеспечение сохранности, целостности, достоверности информации;
- установление уровней доступа и защита от несанкционированного доступа к информации;
- обеспечение эффективного поиска информации по различным критериям;
- создание гибкой системы разделения и распределения информации;
- организация коллективной работы пользователей в локальных и глобальных информационно-вычислительных сетях и средах;
- быстрая и надежная связь и передача информации по каналам связи удаленным пользователям;
- обеспечение дружественного единого интерфейса пользователя со всеми компонентами системы.

Технологии автоматизированного делопроизводства занимают основное место в офисной автоматизации и реализуют следующие функциональные задачи:

- ввод документов с целью их последующей обработки и оформления;

- регистрация и учет входящих, исходящих и внутренних документов;
- распределение документов по исполнителям и при необходимости назначение срока исполнения по установленным руководством нормативам;
- контроль исполнения документов с выдачей напоминаний исполнителям в установленные руководством сроки;
- формирование и выдача руководству справок о состоянии исполнения документов по их видам, исполнителям и другим параметрам;
- ведение оперативных и архивных фондов хранения документов (своевременный перевод данных из одного массива в другой, внесение изменений и дополнений и т. п.);
- поиск документов по запросам;
- поиск архивной информации по запросам;
- составление различных итоговых и сводных документов;
- обмен документами (внешний и внутренний).

Основные функциональные подсистемы офисных ИС

Ввод документов обеспечивается достаточными по мощности текстовыми процессорами, некоторые из которых имеют специальные конверторы форматов для работы с информацией, подготовленной в другой среде. Текстовые процессоры позволяют вводить и редактировать документы сложной структуры; обеспечивают полное соответствие изображения документа на экране его копии на бумаге, что создает дополнительные удобства для пользователей; имеют возможность орфографического, грамматического и стилистического контроля текста с ведением словарей. В некоторых офисных системах, кроме ручного, предусмотрен ввод посредством оптического сканирования печатных или рукописных текстов с использованием программных средств распознавания символов (OCR). Развиваются технологии акустического ввода и распознавания текстов.

Регистрация и учет документов могут быть реализованы средствами СУБД. Для этого необходимо определить перечень и формат представления информации, характеризующих документ (вид документа, дата поступления, входящий/исходящий номер, наименование, имя отправителя/получателя, краткое содержание и т. п.). Для дальнейшего поиска необходимо определить поисковые критерии и ключевые слова.

Распределение документов по исполнителям и разграничение доступа осуществляется либо с помощью систем электронной почты, либо средствами обеспечения коллективной работы (операционные системы, в т. ч. сетевые или СУБД с коллективным доступом). Администратор системы ведет списки исполнителей, объединенных в рабочие группы, с указанием их адресов в локальной сети, номеров телефонов для модемной связи, номеров почтовых ящиков на почтовом сервере. За каждым исполнителем (группой) закрепляется один или несколько документов, объединенных в «папки». Каждому представителю (группе) назначается срок представления или подготовки документа, который вносится вместе с именем исполнителя в базу данных по регистрации и учету документов. Администратор системы назначает уровни доступа к информации (ее части) каждому исполнителю и/или рабочей группе.

Документы ограниченного доступа могут подвергаться шифровке при вводе и передаются санкционированному исполнителю вместе с шифром. Согласно информации, введенной администратором системы, осуществляются маршрутизация, пересылка и доступ к информации каждым санкционированным пользователем. В некоторых офисных системах применяется технология «публикации в сети», позволяющая санкционированным пользователям просматривать и аннотировать (но не редактировать) документ (делать «заметки на полях»).

Контроль исполнения и поиск документов осуществляется на основе информации, внесенной в процессе регистрации и учета документов в БД. С использованием встроенных средств СУБД возможно создание некоторого ряда фиксированных запросов, допускающих изменение критериев поиска, либо применение генератора запросов, позволяющего пользователю проектировать структуру запроса.

Архивирование информации, ведение фондов, а также поиск документов осуществляется либо средствами локальных СУБД для массивов информации ограниченного объема, либо с использованием мощных распределенных баз данных. Во втором случае хранение информации осуществляется на архивных магнитных или оптических дисках.

Формирование и выдача статистических данных осуществляется на основе информации, внесенной в процессе регистрации, учета и исполнения документов в СУБД или электронные таблицы (ЭТ). С использованием встроенных средств СУБД или макросов ЭТ создаются запросы на поиск информации, ее статистическую обработку; выдачу полученных статистических данных на экран или прин-

тер в форматах выходных документов, предварительно спроектированных средствами СУБД или ЭТ.

Подготовка и оформление документов реализуется с использованием текстовых процессоров, электронных таблиц, графических пакетов и настольных издательских систем. Все перечисленные компоненты в той или иной степени позволяют использовать элементы информации, подготовленной в одном модуле, в других компонентах. Например, график, построенный в графическом пакете, с использованием данных, полученных из электронной таблицы, включается в текст, подготовленный текстовым процессором.

Обмен документами с внешними (удаленными) пользователями осуществляется модулями электронной почты, включенными в средства обеспечения коллективной работы. Почтовые системы, кроме рассылки документов внутри системы, дают возможность пересылки информации пользователям других систем электронной почты.

Программные средства офисных ИС

Первоначально в качестве программно-инструментальных средств автоматизированной поддержки делопроизводства использовались текстовые редакторы, электронные таблицы, СУБД, различные языки запросов пользователя, средства графической обработки информации, средства для организации работы пользователей.

С появлением интегрированных пакетов офисных систем, когда все эти средства были объединены в единое целое и их работа была осуществлена под стандартным интерфейсом, каждый из работающих получил возможность иметь доступ ко всем ресурсам таких систем, которые стали использоваться и для поддержки делопроизводства.

Использование программных комплексов офисных систем в сфере автоматизации делопроизводства дает пользователям целый ряд преимуществ, наиболее весомыми из которых являются общность интерфейса входящих в комплект приложений, высокая степень интеграции программ, низкая цена и менее жесткие (за счет использования общих модулей) требования к системным ресурсам. Как правило, такие комплексы включают в себя стандартный базовый набор приложений (текстовый процессор, электронную таблицу и СУБД), иногда расширяемый программами презентации, системами электронной почты и т. д., которые объединены единой средой, способствующей коллективной работе пользователей.

Требования к составу и архитектуре системы:

- *система* должна базироваться на архитектуре «клиент—сервер», обеспечивающей оптимальное соотношение «затраты/производительность»;
- *серверная часть* системы должна функционировать под управлением ОС UNIX или Windows NT (2000);
- *клиентская часть* системы обычно реализуется под управлением Windows 98/2000/XP, что обеспечивает удобный интерфейс пользователя с возможностью его модификации по мере появления более развитых версий этой системы;
- *система должна обеспечивать интерфейс* с наиболее популярными системами электронной почты (MS Mail, cc:Mail);
- *система должна включать технические и программные средства для обработки и хранения документов на различных носителях* (средства сканирования бумажных документов, оптические диски, магнитные ленты и т. д.). Реализация данных механизмов невозможна без ведения специальных хранилищ (архивов) для занесения и поддержания промежуточных и утвержденных документов.

В табл. 4.3 приведены основные сведения по наиболее распространенным программным средствам реализации офисных ИС. Эти программные продукты могут быть использованы для реализации систем электронного документооборота различного назначения на предприятиях любого профиля.

Наиболее целесообразным представляется их применение в условиях массового потока разнотипных документов на различных носителях при большом числе одновременно работающих пользователей, в т. ч. и для обработки и архивирования технической документации с большим удельным весом графического материала (чертежи, эскизы, рисунки и т. д.).

Необходимо отметить, что программные офисные средства Microsoft Office представляют собой набор логически несвязанных между собой, с точки зрения автоматизации электронного документооборота, компонентов инструментального характера. Чаще всего он включает текстовый редактор, систему электронных таблиц, СУБД, системы электронной почты и иногда средства для проведения презентаций.

Программные средства для рабочих групп класса Lotus Notes представляют собой следующий шаг в развитии офисных систем. В этих продуктах решены проблемы взаимодействия и одновременной работы многих пользователей, коммуникации между ними. В их основе присутствуют инструментальные средства для форми-

Таблица 4.3. Сравнительные характеристики основных систем офисной автоматизации

Производитель	Наименование	ОС	Текстовый процессор	Электронные таблицы	СУБД	Электронная почта	Графическая система	Автоматизация документооборота	Планировщик работ
Microsoft	Microsoft Office	Windows; NetWare	MS Word for Windows	Excel for Windows	Access	MS Mail for Windows	PowerPoint for Windows	-	-
Lotus Development	Lotus Notes	Windows; Unix; OS/2; NetWare; Macintosh	Lotus Ami Pro	Lotus 1-2-3	Lotus Approach, Драйверы Datalens доступ к INFORMIX, ORACLE, SQL SERVER и Т. Д.	Lotus cc-Mail/Notes Mail	Lotus Freelance	Work Flow	Lotus Organizer
Data General	AV Object Office	Windows; Unix; NetWare	MS Word/Word Perfect/Ami Pro	Excel	ORACLE, INFORMIX, INGRESS, SYBASE	AV onGO Office	AV Image	Система рассылки Agent;	-
Novell Application Group	Perfect Office	Windows; Unix; NetWare	Word Perfect for Windows	Quattro Pro for Windows	Paradox for Windows	GrupWise	AppWare Презентационная графика Presentation 3.0.	Просмотр и «публикация» документов в сети Envoy 1.0	InfoCentral for Windows
Borland WordPerfect	Borland Office	Windows; NetWare	Word Perfect for Windows	Quattro Pro for Windows	Paradox for Windows	ObjectExchange	-	Work Group Enable Kit	-
Siemens Nixdorf (SNI)	OCIS (Office Communication & Information Systems)	Windows, UNIX / SINIX и сервисные службы UNIX / SINIX	ComfoTex	MS Excel, Wingz, Lotus 1-2-3	ComfoBase на основе стандарта SQL	Mail X, Unix-Mail	MS Paintbrush, MS PowerPoint	OnTime	Doculive

рования документов, их маршрутизации между несколькими пользователями, реализации некоторых этапов технологического цикла обработки документов (визирование документов, контроль исполнения и т. д.). Архитектура большинства продуктов этого класса включает в свой состав базу данных для хранения документов и различных описаний, связанных с ними.

Оптимальным решением для организации электронной обработки и хранения больших массивов документов различных типов на разнообразных носителях является применение интегрированных систем, позволяющих обеспечить следующие условия:

- соответствие типов документов и предлагаемой технологии их обработки требованиям и стандартам, установленным в организации;
- масштабируемость и интенсивность системы к изменению требований пользователей в перспективе;
- базирование системы на наиболее прогрессивных и проверенных временем решениях в области ПО.

Офисный комплекс Lotus Notes

Lotus Notes (LN) — комплекс, разработанный фирмой Lotus Development, поддерживает следующие типы данных: текстовые, числовые, дата/время, ключевые поля; RTF-поля для хранения текста, таблиц, графических, аудио- и видеоданных или OLE-объектов.

Интеграция компьютерных программ осуществляется через: Clipboard; экспорт/импорт файлов; механизм OLE, драйверы Datalens для доступа к внешним базам данных INFORMIX, ORACLE, SQL SERVER (рис. 4.11) и т. д. Для защиты информации в LN используется схема шифрования RSA PublicKey/PrivateKey, обеспечивающая наивысший уровень защиты, доступный приложениям типа «клиент—сервер». Для подтверждения подлинности документов используются электронные подписи. Для защиты особо важной информации предусмотрены дополнительные возможности, среди которых:

- хранение промежуточных версий документов;
- хранение списка пользователей, редактировавших документ;
- шифрование всех данных, передаваемых по сети;
- шифрование хранимых данных, почты;
- система паролей доступа.

Поиск информации осуществляется по ключевым полям. Имеется возможность ведения архивов внутренних и внешних документов.

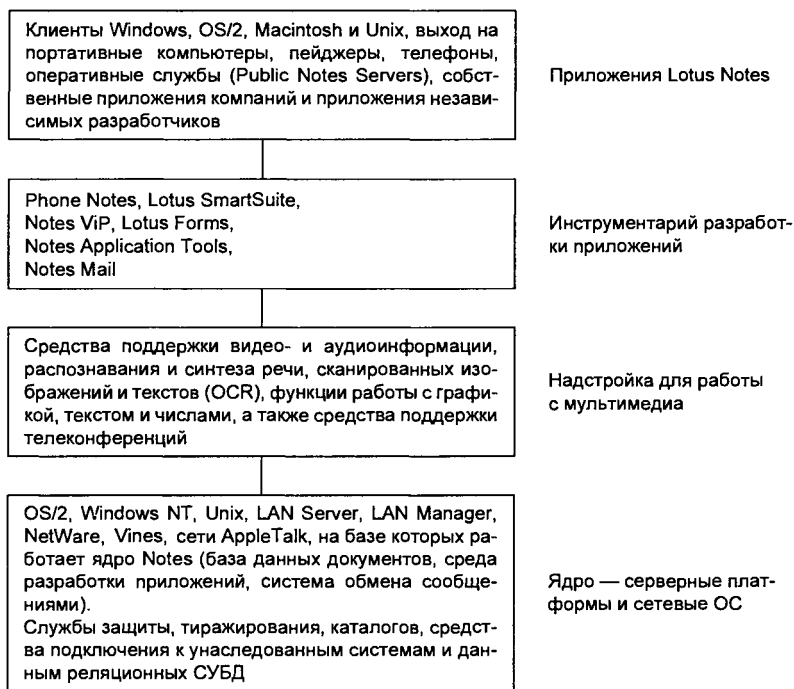


Рис. 4.11. Уровни программного обеспечения Lotus Notes

С программным продуктом поставляются 25 шаблонов, таких, как «Дискуссия», «Обслуживание клиентов», «Хранилище документов» и др. Все они имеют практическое применение и могут быть легко приспособлены к конкретным потребностям. База данных может содержать разноформатные документы системы Notes. Кроме того, базы данных содержат *макросы*, которые автоматизируют работу, *бланки*, которые используются для ввода информации в сложные документы, а также просматриваемые *отчеты* из документов базы данных.

В состав пакета LN входят три основных элемента:

- средства маршрутизации и обработки;
- средства обеспечения защиты/управления;
- архитектура базы данных продукта.

LN обладает развитыми средствами создания интерфейсов (давая возможность встраивать непосредственно в документы командные кнопки); имеются средства текстового поиска, позволяющие строить список документов, соответствующих заданному критерию поиска.

Уровни доступа к данным. LN поддерживает как групповую работу, так и коммуникации между двумя любыми пользователями, при этом допускается, чтобы множество пользователей одновременно обращались к нескольким базам данных. Управление доступом осуществляется на нескольких уровнях: базы данных, формуляра или вида, документа или части документа. С другой стороны, управлять доступом можно также на уровне групп и отдельных пользователей. Для совместной работы с базой данных в большом коллективе ее можно разместить на сервере и задать для отдельных пользователей или групп различные уровни доступа. Всего в LN предусмотрено семь уровней доступа:

1. *Manager* — способен выполнять все операции с базой данных, включая чтение, запись и редактирование документов и форм.

2. *Designer* — имеет те же полномочия, что и *Manager*, но не может управлять уровнем доступа и параметрами тиражирования.

3. *Editor* — уполномочен на чтение, запись и редактирование всех документов в базе данных, но ему не разрешается модифицировать формы и отчеты.

4. *Author* — может создавать новые документы и редактировать их, а также читать (но не редактировать) другие документы.

5. *Reader* — имеет право только на чтение.

6. *Depositor* — вводит новые документы, но не может редактировать существующие.

7. *No Access* — пользователю не разрешается открывать базы данных или сохранять документы.

Кроме того, LN включает в себя систему обмена сообщениями — электронную почту *Notes Mail*, позволяющую посылать корреспонденцию другим пользователям или группам. Он содержит все необходимые для этого компоненты: распределенную базу данных (документов), интегрированную среду разработки приложений и средства передачи сообщений. LN поддерживает как подключенных к сети, так и удаленных пользователей. Он включает в себя компоненты интегрированного программного обеспечения для коллективной работы: среду для разработки и реализации приложений, хранилище объектов, двустороннее тиражирование, средства защиты и обмена сообщениями, межплатформную поддержку и масштабируемость.

Пользователи могут обращаться к сети через графические интерфейсы *Windows*, *Macintosh*, *OS/2* или *Unix*. *Notes* совместим с такими сетями, как *AppleTalk*, *PATHWORKS*, *Novell NetWare*, *TCP/IP*.

База данных LN

БД представляет собой совокупность информации по определенным тематическим направлениям, которая может содержать документы различных типов, включая комментарии, отчеты, коммерческие предложения, финансовые сводки, информацию о контактах, электронные таблицы, изображения или видео- и аудиоинформацию. Несмотря на то, что база данных LN не является реляционной, в ней могут быть установлены связи между записями.

Перечислим некоторые из типов баз данных, которые наиболее часто встречаются в приложениях.

Библиотека документов содержит данные (отчеты, записки, корреспонденции заказчиков, различных форм и запросов) в форме, соответствующей некоторому внешнему представлению, и которые всегда можно отобразить на экране или напечатать на бумаге;

Конференции — базы данных, предоставляющие возможность сотрудникам участвовать в письменном обсуждении текущих вопросов. После создания базы данных уполномоченные пользователи могут добавлять в нее собственные документы. Тот, кто инициировал телеконференцию, автоматически становится ее руководителем, получает все полномочия доступа, а также определяет уровни доступа для других пользователей. Удаленный доступ к конференциям осуществляется по телефонным каналам (через модем) или региональные сети.

Информационные службы позволяют получать свежую информацию по определенным темам (например, служба новостей о производстве может представить возможность узнать о последних событиях на производстве в отрасли или филиалах, расположенных за рубежом).

Одни и те же команды и методы используются для доступа ко всем базам данных, независимо от типа содержащихся в них данных. LN хранит базы данных в отдельных файлах с расширением .nsf. Nsf-файл базы данных содержит следующую информацию:

- формуляры;
- виды;
- собственно документы, хранящиеся в базе данных;
- информацию о доступе пользователей к базе данных;
- справочную документацию по базам данных.

Документ является основной единицей хранения в БД LN. Многие базы данных содержат только один тип документа с заданным форматом (каждый документ состоит из одного или более полей, в которых система сохраняет данные), однако в общем случае

БД могут содержать документы различных типов. Документы делятся на три категории:

- главный документ (рис. 4.12);
- ответ на документ (рис. 4.13);
- ответ на ответ (рис. 4.14).

Тип документа определяется при его создании (создании его формуляра) в меню «Атрибуты формуляра».

ДОГОВОР № 14

коллективного подряда на создание научно-технической продукции по теме (программе)
исследование прошлого.

Российский Государственный Гуманитарный Университет, именуемый в дальнейшем "Заказчик", в лице *Преображенский Е.Н., профессор /Ф.И.О./* должности, с одной стороны, и временным трудовым (творческим) коллективом, именуемым в дальнейшем "ВТК", в лице руководителя *Незабудкин К.Н. /Ф.И.О./*, с другой стороны, заключили настоящий Договор о нижеследующем:

1. ПРЕДМЕТ ДОГОВОРА

1.1. Заказчик поручает, а ВТК берет на себя обязательство выполнить следующую работу (выполнение этапа работы) *исследование архивов /характер и виды работ/*.

1.2. ВТК выполняет работу (этап работы) в соответствии с техническим заданием и календарным планом.

1.3. Сроки выполнения работы (этапа работы):

- начало *18.01.2001*;

- окончание *18.01.2003*.

ВТК имеет право досрочного выполнения работы.

1.4. Результат работы (этапа работы) представляются в виде *обобщающих отчетов/* форма представления работы/.

1.5. Работа (этап работы) считается выполненной после подписания акта приемки-сдачи представителем Заказчика.

1.6. Заказчик обязуется своевременно принять и оплатить выполненную работу (этап работы).

2. УСЛОВИЯ ВЫПОЛНЕНИЯ РАБОТЫ

2.1. Члены ВТК выполняют работу на свой риск в свободное от основной работы время.

2.2. Руководитель определяет состав исполнителей, распределяет между ними обязанности и оплату труда.

2.3. ВТК представляет Заказчику отчеты о выполненной работе (этапе работы) в сроки, установленные в Договоре, (календарном плане).

2.4. ВТК обязан своевременно и качественно выполнить и передать Заказчику промежуточные и окончательные результаты работ.

3. ОПЛАТА ТРУДА

3.1. За выполненную работу (этап работы) по настоящему Договору Заказчик осуществляет оплату труда (авансированную заработную плату) в размере *100 000,13* руб.

Рис. 4.12. Пример документа (главный документ)

УТВЕРЖДАЮ Проректор по научной работе Минаев В.В. /Ф.И.О./ 14.01.2002 от
<u>АКТ</u> к договору №14 о закрытии годового этапа научно-исследовательской работы за 2001 г.
1. Наименование годового этапа <i>исследование архивов.</i> 2. Сроки выполнения годового этапа <i>10.11.2001.</i> 3. Основные результаты выполнения годового этапа <i>обобщены в отчете с приложением к нему ксерокопий особо важных документов</i> /указываются, какие теоретические положения разработаны или уточнены, конкретные полученные результаты исследований, основные технико-экономические параметры, новизна проведенных исследований/. 4. По годовому этапу составлены <i>отчеты</i> /наименования научно-технических документов, разработанных за год/. 5. Стоимость годового этапа <i>120 000,12 руб.</i>
Руководитель темы <i>Незабудкин К.Н.</i> /Ф.И.О./

Рис. 4.13. Пример документа (ответ на документ)

УТВЕРЖДАЮ Руководитель темы Незабудкин К.Н. /Ф.И.О./ от 18.01.2001
<u>ИСПОЛНИТЕЛЬ</u> Личная карточка к договору № 14
1. ФИО исполнителя <i>Вавилов П.К.</i> 2. Должность исполнителя <i>архивариус.</i> 3. Вид выполненной работы <i>классификация архивов.</i> 4. Адрес исполнителя: <i>Москва.</i> 5. Паспорт : <i>V-AB №345356</i> /серия и номер/.

Рис. 4.14. Пример документа (ответ на ответ)

Формуляр (шаблон документа) соответствует каждому типу документа и представляет собой комбинацию *статического текста* (то, что представляет собой неизменный текст в пределах формуляра) и *полей* (области, которые предназначены для ввода данных пользователями).

Различаются два вида полей: *общие и обыкновенные*. Общие поля отличаются тем, что введенные данные будут отражены во всех документах, присоединенных к главному документу и использующие

это поле в своем формуляре. Кроме того, по способу заполнения в LN выделяются следующие типы полей: *редактируемое, вычисляемое, вычисляемое для вывода, вычисляемое при составлении*. На рис. 4.15 и 4.16 соответственно приводятся примеры формуляра и соответствующего ему документа (договор о НИР). Поля выделены курсивом, общие поля отмечены полужирным шрифтом.

ФИО заказчика /ф.И.О./ должности с одной стороны, и временным трудовым (творческим) коллективом, именуемым в дальнейшем "ВТК", в лице руководителя ФИО руководитель /ф.И.О./ с другой стороны заключили настоящий Договор о нижеследующем:'"/>

Рис. 4.15. Пример формуляра LN

Преображенский Е.Н., профессор /ф.И.О./ должности с одной стороны, и временным трудовым (творческим) коллективом, именуемым в дальнейшем "ВТК", в лице руководителя Незавбудкин К.Н. /ф.И.О./ с другой стороны заключили настоящий Договор о нижеследующем:' followed by the section header '1. ПРЕДМЕТ ДОГОВОРА'."/>

Рис. 4.16. Пример документа, соответствующего формуляру LN

Вид — представление документа при поиске, просмотре, печати, в котором фигурируют значения полей документов или вычисленные на их основании выражения. Вид содержит также описания связей между документами, критерии отбора документов для отображения на экране или включения в отчет. Обычно количество видов превышает количество формуляров, что связано с необходимостью получать выдачи документов с различной упорядоченностью и набором полей отображения. Вид аналогичен отчету в реляционных СУБД, где каждому документу при выдаче на печать или экран соответствует одна или несколько строк (при выдаче на экран иногда предусмотрена горизонтальная прокрутка), а каждая колонка отчета обычно соответствует одному или нескольким полям документа (записи) или выражению от полей записи и системных переменных.

Функции и команды LN

Команды и функции представляют собой совокупность инструментальных средств, предназначенных для проектирования интерфейсов и навигации в БД, что осуществляется, в частности, путем составления *формул* и их приписания визуальным элементам экрана (кнопкам и пр.). Функции начинаются с символа «@», функции и выражения отделяются друг от друга точкой с запятой. Многие команды соответствуют пунктам главного меню (см. рис. 4.16) и для их использования в формулах следует пользоваться функцией *@Command([Имя команды])*. Полный список функций и команд содержится в справке LN, вот некоторые из них:

- *@Adjust(время-дата;год;месяц;день;час;минута;секунда)* корректирует указанную величину «время-дата», изменяя каждый следующий аргумент на величину инкремента. Корректирующая величина может быть положительной либо отрицательной;
- *@Contains(строка;подстрока)* определяет, является ли подстрока частью строки, возвращает 1 (*true*), если подстрока содержится в строке; в противном случае возвращает 0 (*false*);
- *@DocNumber* возвращает строку, соответствующую номеру записи в виде документа или категории;
- *@If(условие 1; действие 1; условие 2; действие 2; ...; условие 99; действие 99; иначе_действие)* оценивает условия, и если *true*, то выполняется действие, указанное сразу же после условия. Если же *false*, то происходит переход к проверке следующего условия и т. д. Если ни одно из условий не истинно, выполняется «иначе_действие»;

- *@Prompt([Стиль]; "Заголовок"; "текст"; "по умолчанию"; "выбор"])* открывает диалоговое окно, в котором введенная информация служит аргументом, используется для запроса информации у пользователя и ветвления действия в зависимости от введенной пользователем информации;
- *@Command([NavigateNextMain])* — переход к следующему основному документу в текущем виде; база данных должна быть открыта на уровне вида или документ может быть открыт в режиме чтения или редактирования;
- *@Command([NavigatePrev])* — переход к предыдущему документу в текущем виде;
- *@Command([NavigatePrevMain])* — переход к предыдущему основному документу в текущем виде.

Интерфейс Notes, создание и манипулирование базами данных

Основное рабочее окно (см. рис. 4.16) включает в качестве основных элементов строку меню, линейку пиктограмм SmartIcons, страницы с закладками и строку состояния. При открытии базы данных выводятся дополнительные окна. Один из основных элементов управления — расположенное в верхней части основное меню, включающее в себя как общие пункты (Ред, Окно и Справка), так и специальные меню LN (Файл, Вид, Создать, Сервис и Дизайн).

С помощью меню Файл пользователь может получить доступ к базам данных и административным функциям LN. Пункт Создать Реплику меню Файл создает копию текущей выбранной базы данных.

При создании базы данных необходимо выбрать пункт меню Файл — Создать базу данных. При этом можно использовать поставляемый с Notes шаблон или спроектировать базу данных самостоятельно (выбрав в диалоговом окне Создать базу данных значение Чистая), кроме того, необходимо задать название БД и название файла БД.

Пункт Открыть базу данных создает на выбранной в данный момент странице пиктограмму базы данных или наряду с созданием пиктограммы открывает базу данных.

Для создания формуляра необходимо выбрать пункт меню Дизайн — Формуляры и кнопку Создать. После этого появляется окно типа текстового редактора. Для каждого формуляра требуется заполнить диалог Атрибуты формуляра (пункт меню Дизайн — Атрибуты формуляра, рис. 4.17), выбрать имя и тип, а также заполнить ряд опций. Необходимо ввести статический текст, составляющий непосредственно шаблон и добавить поля. Чтобы добавить поле, не-

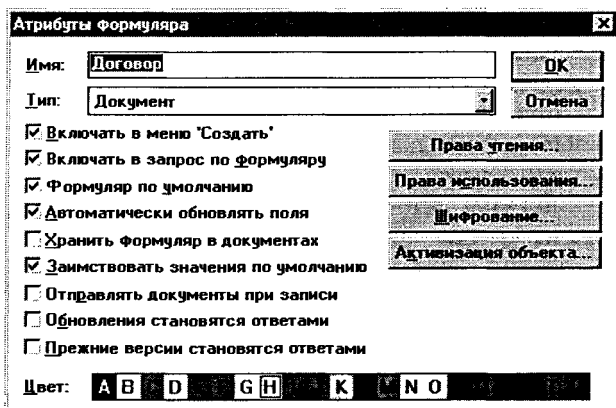


Рис. 4.17. Экран диалога Атрибуты формуляра

обходимо выбрать пункт меню Дизайн — Новое поле или Дизайн — Использовать общее поле, затем определить атрибуты поля. Название поля может состоять из одного слова или нескольких, соединенных знаком «подчеркивание»; если необходимо, поле может быть переопределено как *общее* (пункт меню Дизайн — Сделать поле общим).

Для каждого поля нужно заполнить имя, тип данных, тип отображения, формат и формулу. Кнопка формула необходима для записи функции предварительного заполнения поля, если этого не сделать, то поле по умолчанию остается пустым (рис. 4.18). Кнопка формат определяет формат поля в зависимости от типа данных поля

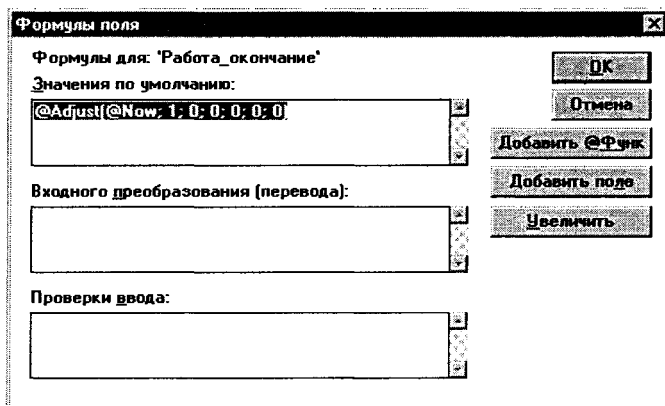


Рис. 4.18. Определение формулы поля данных

(рис. 4.19). Для обеспечения навигации по спискам документов используется *встраивание кнопки в формуляр*. Для добавления кнопки необходимо выбрать пункт меню Ред — Поместить — Кнопку. Далее заполняются текст кнопки и формула для выполнения при ее нажатии (рис. 4.20).

Создание нового вида выполняется с помощью пункта меню Дизайн — Виды и кнопки Создать. Далее заполняется имя вида, тип и некоторые опции в диалоге Дизайн — Атрибуты вида. После определения общих параметров вида нужно добавить колонки для отображения (пункт меню Дизайн — Новая колонка, рис. 4.21) и определить параметры колонки (пункт меню Дизайн — Атрибуты колонки, рис. 4.22). При необходимости отбора документов в выходной отчет необходимо задать критерий (формулу) отбора (рис. 4.23, пункт меню Дизайн — Формула выбора документов); результат отбора представлен на рис. 4.24.

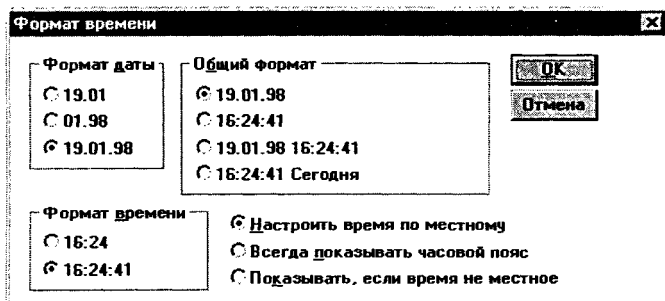


Рис. 4.19. Определение формата поля данных

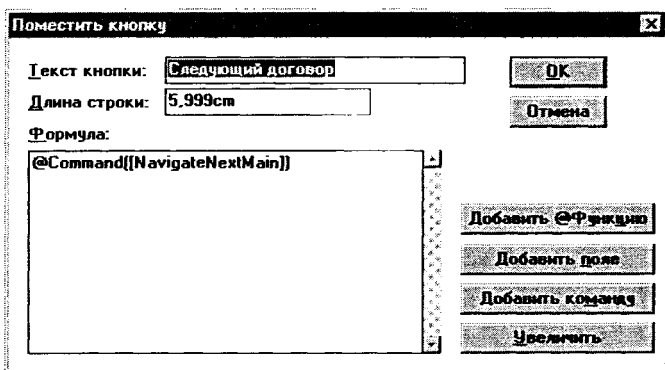


Рис. 4.20. Определение кнопки и соответствующей функции (формулы)

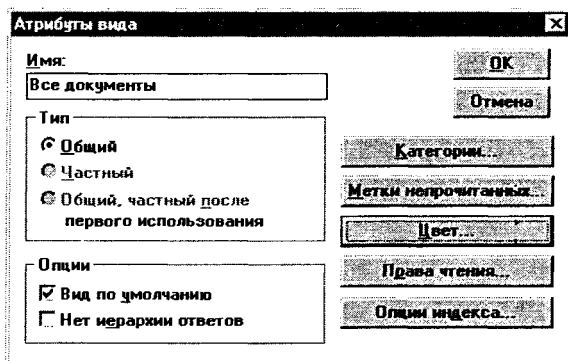


Рис. 4.21. Определение общих параметров вида

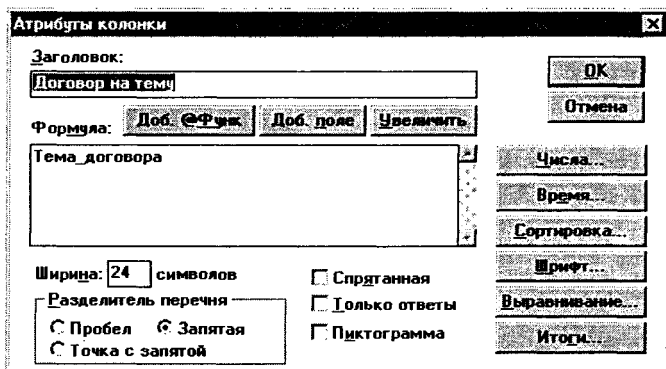


Рис. 4.22. Определение параметров полей вида

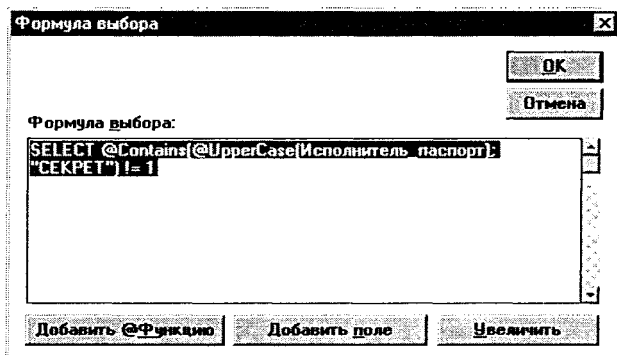


Рис. 4.23. Задание критерия выбора документов для вида

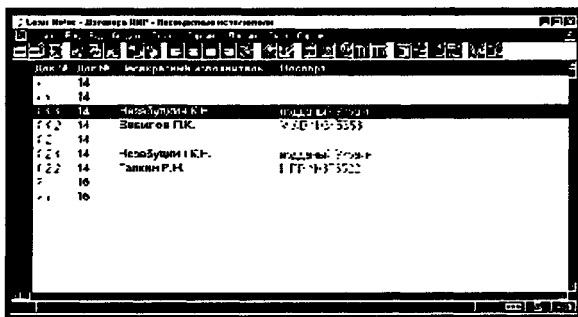


Рис. 4.24. Результат применения критерия отбора к виду документов

Приложения LN и их разработка

Приложения — это системы для коллективной работы, которые разрабатываются для координации деятельности людей, совместно реализующих конкретный бизнес-процесс. Данные процессы могут предусматривать интенсивные коммуникации (например, между заказчиками и поставщиками).

Базовыми стандартными блоками приложений Notes являются формуляры, виды и макрокоманды. Макроязык Notes дает возможность разработчикам управлять выполнением приложений Notes и программировать процессы, функционирующие на сервере. Эти средства одинаковы на всех платформах, где работает Notes. Кроме того, архитектура клиент/сервер реализуется автоматически и разработчику достаточно указать серверу на необходимость создания распределенного приложения.

Notes Help — контекстно-зависимый справочник, доступ к которому можно получить по клавише F1:

- Предметный указатель содержит перечень всех документов в базе данных справочника.
- Сообщения перечисляет все сообщения Notes.
- Функции имеет список всех функций Notes. С его помощью можно изучать синтаксис и просматривать примеры использования функций Notes. Функции Notes позволяют автоматизировать определенные операции и применять этот продукт более гибко.

Ниже приводится фрагмент текста справки о БД по договорам о НИР (функциональный аналог словаря данных).

Конспект базы данных Lotus Notes - создан в 03:10:52 20.01.2001

Общие сведения:

Путь:	D:\NOTES\NIR.nsf
Название:	Договора НИР
Дата создания:	10.01.2001 18:35:33
Дата последнего изменения:	19.01.2001 00:16:43
Дата последнего неинформационного изменения:	20.01.2001 03:10:45
Сжатие текста:	Запрещено
Индексирование:	Запрещено
Имя локального ключа к шифру:	[Не назначено]

Информация о занимаемом пространстве:

Размер на диске:	512 Килобайт
Документы данных:	12
Общее пространство:	435 Килобайт
Занятое пространство:	359 Килобайт
Свободное пространство:	17.5%

Информация о реплицировании:

ID-реплики:	C3256588:0055A708
Приоритет:	Средний
Распространить удаления:	Разрешено
Удаление устаревших документов:	Запрещено
Усекать длинные документы:	Запрещено
Реплицировать имена шаблонов дизайна:	Запрещено

Другие сведения:

Указывать в каталоге баз данных:	Разрешено
Показывать в диалоге 'Открыть базу данных':	Разрешено
Спрятать дизайн базы данных:	Запрещено
Фоновые макропрограммы:	Разрешено

Информация списка прав доступа:

Имя пользователя/группы:	-Default-
Уровень доступа:	Проектировщик
Может удалять документы:	Да
Роль(и):	[Не назначено]
Имя пользователя/группы:	OtherDomainServers
Уровень доступа:	Менеджер
Может удалять документы:	Да
Роль(и):	[Не назначено]

Информация об общем поле:

Поле:	Акт_дата
Тип данных:	Время
Формат времени:	20.01.2001
Справочное описание:	Дата акта
Тип поля:	Редактируемое
Для обновления требуется доступ редактора:	Нет
Формула значения по умолчанию:	@Now
Дата последнего изменения:	10.01.2001 23:27:09

Поле:	Исполнитель_ФИО
Тип данных:	Текст
Справочное описание:	ФИО исполнителя этапа
Тип поля:	Редактируемое
Для обновления требуется доступ редактора:	Нет
Дата последнего изменения:	10.01.2001 20:08:00
Поле:	Номер_договора
Тип данных:	Число
Формат чисел:	Общий
Процент (значение * 100)%:	Нет
Разделители тысяч:	Нет
Справочное описание:	Номер договора
Тип поля:	Редактируемое
Для обновления требуется доступ редактора:	Нет
Формула значения по умолчанию:	Номер_договора
Дата последнего изменения:	18.01.2001 23:30:43

4.3. Автоматизированные информационные системы электронной коммерции

Электронную коммерцию называют технологией третьего тысячелетия. Это новое направление экономики, которое появилось в 90-х гг. благодаря огромной популярности сети Интернет, развивается невероятными темпами.

Понятие *электронной коммерции* (Э-торговля, electronic commerce, e-commerce) появилось достаточно давно и рассматривается как один из современных инструментов бизнеса, позволяющий сократить затраты организаций-производителей, торговых фирм и покупателей при одновременном улучшении качества товаров и услуг и повышении скорости поставки.

Электронной коммерцией является такая форма поставки продукции, при которой выбор и заказ товаров осуществляется через компьютерные сети, а расчеты между покупателем и поставщиком — с использованием электронных документов и/или средств платежа. При этом в качестве покупателей товаров (или услуг) могут выступать как частные лица, так и организации.

Глобальная сеть Интернет сделала электронную коммерцию доступной для фирм любого масштаба. Если раньше организация электронного обмена данными требовала заметных вложений в коммуникационную инфраструктуру и была по плечу лишь крупным компаниям, то сегодня использование Интернет позволяет вступить в ряды «электронных торговцев» и небольшим фирмам.

Электронная витрина в WWW дает любой компании возможность привлекать клиентов со всего мира. Подобный онлайн-бизнес формирует новый канал для сбыта — «виртуальный», почти не требующий материальных вложений. Информация, услуги или продукция (например, программное обеспечение) могут быть поставлены через сеть, кроме того, весь процесс продажи (включая оплату) может происходить в онлайн-режиме.

Под определение электронной коммерции подпадают не только системы, ориентированные на Интернет, но также и «электронные магазины», использующие иные коммуникационные среды — BBS, VAN и т. д. В то же время процедуры продаж, инициированных информацией из WWW, но использующих для обмена данными телефон, факс и т. п., могут быть лишь частично отнесены к классу электронной коммерции. Отметим также, что, несмотря на то, что WWW является технологической базой электронной коммерции, в ряде систем используются и другие коммуникационные возможности. Например, запросы к продавцу для уточнения параметров товара или для оформления заказа могут быть посланы через электронную почту.

На сегодняшний день доминирующим платежным средством при онлайн-покупках являются кредитные карточки. Однако на сцену выходят и новые платежные инструменты: смарт-карты, цифровые деньги (digital cash), микроплатежи и электронные чеки.

Электронная коммерция предполагает не только онлайн-транзакции. В область, охватываемую этим понятием, необходимо включить следующие виды деятельности:

- проведение маркетинговых исследований;
- определение возможностей и партнеров;
- поддержка связей с поставщиками и потребителями;
- организация документооборота и пр.

Таким образом, электронная коммерция является комплексным понятием и включает в себя электронный обмен данными как одну из составляющих.

Рынок электронной коммерции

Одним из «измерений», характеризующих многомерный мир электронной коммерции, является тип потребителя продукции. Требования, вытекающие из задачи обслуживания частных лиц, сильно отличаются от тех, что возникают при работе с корпоративными клиентами — фирмами и компаниями. Корпоративные клиенты — бизнес-партнеры — должны быть определены заблаговре-

менно, расширение их круга — достаточно сложный процесс. В то же время, целесообразность увеличения числа частных клиентов не вызывает сомнений — чем их больше, тем выше прибыль. Так, электронный магазин, торгующий цветами без каких-либо предварительных условий должен принимать заказы от любого, кто способен заплатить.

В 2000 г. объем электронных сделок составил 6,6 млрд долларов. Несмотря на то, что эти объемы составляют сравнительно небольшую часть от общего объема розничного рынка, составляющего около 2 трлн долларов, их абсолютные значения служат достаточным основанием для инвестиций в электронную коммерцию.

Основным типом товаров, продающихся через Интернет, являются программные продукты и средства вычислительной техники. Далее следуют путешествия и финансовые услуги. В табл. 4.4. представлены сектора рынка в соответствии с двумя различными критериями: объемом продаж в денежном выражении и количеством проданных экземпляров того или иного товара (или обслуженных клиентов).

Таблица 4.4. Ранжирование рынка электронной коммерции

Место	Ранжирование	
	Денежный объем	Количество продаж
1	Недвижимость	Программное обеспечение
2	Компьютеры и комплектующие	Звуковоспроизводящая аппаратура
3	Программное обеспечение	Разное
4	Туристическое обслуживание	Компьютеры и комплектующие
5	Звуковоспроизводящая аппаратура	Туристическое обслуживание
6	Финансовые услуги	Финансовые услуги

Сектор программного обеспечения составляют главным образом не средства, ориентированные на обычного пользователя, а высокотехнологичные системы: комплексы САПР, программы для медицины и промышленности, средства разработки программного обеспечения.

Исследования, проведенные Mentis Corporation (США) дают следующую картину продаж, связанных с Интернет (табл. 4.5).

Идея организации закупок у поставщиков с использованием технологий компьютерных коммуникаций не нова. Компания по розничной торговле WalMart реализовала обмен информацией в ре-

Таблица 4.5. Оценка рынка электронной коммерции в 2000 г.

Характеристика рынка	Оценка
Общий объем Интернет-продаж	\$4,5 млрд – 6 млрд
Объем продаж на одного покупателя	\$600–800
Средний размер Интернет-транзакции	\$25–30
Общее количество Интернет-транзакций	130–200 млн
Доля онлайн-продаж товаров	60–70 %
Доля продаж доставляемых покупателю товаров	30–40 %

жиме реального времени по сетям связи между своими компьютерами и компьютерами поставщиков, обеспечив немедленное выполнение получаемых ею заказов, что привело к радикальному увеличению доходов WalMart.

Из технологий, на которых может базироваться электронная коммерция, наиболее отработанной на сегодняшний день является электронный обмен данными — EDI (Electronic Data Interchange). Этот метод кодировки последовательных транзакций и их обработки в режиме онлайн используется уже 25 лет и представляет собой индустрию объемом в 45 млрд долларов. Согласно данным Giga Information Group, только лишь компании США осуществляют в электронном режиме закупки на сумму до 500 млрд долларов в год.

Таблица 4.6 иллюстрирует рост распределения доходов по секторам рынка (источник — Forrester research).

Таблица 4.6. Объемы онлайн-продаж по секторам рынка (млн долларов)

Сектор	1996	1997	1998	1999	2000
Компьютеры и комплектующие	140	323	701	1,228	2,105
Путешествия и туризм	126	276	572	961	1,579
Развлечения	85	194	420	733	1,250
Одежда	46	89	163	234	322
Цветы и подарки	45	103	222	386	658
Продукты и напитки	39	78	149	227	336
Прочее	37	75	144	221	329
ИТОГО:	518	1,138	2,371	3,990	6,579

Информационные системы биржевой торговли (историческая справка)

Первая в РФ автоматизированная ИС для поддержки биржевой торговли была реализована на Российской товарно-сырьевой бирже (РТСБ) в 1990—1991 гг. Брокерские конторы (БК), зарегистрированные на бирже, подготавливали предложения на продажу, которые объединялись в тематические списки (табл. 4.7); эти списки затем рассылались по брокерским конторам в печатном виде, на дискетах или в режиме электронной почты.

Таблица 4.7. Фрагмент сводной распечатки предложений на продажу РТСБ

(С) Российская ТоварноСырьевая Биржа. Информация АО РТСБ не может служить основанием для котировок РТСБ (Продажа) Дата: 06.08.91 Время: 12.30 Место: 1 Секция 32											
N	БК	В Т	Т	Наименование товара	Кол-во	Ед. изм.	Цена ед.	Валю- та	Форма опл.	Место нхжд	Ср.по ст
43	Б047	9	9	Черепица биметаллическая 1200×270×0,5	42510	шт	16.00	руб	Пл. поруч	Рига	Не-медл
44	Б047	5	5	Ртуть марка Р-1	5000	кг	300.00	руб	Пред-опла	Москва	3 нед
45	Б047	4	4	Проволока алюминиевая марки А5 Ф1.6, АД1 Ф2	2800	кг	9.00	руб	Пред-опла	Москва	15 дн
46	Б047	4	4	Плита чугуи. стандов. 1200×2400×200 (литье) необработ	34	шт	3500	руб	Пред-опла	Москва	15 дн
47	Б063	9	9	Свинец УС ГОСТ 1292-81 в чушках	5	т	8000	руб	Пред-30 %	Смоленск	Не-мед
48	Б070	9	9	Профнастил оцинкованный 0.8×1000×2000	810	лист	100	руб	По факту	Самара	3 д
49	Б070	6	6	Сталь Х-12Ф-1 квадрат 50×50	20	т	1700	руб	Пред-опла	Самара	7 дн
50	Б070	4	4	Полоса стальная легированных углеродистых марок	370	т	1700	руб	Пред-40 %	Оренбург	3 кв 91
51	Б075	7	7	Круг d = 50,70,90,100 ОХ18Н10Т l = 6 м	100	т	8000	руб	Пред-опла	Моск. обл.	1 нед
52	Б076	9	9	Рельсы Р-65	140	т	2400	руб	По факту	Таллинн	Не-мед

Окончание табл. 4.7

53	Б076	9	9	Рельсы Р-50	140	т	2500	руб	по факту	Таллин	Не-мед
54	Б082	4	3	Лом и отходы пров. мед. электр. ГОСТ2112–79 в брикетах	30	т	14 500	руб	Пл. поруч	Самара	3 дня
55	Б086	1	1	Сталь кровельная листовая 08-10КП толщ. 0,8–1,2 мм	70	т	3100	руб	Пред-опла	Омск	10 дн
59	Б091	7	7	Лом алюминиев. пищевой марка АД-О, масса ящика 2,2 кг	10	т	6500	руб	Пред-100	Шахты	Не-мед

Аналогично обрабатывались заявки на приобретение товаров. Список включает следующие данные: номер заявки (лот), код брокерской конторы, номер торгов и общее количество торгов, в которых выставляется лот, наименование, количество, цена, товара, вид оплаты, место нахождения (доставки) товара и пр. сведения. Для заключения сделок БК устанавливали связь по телефону, по заключенным сделкам РТСБ выпускала итоговый отчет. В дальнейшем в состав программного и информационного обеспечения данной и подобных систем были включены шаблоны договоров и средства подготовки платежных документов.

Интернет-магазин

Рассмотрим конкретный пример создания и работы магазина в Интернете. Некоторая компания торгует товаром, который достаточно дешев по сравнению с аналогичными в других магазинах и поэтому он быстро «уходит» и успевает продаться со склада, пока покупатель ехал в магазин. В то же время количество звонков с вопросом о наличии товара на складе постоянно возрастает. В такой ситуации решением может послужить возможность просмотра цен и предварительного резервирования товара через сеть Интернет.

Модель работы выглядит следующим образом:

- покупатель видит только тот товар, который может купить и о котором может получить дополнительную информацию;
- при оформлении Интернет-заказа происходит резервирование товара;
- Интернет-заказ обрабатывается как полноправный документ в используемой системе автоматизации;

- резервируемый товар может быть возвращен на продажу в рамках логики системы автоматизации.

Предполагается, что только часть выписанных заказов будет оплачена. Поэтому по мере того, как товар продается со склада или растет время его резервирования, ответственное лицо может вернуть товар в продажу. При этом никаких вопросов покупателям не задается.

Интернет-магазин изначально задумывается как шаг навстречу покупателю, позволяющий ему самостоятельно и в удобное для него время изучить и отобрать себе товар. Интернет-магазин работает только с конечными покупателями.

Реклама продаваемых товаров проводится за рамками Интернет-магазина на отдельном web-сайте. В дополнение к этому для увеличения интереса к таким продажам покупателям, выписавшим Интернет-счета, предлагаются скидки. Анализ динамики и структуры продаж проводится средствами Oracle Browser и Oracle Discoverer.

Интернет-торговля позволяет привлекать новых покупателей и разгружает отдел безналичных расчетов. Цена эксплуатации достаточно низкая и складывается из цены на отдельный компьютер, стоимости электричества и обслуживания.

Для реализации расчетов по кредитным картам была выбрана схема, сводящая риски участников сделки к таким, которые характерны для обычных расчетов в магазине. Это подразумевает следующее:

1. Покупатель товара получает товар в обмен на правильно заполненный и подписанный slip-чек.

2. Продавец товара при оформлении Интернет-счета с оплатой по кредитной карте резервирует сумму покупки на счете покупателя. Далее продавец и покупатель вступают в контакт, и происходит обмен товара на slip-чек.

3. Банк переводит деньги со счета покупателя на счет продавца только после получения slip-чека, подтверждающего покупку от продавца. В данной схеме покупка товара в Интернет-магазине с оплатой по кредитной карте по степени риска не отличается от оплаты ужина в ресторане.

Для удобства пользователей-покупателей были введены следующие возможности:

- работа из любого WWW-браузера;
- высокая скорость работы на медленных каналах;
- стабильная работа на ненадежных каналах;
- встроенные средства поиска по товарному пространству;
- поэтапное формирование заказа за несколько сеансов связи;

- поддержка нескольких владельцев товара в рамках одного Интернет-магазина;
- автоматическое уведомление покупателя, владельца товара и владельца Интернет-магазина о факте совершения сделки;
- получение различных отчетов и справок о работе магазина по электронной почте;
- оплата товара через систему расчетов CyberPlat;
- возможность размещения Интернет-магазина на сервере провайдера услуг Интернет.

До недавнего времени отечественные электронные магазины больше напоминали виртуальные витрины: можно посмотреть товар, узнать его особенности и даже оформить заказ. Но, чтобы заплатить, приходилось идти в банк, на почту, в фирму, владеющую магазином или вызывать курьера. В результате Интернет-магазины теряли своих потенциальных клиентов, так как такие способы оплаты доставляли одни неудобства, пока не появились системы онлайн-оплаты покупок в Интернет-магазинах. Магазин, предлагающий своим посетителям возможность быстро оплатить покупку по Интернет, скорее всего, понравится покупателям, не желающим тратить время на поход по торговым точкам. Что касается магазинов, расположенных в регионах, то они смогут охватить большую территорию и привлечь удаленных покупателей.

Рассмотрим одну из типовых систем. Основное ее достоинство — безопасность как с финансовой, так и с юридической точки зрения.

Во-первых, система основана на широко распространенной технологии криптографической защиты информации с помощью цифровой подписи. Эта технология позволяет не только предохранить информацию, но и однозначно определить автора любого сообщения, а также убедиться, что сведения не были изменены во время передачи по Интернет.

Во-вторых, система основана на комплексе юридических договоров, полностью соответствующих российскому законодательству. Это дает продавцам и покупателям твердую уверенность в том, что существует четкая и определенная законом процедура решения всех спорных вопросов, возникающих при расчетах между участниками сделки. Оператором этой системы является один из крупнейших банков России.

Принцип работы системы. Если покупатель использует кредитные карты Visa, Eurocard/MasterCard, Diners Club или JCB, ему нужно, прежде всего, зарегистрироваться в системе. Для этого необязательно лично являться в банк — достаточно зайти на Web-сайт системы и заполнить форму, указав свои персональные данные (фамилию, имя,

отчество, паспортные данные, адрес электронной почты, почтовый адрес, телефон) и реквизиты своей карточки (название платежной системы, номер карты, дату окончания срока действия карты, имя держателя в том виде, как указано на карте). Вся эта информация передается в защищенном виде в банк для регистрации. Следует подчеркнуть, что за сохранность этих сведений отвечает банк, который не предоставляет их магазину при финансовых операциях покупателя. Однако до тех пор, пока клиент лично не удостоверит свою карточку, явившись в банк или в организацию-агент, он может совершать покупки лишь в ограниченном объеме.

Если у покупателя нет кредитной карты, он также удаленно регистрируется на Web-сайте системы, при этом ему автоматически открывается счет в банке. Средства на счет можно зачислить наличными через кассу банка или переводом через любой другой банк. Деньги, размещенные таким способом на счете покупателя, могут использоваться не только для расчетов в системе.

В процессе регистрации в системе пользователь устанавливает на своем ПК специальное ПО «Покупатель», необходимое для работы с системой платежей. Эту программу можно получить на магнитном носителе или загрузить по Интернет (бесплатно). Банк гарантирует необходимые телефонные консультации для установки и обслуживания этого ПО в течение срока действия договора.

Необходимо подчеркнуть, что регистрация в качестве клиента — частного лица в системе и установка необходимого ПО осуществляются в режиме on-line непосредственно через Интернет и занимают несколько минут.

Покупатель может заключить договор также в качестве корпоративного клиента. Такая услуга пригодится организациям, которые хотят предоставить своим сотрудникам возможность выполнять платежи по Интернет. В данном случае организация переводит в систему определенную сумму и распределяет ее по счетам сотрудников. Предназначенное для этого ПО «Корпоративный клиент» позволяет администратору распределять общие деньги по нескольким счетам. Сотрудники сами распоряжаются своим счетом в рамках установленного лимита.

Владельцы Интернет-магазинов, которые хотят получать оплату за товары с помощью системы, должны заключить соответствующий договор с банком и установить ПО «Магазин — расчеты». ПО «Магазин — расчеты» легко интегрируется с любым Интернет-магазином а его версии сегодня могут функционировать практически на всех используемых в России аппаратных платформах и под управлением большинства операционных систем.

Расчеты ведутся между тремя участниками сделки: покупателем, виртуальным магазином и расчетным банком. Основой взаимодействия всех трех сторон является электронный документ (чек) с цифровой подписью, так как в системе подлинность чека (принадлежность его клиенту и отсутствие в нем несанкционированных изменений) определяется по электронно-цифровым подписям (ЭЦП) при авторизации чека в банке.

Технология криптографии с помощью открытого/закрытого ключа, применяемая для цифровой подписи, служит и для защиты секретной информации (реквизитов кредитной карты и выписки со счета). Этот популярный метод предохранения информации от несанкционированного доступа со стороны посторонних лиц широко используется в электронной коммерции.

Предположим, вы уже выполнили все предварительные действия (зарегистрировались, заключили договор, получили ключи) и стали клиентом системы. Теперь начинается то, ради чего вы все это сделали — посещение электронных магазинов и оплата покупок с помощью системы.

Через Интернет покупатель подключается к Web-сайту магазина, формирует корзину заказа и посылает магазину запрос на выставление счета (рис. 4.25). Магазин в ответ на запрос покупателя направляет ему подписанный своей цифровой подписью счет, в котором указывает наименование и стоимость товара, код магазина, а также время и дату совершения операции. Покупатель подписывает этот счет своей цифровой подписью и отправляет его обратно в магазин. С момента подписания покупателем выставленного счета договор между ним и магазином считается заключенным. Чек, подписанный двумя цифровыми подписями (магазина и покупателя), направляется магазином в банк для авторизации. Банк обрабатывает подписанный чек: проверяет наличие покупателя и магазина в системе, их цифровые подписи, а также остаток и лимиты средств на счете покупателя. После этого копия чека сохраняется в базе данных. Результатом обработки является разрешение или запрет на проведение платежа. При разрешении платежа банк переводит деньги со счета покупателя на счет магазина, а затем направляет магазину разрешение на отпуск товара. Весь этот процесс занимает несколько секунд. В результате магазин оказывает покупателю требуемую услугу и продает нужный товар. При запрете платежа банк передает магазину отказ от проведения платежа с указанием причины. Как положительные, так и отрицательные ответы банка подписаны его электронно-цифровой подписью.

Следует подчеркнуть, что с покупателя за проведение платежа не взимается никаких отчислений. Кроме того, он полностью контролирует процесс совершения покупки. Клиент имеет возможность проверять свои платежи, сравнивать выписку по своему счету с результатами прохождения платежей в магазинах. Для этого он запрашивает выписку о состоянии своего счета, подписав запрос собственной цифровой подписью. Банк проверяет код покупателя и его подпись и, если все окажется в порядке, в защищенном виде направляет выписку покупателю. Получив это сообщение, покупатель в свою очередь проверяет подпись банка и декодирует выписку. Это гарантирует подлинность и конфиденциальность информации о состоянии счета. Если же покупатель захочет получить информацию о проведенных им в данном магазине платежах, он должен указать свой код в системе и посылать запрос в магазин, который и предоставит ему нужные сведения. Таким образом, система платежей является прозрачной для покупателей в том смысле, что они всегда могут получить информацию о состоянии своих платежей.

Перечисленные действия на первый взгляд кажутся слишком сложными, однако при более внимательном рассмотрении становится ясно, что без тщательной проверки и контроля покупателя и продавца не могут быть уверены в безопасности электронных сделок. Ведь именно боязнь мошенников является главной причиной, тормозящей развитие электронной коммерции.

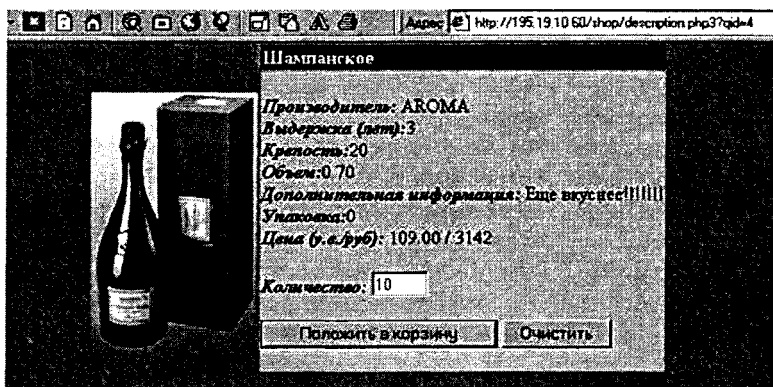
Агентская сеть. Банк создает сеть организаций-агентов, которые помогают обслуживать клиентов системы CyberPlat. Такие партнеры выполняют следующие функции:

Заверение регистрационных данных. Это необходимо, чтобы подтвердить подлинность электронной цифровой подписи или кредитной карты клиента. Для этого клиент системы должен лично явиться в организацию-агент, предъявить документ, удостоверяющий личность (паспорт или военный билет), а также принести дискету со своей электронной цифровой подписью. Эта подпись распечатывается и сверяется с зарегистрированной в системе. Полученная распечатка заверяется собственноручной подписью клиента. Если клиент использует пластиковую карту, ее реквизиты тоже сверяются с зарегистрированными, и после этого с клиента снимаются предельные ограничения по платежам, установленные при его регистрации в системе.

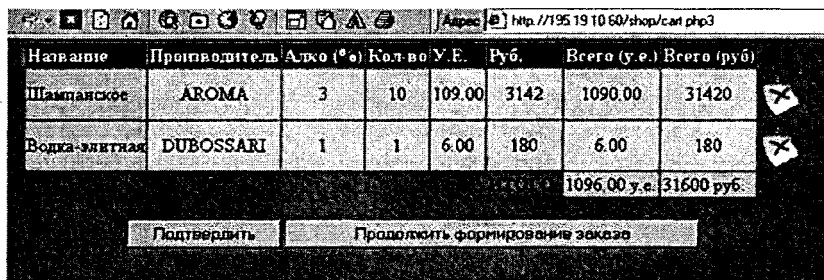
Привлечение магазинов. В роли агентов выступают компании, которые создают Интернет-магазины или сдают в аренду свой большой Интернет-магазин, так как для оплаты в этих магазинах приме-

няется указанная выше система, эти агенты получают свою долю от прибыли.

Услуги банка-агента. В этом случае клиент не переводит деньги в банк системы, а обращается в свой банк (партнер банка системы) и подает заявление о том, что он намерен тратить определенную сумму в месяц на оплату покупок по системе. В результате клиент получает такую же возможность использовать систему, как если бы он открыл счет в банке. Такого рода услуги особенно актуальны для регионов. При этом все риски берет на себя банк, однако банки-агенты должны установить корреспондентские отношения с этим банком и держать в нем определенное денежное покрытие. Помимо предоставления своим клиентам новых услуг, такой банк-агент также получает комиссионное вознаграждение от работы системы.



а)



б)

Рис. 4.25. Примеры экранов интерфейса системы «Электронный магазин»: а — просмотр и заказ товаров; б — промежуточный отчет о закупке

Платежная система «Элит»: общее описание

Платежная система «Элит» предназначена для осуществления электронных платежей по пластиковым картам систем Union Card, Visa, Master Card через сеть Интернет. Она может использоваться как самостоятельно, так и в составе систем электронной коммерции — в частности, по оформлению заказов на товары и услуги через Интернет.

Основные характеристики платежной системы «Элит»:

1. Возможность оформления платежей независимо от расположения и режима работы персонала электронного магазина, банка-эквайера. Клиент может осуществлять платежи круглосуточно, «участники» платежа (клиент, продавец, банк) могут находиться в разных городах.

2. Полное соответствие правилам банковского эквайринга.

3. Предоставление пользователю оперативной информации о результатах авторизации карточки.

4. Автоматическое направление информации о проведении платежа и авторизации карточки в торговую систему (менеджеру конкретного электронного магазина).

5. Независимость от банков-эмитентов карточек. Пользователю платежной системы «Элит», осуществляющему платежи с помощью пластиковой карты, не нужен специальный договор с каким-либо банком, система так же открыта, как и любая карточная система (Union Card, Visa, Master Card). Договоры на Интернет-эквайринг заключаются только между электронным магазином (продавцом) и банком-эквайером, производящим авторизацию пластиковых карт, по которым платежи проходят через этот магазин.

6. Простая настройка под возрастающие количественные и качественные требования бизнеса производителей и поставщиков товаров и услуг, использующих электронные магазины. Система может функционировать совместно с любыми серверами авторизации пластиковых карт.

7. Максимальная защита интересов всех участников (клиента, продавца и банка) от возникновения технических и юридических проблем.

8. Встроенные средства обеспечения конфиденциальности информации при ее хранении и передаче по каналам связи, проверки на наличие в электронных платежных документах недекларированной информации, разграничения доступа (несколько уровней доступа менеджеров платежных терминалов и пр.).

9. Надежность — система построена в соответствии с требованиями обеспечения бесперебойного круглосуточного функционирования.

10. Простота использования — все компоненты системы ориентированы на пользователей, не имеющих специальной подготовки — как клиентов, осуществляющих платежи с помощью пластиковых карт, так и менеджеров-продавцов, осуществляющих контроль проведения платежей, оформление заказов и доставку товаров.

11. Открытость — система может быть интегрирована с любыми бухгалтерско-складскими программными комплексами, дополнена модулями статистического анализа информации.

12. Возможность удаленного администрирования через Интернет — рабочие места менеджеров-продавцов и администраторов системы, так же как и клиентов, осуществляющих платежи, функционируют под управлением браузеров Интернет.

13. Идентификация клиента, а не его рабочего места — так как клиент при регистрации получает уникальные имя и пароль для каждого электронного магазина, то в дальнейшем он может осуществлять платежи с любого компьютера, имеющего выход в Интернет.

Общие принципы функционирования. Если клиент желает расплатиться с помощью пластиковой карты за очередной заказ, оформленный в конкретном электронном магазине, то он вызывает платежную систему «Элит» и заполняет предложенную форму электронного платежного поручения, после чего отправляет ее на обработку (рис. 4.26).

По указанию менеджера-продавца, контролирующего проведение оплаты заказа, система совместно с сервером проводит авторизацию пластиковой карточки и блокирует средства (в объеме стоимости заказа) на карточном счете клиента.

Менеджер-продавец контролирует все этапы прохождения платежа, в необходимых случаях (недостаточно средств на счете, неудачная попытка авторизации, отрицательный результат авторизации) отказывая клиенту в принятии платежа (при этом доставка/отгрузка товаров по данному заказу не производится).

Менеджеру-продавцу предоставляются средства уведомления клиента о принятии платежа или отказе от оплаты.

При желании клиент может следить за процессом авторизации карточки с помощью предназначенных для этого средств платежной системы. Для этого он должен указать уникальный номер платежной операции (транзакции). Уникальный номер генерируется системой для каждой платежной операции в момент отправки клиентом электронного платежного поручения на обработку.

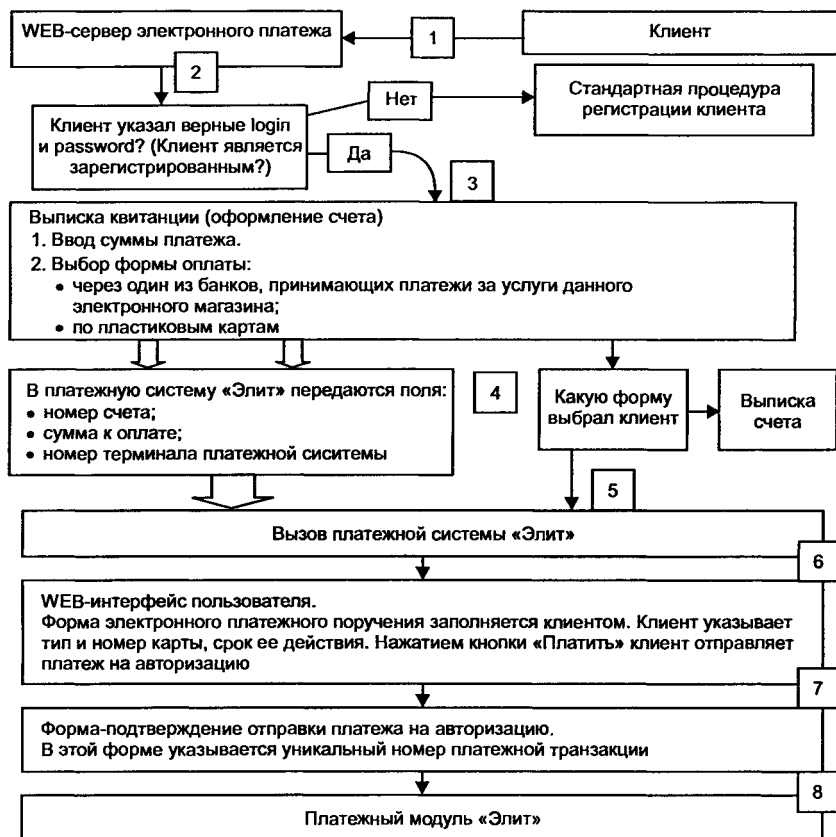


Рис. 4.26. Схема инициализация платежа

При положительном результате авторизации пластиковой карты менеджер-продавец оформляет доставку заказа клиенту (доставка осуществляется собственной службой доставки, службой экспресс-доставки, почтовым отправлением).

При успешной доставке клиенту оформляется документ, подтверждающий получение заказа (документ, подписанный клиентом — слип/карточка службы экспресс-доставки/почтовое извещение).

Поэтапный контроль прохождения платежа и процесса доставки заказанных услуг, осуществляемый менеджером-продавцом, обеспечивает важное преимущество платежной системы «Элит» — перечисление заблокированных средств с карточного счета клиента на

счет продавца осуществляется по указанию менеджера-продавца *только* после фактической доставки товара. В случае отказа клиента от получения доставленного товара менеджер-продавец в платежной системе осуществляет операцию разблокирования средств на карточном счете клиента. Такая схема позволяет для всех участников операции (клиент, продавец, банк) минимизировать или свести к нулю риски, связанные с проведением платежей через Интернет.

Взаимодействие элементов. Принята следующая последовательность операций при оформлении и прохождении платежа.

1. Регистрация клиента. Если клиент ранее не регистрировался в системе, то ему необходимо пройти полную процедуру регистрации в электронном магазине. При этом клиенту либо будет предложено ввести свои имя и пароль, либо они будут автоматически присвоены.

2. На Web-странице формирования заказа клиент окончательно проверяет его. При выборе способа оплаты «по пластиковой карте» активизируется платежная система «Элит», в которую передаются значения «закрытых» полей электронного платежного поручения — номер счета, сумма к оплате, номер терминала платежной системы «Элит» (зарегистрированного в Автобанке для данного электронного магазина).

3. Заполнение формы электронного платежного поручения. После инициализации платежной системы «Элит» в Интернет-браузере клиенту предлагается форма электронного платежного поручения, в которой требуется указать реквизиты пластиковой карты (вид карты — Visa, Master, ..., номер и срок действия, PIN-код не указывается). Если клиент укажет свой E-mail, то по результату авторизации платежная система автоматически направит клиенту соответствующее сообщение.

4. Отправка электронного платежного поручения на обработку. После ввода необходимых данных клиент «нажимает» кнопку «Оплатить». При этом электронное платежное поручение отправляется на обработку в платежной системе «Элит», клиенту предьявляется web-форма сообщения, подтверждающего факт отправки очередного платежа на обработку. В этой форме клиенту предьявляется уникальный номер платежной транзакции, при желании клиент может запомнить этот номер и затем во время прохождения платежа следить за этим процессом с помощью предназначенного для этого Web-интерфейса.

5. Передача данных электронного платежного поручения по каналам связи. Данные электронного платежного поручения в защищенном виде (конфиденциальность обеспечивается программно-аппаратными средствами платежной системы «Элит») последователь-

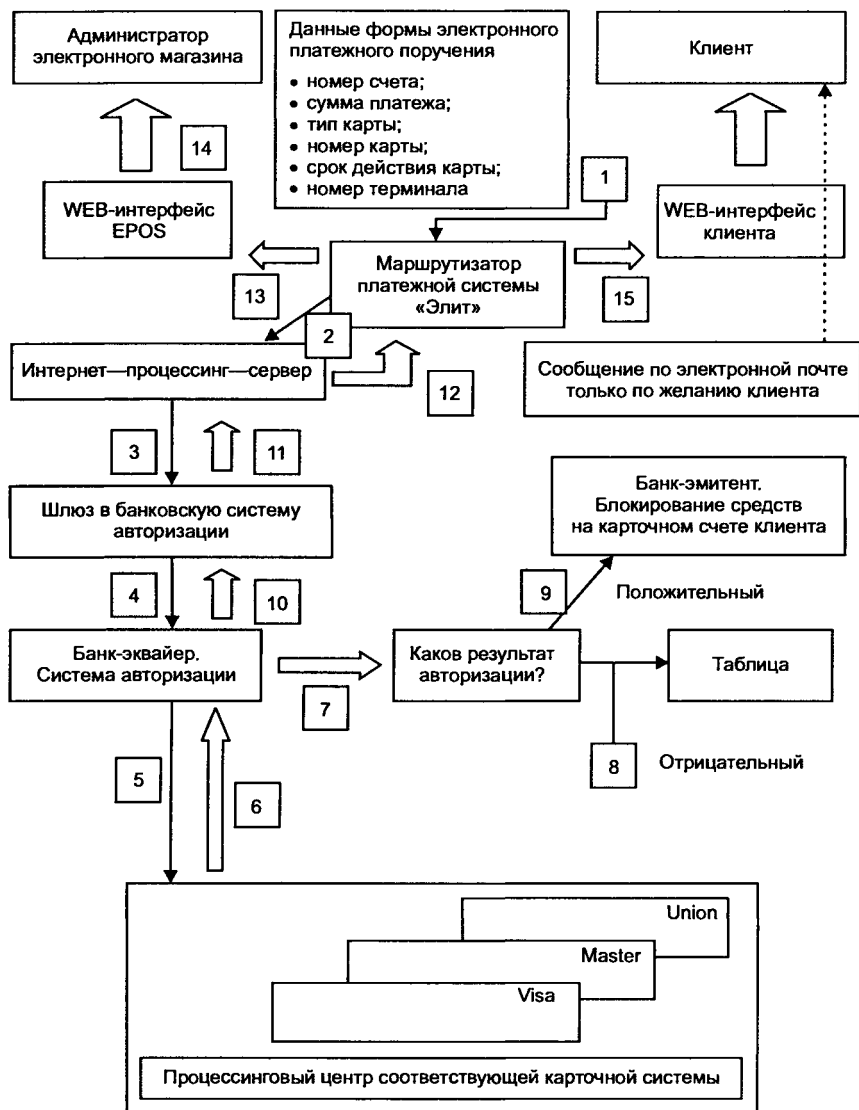


Рис. 4.27. Проведение авторизации платежа:

- Прохождение запроса на авторизацию
- Прохождение результата авторизации

но предаются на Internet Processing Server платежной системы «Элит», а затем — на процессинговый сервер Автобанка (процессинговый центр Автобанка).

6. Авторизация. Процессинговый сервер банка-эквайера (Автобанка) осуществляет запрос в процессинговый центр соответствующей международной карточной системы (рис. 4.27).

Сведения о положительном или отрицательном результате авторизации последовательно возвращаются в:

- процессинговый сервер банка-эквайера;
- шлюз между банковской системой авторизации и платежной системой «Элит»;

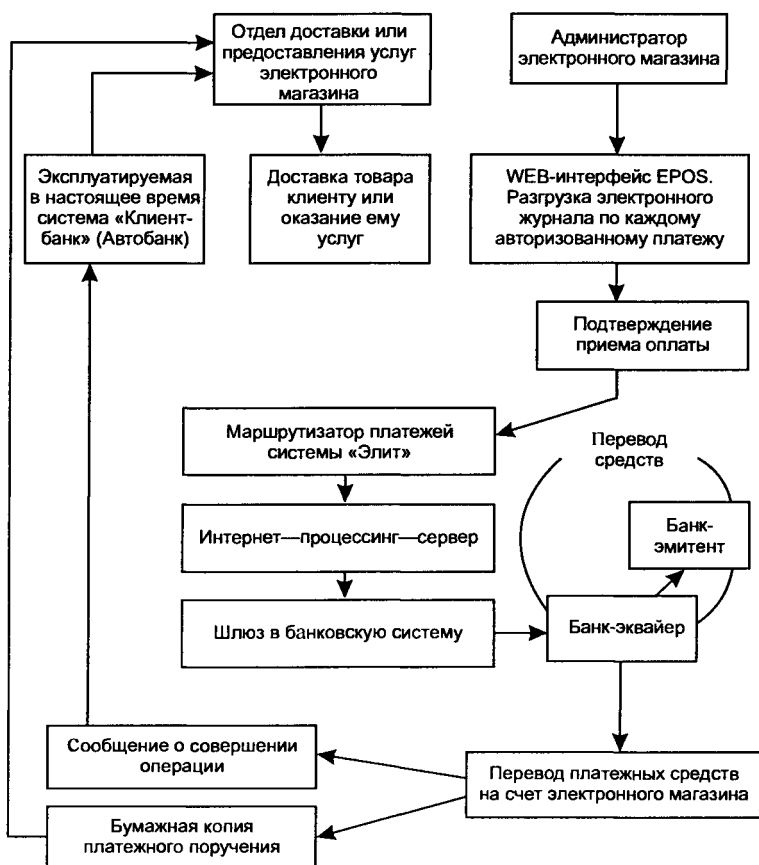


Рис. 4.28. Перевод денежных средств на счет электронного магазина

- Internet Processing Server платежной системы «Элит»;
- маршрутизатор платежной системы «Элит»;
- в Web-интерфейсы EPOS (терминал платежной системы «Элит» — рабочее место администратора) и клиента.

При положительном результате авторизации средства в размере суммы платежа блокируются банком-эмитентом на карточном счете клиента.

7. Перевод денежных средств на счет электронного магазина. В установленный период администратор проводит операцию разгрузки электронного журнала платежей. Данная операция заключается в подтверждении администратором факта успешного прохождения каждого положительно авторизованного платежа и обеспечивает зачисление этих средств на счет электронного магазина (рис. 4.28).

8. Подтверждение администратора является сигналом к началу процесса перевода денежных средств (суммы платежа) с карточного счета клиента на счет электронного магазина.

Далее перевод средств осуществляется в соответствии с правилами банковской системы.

9. Администратор получает подтверждение о переводе средств на счет электронного магазина различными традиционными средствами (получив копию платежного поручения, а так же с помощью системы клиент—банк). Далее администратор в соответствии с принятой процедурой зачисляет поступившие средства на учетный счет клиента.

Российская Торговая Система

Российская Торговая Система — первая и до сих пор единственная реально функционирующая система внебиржевой торговли ценными бумагами в России. Более того, Российская Торговая Система — самая крупная торговая площадка, специализирующаяся исключительно на работу с корпоративными ценными бумагами.

Торговля в РТС осуществляется следующим образом. Участники торгов через удаленные терминалы вводят в центральный компьютер свои предложения на покупку и продажу акций. Введенные предложения отображаются на терминалах всех участников торгов. Если кто-нибудь из участников торгов пожелает совершить сделку на предложенных условиях, то он связывается с трейдером, выставившим заявку, по телефону и договаривается с ним о сделке. Таким образом, РТС в современном виде — это своеобразная доска

объявлений, доступ к которой регламентируется определенными правилами.

Необходимо отметить также, что достаточно жестко регламентирован сам доступ к торговле в РТС: любая компания, желающая присоединиться к системе, должна прежде всего быть членом НАУФОР (Национальной Ассоциации Участников Фондового Рынка) и принять торговое соглашение партнерства, в котором раскрываются основные обязательства и принципы торговли через РТС. Ключевыми в данном документе являются:

1) положение об обязанности исполнения и гарантий сделки на определенных условиях. Данное положение, хотя не в состоянии гарантировать, призвано максимально защитить участников Партнерства от рисков, связанных с неисполнением сделки;

2) положение о претензионном урегулировании споров через Третейский суд НАУФОР. Оно гласит о том, что все участники обязуются все противоречия решать путем первоначального обращения в Третейский Суд НАУФОР; таким образом, РТС берет на себя роль арбитра в урегулировании спорных вопросов.

Ключевые моменты правил торговли в РТС.

1. Все предложения (котировки), введенные в систему, являются «твердыми», т. е. обязательными для исполнения.

Это принципиальное отличие РТС от подобных ей электронных систем, не являющихся по сути торговыми, а лишь информационными.

2. Минимальный объем лота, выставяемого в системе, — 10 тыс. долл. США.

Минимальный размер лота, однако, не означает, что сделка не может быть заключена на меньшую сумму. Следует отметить также, что в РТС-2 минимальный размер лота составляет 5 тыс. долл.

3. Участники торговой системы разделены на несколько категорий, среди которых выделяются:

- *члены торговой системы с просмотрным режимом*, которые торгуют на общих основаниях, однако лишены возможности самостоятельно выставять котировки. Для получения данного статуса необходимо выполнить минимальные требования по минимальному ограничению уставного и ликвидного чистого капитала, наличия лицензии профессионального участника рынка ценных бумаг;
- *члены торговой системы с торговым режимом*, обладающие возможностью самостоятельно выставять котировки; однако для получения торгового статуса необходимо соблюдать более жесткие условия касательно минимального размера ликвидно-

го чистого капитала и наличия в персонале компании профессиональных прошедших аттестацию НАУФОР трейдеров;

- *маркет-мейкеры*, одной из основных задач которых является поддержание ликвидности рынка. В обязанности маркет-мейкеров входят постоянное поддержание двусторонних котировок (на покупку и на продажу) по акциям как минимум трех эмитентов и односторонних котировок (либо на покупку, либо на продажу) как минимум двух эмитентов; объем выставляемых ими лотов должен быть не менее 30 тыс. долл. США. Маркет-мейкеры имеют определенные льготы. Так, им предоставляется право, если к ним обратится «простой» участник торгов, потребовать совершения сделки на условиях предоплаты вместо стандартных условий предпоставки.

Начиная с момента начала торговли в РТС в 1995 г. был введен индекс РТС, сводный показатель деловой активности и настроения рынка, рассчитываемый по средневзвешенным ценам дневных сделок по акциям, входящим в котировальный лист первого уровня. Индикаторы, удобные и привычные для всех фондовых рынков, позволяют участникам торговли оценить картину, происходящую на рынке в целом.

Технологическое развитие РТС. Этап «ПОРТАЛ». Торговая система ПОРТАЛ была запущена в декабре 1994 г., и в ее работе участвовали шестнадцать компаний. Программное обеспечение было заимствовано у американской фирмы NASDAQ (информационное обеспечение внебиржевого рынка) и доработано под российские требования, в частности, появились дополнительные поля по регистрации и расчетам сделок.

Этап «РТС». С 5 июля 1995 г. терминалы РТС появляются не только в Москве, но и в Санкт-Петербурге, Новосибирске и Екатеринбурге. Становится очевидной необходимость доработки программного обеспечения, прежде всего для оптимизации использования ресурсов сервера и повышения надежности работы.

Этап «Continuum». В феврале 1996 г., когда число терминалов превысило 150, был установлен новый более мощный сервер Stratus модели Continuum 610, который работает и сейчас. В начале 1996 г. было принято решение о разработке нового программного обеспечения РТС.

Вот несколько цифр, характеризующих центральный сервер и работу ПО:

- количество транзакций — более 1 000 000 в день;
- количество одновременно работающих процессов — более 1200.

Программно-техническое обеспечение РТС имеет иерархическую структуру (рис. 4.29). В центре системы находится сервер Stratus Continuum, на который ложится обработка информации о котировках и сделках. Все проверки прав доступа участников, контроль за правильностью вводимых торговых лотов, сведение отчетов о сделках, ведение журналов ввода котировок и сделок, первичная обработка статистической информации проводится на центральном сервере Stratus. Этим обеспечивается надежность работы торговой системы.

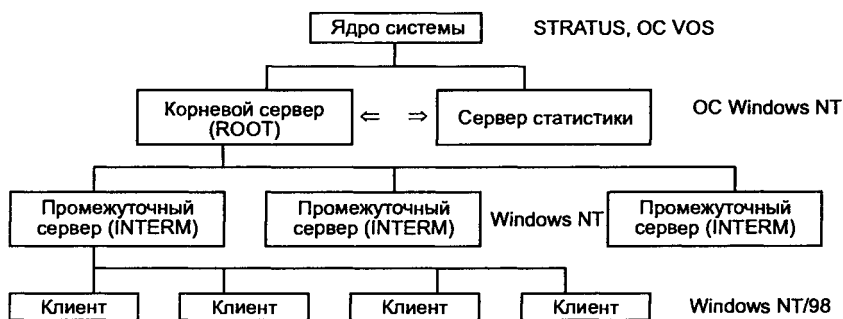


Рис. 4.29. Архитектура программного обеспечения РТС

Сервер, обозначенный на диаграмме как ROOT, выполняет роль интерфейса между центральным сервером и рабочими станциями. Его функциями являются:

- проверка прав доступа и обеспечение защиты передаваемой между сервером и рабочими станциями информации;
- получение от центрального сервера информации о котировках и сделках и перевод ее в формат таблиц реляционных баз данных, необходимый для дальнейшей репликации на рабочие станции РТС.

Сервер ROOT, как и промежуточные серверы INTERM, не производит обработку вводимых данных. Таким образом, сообщение, сформированное на рабочей станции РТС, без каких-либо изменений доставляется на центральный сервер Stratus, где обрабатывается. При этом достигается высокое быстродействие и надежность работы торговой системы.

Промежуточный сервер INTERM является в сущности копией сервера ROOT, из которого исключены функции проверки прав доступа и обмена сообщениями с сервером Stratus. За счет использования нескольких серверов INTERM достигается масштабирование

ние системы и оптимизация потоков данных. Промежуточный сервер может обслуживать как регион, так и одну или несколько компаний-участников и позволяет оперативно (без необходимости пересылки всей информации из центра) восстанавливать актуальность базы данных на рабочей станции РТС. Промежуточные серверы используют также и для замены клиентского программного обеспечения.

Рабочая станция РТС обеспечивает прием поступающей от центрального сервера (через ROOT и INTERM) информации на персональный компьютер клиента, ее отображение на экране и резервное копирование в локальную базу данных, а также формирование сообщений об изменении котировок и отчетах о сделках. Рабочая станция РТС проводит предварительную проверку информации, содержащейся в сообщении. Окончательная проверка производится на центральном сервере Stratus, после чего происходят соответствующие изменения в базе данных.

Функциональные возможности рабочей станции РТС (RTS). Экран рабочей станции разделен на семь основных областей и две вспомогательные (рис. 4.30).

Вспомогательные области:

- линейка меню (Menu);
- строка отображения состояния программы (Status Line).

Основные области:

1. *Сводная информация (Market Minder)* — сводные списки лучших котировок на продажу и покупку по интересующим пользователя группам эмитентов бумаг. Помимо цифровых значений, отображаются тенденции к понижению или повышению цен покупки/продажи.

Menu		
Market Minder	[Quote Retrieval] [Issues][Participans][Trades] [Trade Reports][News] Detailed information	Pending Trades
Company Minder	Hotline [Quote Updates] [Trade Report][Communication] Dealer	Selector

Рис. 4.30. Экран рабочей станции РТС

2. *Котировки компаний-участников торгов (Company Minder)* — полные списки котировок по всем эмитентам, котируемых интересующими пользователя участниками.

3. *Детальная информация (Detailed Information)*, где возможен просмотр следующих данных:

- список котировок по интересующим пользователя эмитентам (Quote Retrieval);
- полный список котируемых в РТС эмитентов (Issues);
- полный список участников РТС (Participants);
- список последних сделок, зарегистрированных в РТС (Trades);
- список сделок компании пользователя (Trade Reports);
- новости РТС или агентства «Интерфакс» (News).

Эти виды информации доступны не одновременно, а в зависимости от выбора рубрик (закладок).

4. *Область отображения важных сообщений (Hot Line)*.

5. *Область диалога пользователя (Dealer)*, предназначена для следующих действий:

- ввод и изменение котировок (Quote Update);
- просмотр отчетов о сделках (Trade Report);
- переговоры между участниками торгов на РТС (Communications).

6. *Список неподтвержденных сделок (Pending Trades)* — отображает отчеты по сделкам, введенным пользователем, но неподтвержденным контрагентами, или наоборот.

7. *Коммутатор переговоров (Selector)* — окно, дополняющее функцию Communication.

Расположение окон фиксированное, размеры же могут изменяться за счет перемещения подвижных перегородок. Многофункциональные окна имеют несколько рубрик (закладок), обеспечивающих доступ к выбранной информации; выбор режима осуществляется одним из следующих способов:

- выбором курсором манипулятора мышью корешка закладки;
- нажатием функциональной клавиши;
- выбором из меню (строка Menu).

Основным методом взаимодействия пользователя с системой является *буксировка с помощью манипулятора мышью*. При этом выбирается (указывается) объект в одном окне и перемещается в другое окно (или в пределах этого же окна). Общий принцип разрешенных перемещений таков: если объект содержит информацию об эмитенте, его можно поместить в окна, где содержатся данные об эмитентах (Market Minder или Quote Retrieval) и т. п. Некоторые из разрешенных буксировок приводятся в табл. 4.8.

Таблица 4.8. Операнды и результат некоторых из «буксировок» в RTS

Откуда	Куда	Перемещаемый объект	Результат
Issues	Market Minder	Элемент списка эмитентов	Добавление нового эмитента в список лучших котировок
Participants	Company Minder	Участник РТС	Замена списка котировок одной компании на список выбранной компании
Market Minder	Quote Retrieval	Элемент списка лучших котировок по эмитентам	Отображение полного списка котировок по выбранному эмитенту
Company Minder, Quote Retrieval	Quote Update	Элемент списка котировок	Заполнение формы Quote Update данными о котировке по выбранному эмитенту
Quote Retrieval	Trade Report	Элемент списка котировок	Заполнение формы Trade Report данными о предполагаемой сделке
Quote Retrieval Trade Reports	Communication	Элемент списка котировок или сделок	Выбор компании для переговоров
Selector	Communication	Элемент списка коммутатора	Включение в режим переговоров

После загрузки программа RTS реализует процедуру импорта БД с центрального компьютера РТС, после чего содержимое БД пользователя отображает состояние центральной БД. Пользователь приступает к работе в *основном рабочем режиме* (см. рис. 4.30), что позволяет:

- оперативно добавлять новые эмитенты в списки Market Minder;
- выбирать эмитенты из Market Minder для детального просмотра котировок в окне Quote Retrieval;
- выбирать из окна Quote Retrieval наименования просматриваемых в Company Minder компаний;
- изменять котировки в окне Quote Updates;
- вводить в систему отчеты о сделках в окне Trade Reports. Все эти операции выполняются путем буксировки объектов.

Окно Market Minder включает следующую информацию (рис. 4.31, а):

- аббревиатуру (символ) эмитента (Issue);
- лучший на данный момент запрос на покупку (BID) по данному эмитенту;
- лучшее на данный момент предложение (ASK) о продаже по эмитенту.

Тенденции по повышению или понижению котировок отражаются путем изменения цветов цен (зеленый цвет — повышение, красный — понижение).

Issue	BID	ASK
EESR	0.3961	0.3975
GUMM	4.0000	4.5900
KNFT	5.5000	6.5500
KOLE	0.5410	0.5650
LKOH	24.2000	24.4000

а)

Issue	BID	ASK
BEGY	0.7900	0.8250
BEGYP	0.7600	0.9900
EESR		0.3990
ESIR	1.7500	2.0500
GAZA		132.000

б)

Рис. 4.31. Фрагменты окон Market Minder (а) и Company Mindek (б)

Окно *Company Minder* может поддерживать несколько списков сводных котировок по компаниям и включает три колонки (рис. 4.31, б):

- символ (аббревиатуру) эмитента (Issue);
- котировку данного участника на покупку (BID) по каждому из котируемых компанией эмитентов;
- котировку выбранного участника на продажу (ASK) по каждому из котируемых компанией эмитентов.

Окно *Detailed Information* в режиме *Quote Retrieval* содержит подробную информацию по интересующим пользователя эмитентам. Описание каждого эмитента (в окне может отображаться 1—6 в зависимости от заданного режима), включает следующие данные (рис. 4.32):

- символ (аббревиатуру) эмитента;
- цену, дату и объем последней сделки по данному эмитенту (Last Sale);
- цену последней сделки на момент закрытия предыдущей торговой сессии (Previous Close);
- суммарное количество акций, проданных в течение текущей торговой сессии (Today's Volume);

EESR	Last Close 0.3950 05 Sep 97 100T		
	Prev Close 0.3940 Today's Vol 42894T		
UNIB 0.3951 B 3 100T	CSFB 0.3975 S 3 500T		
BROK 0.3950 B 3 100T	GLBX 0.3975 B 3 200T		
FTCIM 0.3950 B 3 70T	UNIB 0.3977 J 3 100T		
BRIC 0.3947 B 3 100T	TORI 0.3978 B 3 100T		
ERGO 0.3945 B 4 100T	RUSO 0.3980 J 3 200T		
TRDY 0.3945 S 3 100T	RMM 0.3985 B 3 100T		

Рис. 4.32. Окно просмотра списка котировок (Quote Retrieval)

- два списка текущих котировок, представленных отдельно на покупку (BID — слева) и продажу (ASK — справа).

Каждая строка включает:

- символ предложившего котировку участника;
- цену, установленную участником;
- условия платежа (вид валюты: В — рубли, S — доллары, J — любая валюта; и срок перерегистрации сделки в днях);
- размер предлагаемой сделки в акциях.

Списки отсортированы от лучшей котировки к худшей.

Окно Detailed Information в режиме Issues отображает подробную информацию по котируемым в РТС эмитентам (рис. 4.33):

- символ (аббревиатуру) эмитента латинскими буквами (Issue);
- полное название эмитента на русском языке (Name);
- регистрационный код эмитента (Reg.Number);
- код валюты, в которой котируется эмитент (Curr.).

Issue	Name	Reg. Number	CURR
BEGY	Башкирские ОАО энергетики	RU2354231	USD
BEGYP	Башкирские ОАО энергетики прив.	RU2345623	USD
CHGZ	АООТ Черногорнефть	RU5647182	USD
CHNGP	АООТ Челябинерго	RU6789431	USD
EESR	РАО Единая Энергетическая Система	RU5643241	USD
ESIR	АО Электросвязь Иркутской обл.	RU7648390	USD

Рис. 4.33. Просмотр информации по эмитентам ценных бумаг (Issues)

Окно Detailed Information в режиме Participants отображает подробную информацию по зарегистрированным в РТС компаниям-участникам (рис. 4.34):

- символ (аббревиатуру) компании-участника РТС (Name);
- телефон (Phone);
- статус участника в РТС с одним из трех возможных значений: Q (Query Only), Reporting и Dealer;
- полное наименование компании-участника на русском языке (Name).

Name	Phone	Type	Full Name
ABSIM	7-095-785-02-42	Dealer	Абсолют Инвест ИК ЗАО
ACADM	7-095-247-91-35	Dealer	Академия ИК ЗАО
ACCOM	7-095-291-50-36	Reporting	Аккорд М АОЗТ
ACPTP	7-8469-37-70-75	Reporting	Акцент ЛТД ЗАО
ADVNS	7-812-327-78-56	Dealer	Эдванс ИК АОЗТ

Рис. 4.34. Просмотр информации по участникам РТС (Participants)

Информация отсортирована в алфавитном порядке в соответствии с полем Name.

Окно Detailed Information в режиме Trades представляет подробную информацию по совершенным сделкам (рис. 4.35):

- символ эмитента (Issue);
- дату и время регистрации сделки (Moment);
- цену (Price);
- объем сделки (Qty);
- валюту (Curr.);
- статус сделки (Stat).

Issue	Moment	Price	Qty	Curr	Stat
EESRP	09/09/97 20:27	0.3140	100000	USD	COM
LKOH	09/09/97 19:06	24.7500	5000	USD	
EESR	09/09/97 18:43	0.4000	100000	USD	COM
SVER	09/09/97 18:38	1.0000	50000	USD	COM
EESR	09/09/97 18:35	0.4080	100000	USD	COM
EESR	09/09/97 18:28	0.4127	100000	USD	COM
MSNG	09/09/97 18:17	1.6050	50000	USD	

Рис. 4.35. Окно просмотра списка совершенных сделок (Trades)

Полный список можно отфильтровать по эмитенту, выбрав его из списка по коду Issue.

Окно Detailed Information в режиме Trade-Reports содержит информацию о сделках, в которых компания пользователя является одним из контрагентов сделки:

- название эмитента (Issue);
- вид операции: (Act) — символы, P — покупка или S — продажа, которые могут дополняться символом Q в том случае, если сделка заключена на условиях, указанных в котировке;
- статус сделки (Stat) — аббревиатура, имеющая следующие возможные значения:

ADV — сделки, введенные контрагентом, но не подтвержденные пользователем;

COM — сделки, подтвержденные обеими сторонами;

CXL — ранее подтвержденная сделка, по которой обе стороны договорились об ее снятии;

CPX — ранее подтвержденные сделки, аннулированные контрагентом;

DEL — сделки, аннулированные пользователем;

CLR — сделки, введенные контрагентом, но снятые до подтверждения пользователем;

UNC — сделки, введенные пользователем, но не подтвержденные контрагентом;

VXL — ранее подтвержденные сделки, аннулированные пользователем. Сделки, требующие подтверждения (UNC, ADV), отображаются также в окне Pending Trades;

- символ (аббревиатуру) контрагента (Contra);
- цену (Price);
- объем сделки (Qty);
- дату проведения расчетов по сделке (Settle);
- номер транзакции, присваиваемый системой при регистрации сделки (Trans);
- дату и время регистрации сделки (Moment);
- общую сумму сделки (Amount);
- валюту (Curr).

Полный список отчетов по сделкам может быть отфильтрован как по эмитентам, так и по контрагентам.

Форма для ввода и коррекции котировок (режим Quote Update окна Dealer). Форма ввода информации имеет поле для выбора эмитента и поля для ввода котировок, объединенных в группы BID и ASK. Каждая группа состоит из редактируемой строки, отображающей текущие данные о котировке, и редактируемых полей, куда вводятся обновленные данные — цена, код валюты, срок расчетов, объем сделки.

При изменении котировки может вводиться как новая абсолютная величина цены, так и относительные изменения с помощью кнопок типа [-0.01], [+0.01] и [-5%] или [+5%].

Информация о приеме изменений системой появляется в строке статуса в виде сообщения *accepted*, сопровождаемого датой и временем приема котировки, при наличии ошибок при заполнении полей формы система не принимает измененную корректировку, сообщая *reject invalid update*.

Форма для ввода отчета по сделке (режим Trade Report окна Dealer). В PTC предусмотрены следующие варианты заключения сделок:

- в точности на условиях, обозначенных компанией, предложившей котировку (сделка по котировке). Необходимо отметить отличие сделки по котировке от биржевой торговли. Существенным элементом является то, что в PTC участник при совершении сделки имеет право выбора контрагента, а не совершает ее анонимно, как это происходит на бирже;

- продукт переговоров между участниками РТС (договорные сделки);
- между участниками РТС и клиентами, не имеющими в РТС полномочий дилеров (односторонние сделки).

Форма позволяет вводить отчеты по сделке, подтверждать отчеты, введенные контрагентами, удалять ранее введенные в систему сделки. Форма включает следующие элементы (поля):

- вид операции (I sell — продажа, I Purchase — покупка);
- список для вывода символа эмитента;
- список для выбора имени контрагента (Contra);
- поле для ввода цены (Price);
- поле для ввода размера сделки (Quantity);
- поле статуса сделки (трехбуквенная аббревиатура, см. выше).

Здесь добавляются следующие коды:

- B/Q — сделка по котировке;
- NEW — новая, еще не введенная в систему сделка;
- поле для ввода даты расчетов (Settle);

Форма содержит также следующие кнопки:

[Submit] — ввод отчета по сделке в систему или подтверждение отчета, введенного контрагентом;

[Edit] — редактирование сделки по котировке (B/Q) или ранее введенного отчета о сделке, не подтвержденного контрагентом (UNC);

[Delete] — удаление отчета по сделке;

[Connect] — связь с контрагентом, переход в режим Communication;

[Reset] — очистка формы.

Таким образом, данный режим работы обеспечивает следующие операции пользователя:

- ввод сделок по котировке;
- ввод договорной сделки;
- ввод односторонней сделки;
- подтверждение сделки, введенной контрагентом;
- внесение изменений в отчет по сделке;
- удаление сделки.

Режим переговоров (окна Communication и Selector). АРМ предоставляет возможность ведения переговоров между участниками РТС в режиме обмена текстовыми сообщениями; реализуются следующие функции:

- выбор абонента для соединения (из списка участников РТС);
- установление соединения;
- прием—передача текстовых сообщений;
- временный разрыв соединения;

- одновременный диалог с несколькими абонентами;
- возвращение к прерванному разговору;
- разрыв соединения с абонентом.

Мировая практика показывает, что роль внебиржевой торговли в условиях развития технических средств все более возрастает, и, по-видимому, это является одной из причин того, что практически весь рынок корпоративных ценных бумаг является внебиржевым.

4.4. Автоматизированные библиотечные информационные системы (АБИС)

Технологический процесс и автоматизация библиотек, библиотечного обслуживания, библиотечных процессов

В общем виде технологический процесс библиотеки включает следующие основные звенья (рис. 4.36):

- заказ литературы и комплектование фондов;
- обработку литературы — каталогизация, описание, индексирование, составление каталога;
- обслуживание пользователей — регистрация читателей, прием заказов на литературу, выдача/прием литературы.

Эти процессы используют как правило, следующие основные массивы информации (файлы, базы данных):

1. Электронный каталог (реквизиты изданий).
2. Читатели (реквизиты зарегистрированных читателей, сведения о выдаче-возврате литературы).
3. Заказы (очередь текущих заказов на выдачу литературы).
4. Поставщики изданий (издательства, магазины) и библиотеки, с которыми осуществляется взаимодействие по межбиблиотечному абонементу (МБА).
5. Словарные массивы.



Рис. 4.36. Основные технологические процессы и информационные ресурсы АБИС

На рис 4.37—4.40 приведены структуры типичных записей из файлов (БД) регистрации читателей, статей, периодических изданий, книг.

Перечисленные процессы являются укрупненным взглядом на деятельность АБИС и при более подробном рассмотрении могут быть выделены функциональные группы задач, которые принято группировать в функциональные программные модули, интерфейсы или АРМы (автоматизированные рабочие места), предназначенные

Имя читателя :Bivins, Kristina
Форма обращения :Kristina Bivins
Степень: Distributor Manager
Отдел: Sales & Marketing
Сектор : Distributor Network
Номер комнаты: 14
Местонахождение: Suite 203 — West
Номер читателя: 2045
Класс читателя: staff
Максимальный период выдачи: 21
Ограничение на выдачу: 100
Стандартный код штрафа: 1
Локальный максимальный период выдачи: 28
Локальное ограничение на кол-во изданий: 50
Всего изданий выдано: 2
Маршрутный лист: ASLIB Information/1
Маршрутный лист: C & L Applications/1
Имя профиля ИРИ: Bivins, Kristina - LAW -
 Serials

Заглавие статьи: The function of the library system specification
Автор: Hamilton, R
Классификационный индекс: 021.32
Название периодического/продолжающегося издания: Special Libraries
Номер тома/части: 19
Дата: 1999
Число экземпляров: 1
Инвентарный номер: A126
Шифр хранения: Special Collection Room
Расстановочный шифр: LIB
Класс издания: special collection
Стандартный код штрафа: 3
Максимальный период выдачи: 1
Дата каталогизации:28/09/2000

Рис. 4.37. Запись файла регистрации (картотеки) читателей на примере АБИС TINLIB

Рис. 4.38. Запись электронного каталога (регистрация статей) на примере АБИС TINLIB

Периодич./Продолж. издание: Automated Library and Information Systems
Альтернативное заглавие: Program
Полочный шифр: COM/SER
Наличие издания: 1960 - 1999
Подписка на издание Y/N?:N
Номер копии: 1
Шифр хранения: Reading Room
Примечание: Permanent
Указания по переплету: Bound in yearly volumes
Периодичность: Monthly
Кл. слова: Automated; Information; Library; Systems
Дата каталогизации: 28/07/2001
Дата редактирования: 24/08/2002
Имя каталогизатора: IME DEV

Рис. 4.39. Запись электронного каталога (регистрация периодики) на примере АБИС TINLIB

Заглавие книги: 100 ways to be a better manager

Автор: Dennis, K

Сведения об издании: 3rd.

Издательство: Heinemann

Место издания: London

Дата издания: 1989

ISBN : 0234586719

Число стр : 300р.

Цена: 23.00

Вид издания: Handbook

Инвентарный номер: A4

Класс издания: Loan

Шифр хранения: Library

Классификационный индекс: 309.04

Расстановочный шифр: MAN

Предметная рубрика: Management

Ключевое слово, дополн. заглавие: Management

Ключевые слова: 100; Better; Management; Manager; Ways

Рис. 4.40. Запись электронного каталога (регистрация книг)
на примере АБИС TINLIB

для обслуживания определенной группы пользователей, структурного подразделения библиотеки, либо связную совокупность технологических операций.

В табл. 4.9 приведены некоторые наиболее типичные АРМ.

1. *Комплектование и поддержка движения фонда литературы* включает:

- планирование комплектования;
- заказ книг и подписку на периодические издания;
- контроль поступления и оформление рекламаций;
- распределение литературы между отделами или филиальными библиотеками;
- приобретение, передачу, списание документов (составление актов);
- инвентарный и суммарный учет поступлений и списанной литературы;
- статистический анализ и учет результатов и хода комплектования;
- сверку на дубль.

2. *Обработка (регистрация, техническая обработка, каталогизация и индексирование, аналитико-синтетическая обработка) включает в себя:*

- поддержку описаний различных видов и форматов документов (книги, статьи, журналы, диссертации, патенты, каталоги,

Таблица 4.9. Основные АРМ (пользовательские интерфейсы) АБИС

Условное наименование АРМ	Основные функции	Решаемые задачи
«Администратор»	АРМ специалиста, выполняющего системные операции над базами данных в целом	Настройка системы на конкретные требования библиотеки, обслуживание БД – инициализация, актуализация, объединение/разъединение баз данных, архивация/восстановление, ведение списка пользователей и системы паролей, загрузка документов или данных из других каталогов и в другие каталоги или базы данных (иногда оформляются в отдельный АРМ «Экспорт-Импорт»)
«Комплектование», иногда объединяется с «Обработкой»	АРМ библиотечного работника, выполняющего функции по формированию, пополнению и корректировке ЭК	Предварительный заказ литературы, подписка на периодические издания, учет поступления и формирование записей в ЭК (ввод нового документа, корректировка документа и пр.), инвентаризация литературы, формирование и печать отчетов и выходных документов
«Обработка» («Каталогизация»)	- - -	Ввод библиографических описаний в ЭК, поиск и отбор (отметка) документов для обработки, обработка (систематизация, формирование карточек для каталога), присвоение шифров хранения, просмотр и печать документов (выдача библиографических справок, ведение справочников и рубрикаторов)
«Читатель»	АРМ конечного пользователя электронного каталога	Поиск, просмотр и печать найденной информации, формирование заказа на литературу, просмотр и печать результатов поиска
«Книговыдача» («Абонемент, Циркуляция, Кафедра выдачи»)	АРМ библиотечного работника, осуществляющего выдачу и прием литературы	Регистрация читателя, поиски по тематикам и библиографическим описаниям, формирование и печать читательских требований, выдача данных о местонахождении литературы, заказ литературы, очередь невыполненных заказов, прием книг и обработка возврата книг читателем, контроль книговыдач, межбиблиотечный абонемент, статистика книговыдач (справки о задолженностях, выданных книгах и пр.), статистика движения литературы
Специализированные АРМ	«Специализированный отдел» (обычно полный набор функций)	Учет особенностей библиотечных технологий обработки нетипичных материалов – редкие книги, графические материалы, искусство, иностранные языки и пр.

нормативно-технические, неопубликованные, нотно-музыкальные и пр.);

- аннотационную, реферативную или фактографическую обработку документов;
- редактирование ранее составленных или импортированных описаний документов;
- распечатку выходных форм (каталожных карточек, формуляров и пр.).

Программные средства автоматизированных библиотечных систем

В табл. 4.10 приведены основные данные о некоторых наиболее распространенных программных продуктах, используемых в АБИС.

Таблица 4.10. Характеристики некоторых программных продуктов АБИС

Разработчик	Наименование	АРМы, модули, подсистемы	Операционные системы, СП, СУБД	Количество экз.	Базы данных/ массивы
ЮНЕСКО (UNESCO)	ISIS	Уровни доступа	UNIX; Windows; MS DOS	Более 7000	Общая БД
IME (Великобритания)	TriLib	Пользователь; Оператор; Администратор	UNIX; MS DOS	Более 2000	Общая БД
Библиотека по естественным наукам (БЕН РАН, РФ)	SOLAR, DISCAT, СОЧИ	Комплектование; Обслуживание читателей; Каталог; ИРИ	MS DOS; Clipper; Turbo-C;	Более 15	Библиотеки; Читатели; Абоненты
Государственная публичная научно-техническая б-ка (ГПНТБ, РФ)	ИРБИС (ГПНТБ)	Администратор; Каталогизатор; Читатель; Книговыдачи	MS DOS; Windows; ISIS-Pascal ISIS	Более 20	Электронный каталог; Читатели; Заказы
НПО Информсистема (РФ)	МАРК	Администратор; Комплектование; Обработка; Абонемент; Хранение; Поиск;	MS DOS	Более 30	Каталог; Читатели; Фактографические БД
МГУ им. Баумана (РФ)	БКС	Комплектование; Обработка; Заказ; Абонемент; Библиограф	MS DOS; Clipper	Около 10	Издания; Абоненты; Новые поступления; Формуляры; Статистика
ВЦ Минкультуры РФ	АС-Библиотека АС-ЦБС «Библиофил»	Администратор; Комплектование; Обработка; Абонемент; СпецОтдел; Обмен информацией	MS DOS; Rbasic FoxPro Revelation	Более 40	Читатели Подписка МБА
Государственная центральная медицинская научная б-ка (ГЦМНБ, РФ)	«ДИТ-ИБИС»	Модули: Комплектование; Каталогизация; Поиск и заказ; Выдача и прием; Регистрация читателей; Справочная; Лингвистическая; ИРИ, издания; Обмен данными	MS DOS ISIS	Более 12	БД ISIS
МГУ (РФ)	АИБС «Библиотека»	Единый АРМ: ввод; редактирование; поиск; вывод; экспорт/импорт	MS DOS; Clipper	Более 10	Книги; Комплектование; Статья; Периодика; Подписка; Дисциплины; Читатели

Как следует из табл. 4.10, наиболее распространенным программным продуктом для библиотечной автоматизации является ISIS (Integrated Set of the Information Systems), который существует длительное время, распространяется на некоммерческой основе и реализован для разнотипных вычислительных платформ (ЭВМ и операционных систем).

Продукт CDS/ISIS/M разработан в отделе развития и применения программного обеспечения бюро информационных программ и служб ЮНЕСКО специально для библиотек и архивов и является меню-ориентированной информационной системой для хранения и поиска информации в нечисловых БД. Основные его преимущества — возможность обработки неограниченного числа БД и управление полями переменной длины. Кроме того, существует библиотека функций, расширяющих систему программирования Паскаль (ISIS — Паскаль), позволяющая из программ на ЯП Паскаль осуществлять вызовы ядра ISIS, что дает возможность пользователям самостоятельно развивать систему.

Фирма IME Systems Inc. (США), ранее известная как Information Management and Engineering, была организована в 1984 г. с целью разработки и продажи программных продуктов автоматизации библиотек. Основной продукт фирмы — система Tinlib. Это модульная интегральная библиотечная система для выполнения процессов каталогизации, контроля по автору, интерактивного доступа к каталогу, контроля за обращением и периодических изданий, МБА. Дополнительные модули системы обеспечивают регистрацию импорта — экспорта, генерацию отчетов, обмен данными.

Программный продукт работает с рефератами, аннотациями и другими текстовыми блоками в библиографическом описании, используется в более чем 2 тыс. организациях почти 40 стран мира. В основном это специальные библиотеки (фирменные, правительственные, юридические, медицинские) и информационные центры, 20 % составляют вузовские библиотеки.

Функции и структура АИБС TinLib

Система Tinlib выбрана здесь для более подробного рассмотрения как в связи с широкой распространенностью, реализацией на различных платформах (см. табл. 4.10), так и в связи с развитостью функций и типичной структурой: система поддерживает три АРМА (функциональных набора) — рис. 4.41:

- оператор (сотрудник, ответственный за рабочие функции);



Рис. 4.41. Структура интерфейсов Tinlib

- читатель (пользователь, проводящий свободный поиск в каталогах);
- администратор (лицо, ответственное за поддержание БД).

В системе реализована 3- (4-) уровневая система навигации — Основное меню — Меню — Список записей — Содержание записи (рис. 4.42).

Основное меню системы есть форма, появляющаяся после запуска соответствующего АРМа, которая содержит перечень основных библиотечных процессов, выполняемых данным АРМ, причем каждый процесс можно вызвать для работы, указав его курсором и нажав <Enter> (например, рис. 4.43 — основное меню АРМ читателя).

Меню (подменю) появляется на экране, представляя более детальный перечень операций (например, рис. 4.44 — варианты поиска по ключевым словам).

1. ОСНОВНОЕ МЕНЮ (рубрика)

1.1. Меню (рубрика)

1.1.1. Подменю (рубрика)

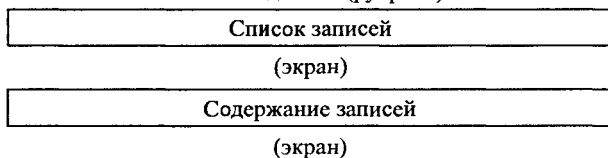


Рис. 4.42. Схема уровней интерфейса пользователя TinLib

Подсказка: Основное меню – Перемещайте стрелку вверх и вниз – нажмите <Enter>

ВЫБЕРИТЕ тип ПОИСКА, используя клавиши "стрелка вверх" или "стрелка вниз", нажмите <ENTER>.

==>

АВТОРЫ
ЗАГЛАВИЯ
ПРЕДМЕТНЫЕ РУБРИКИ
ПОИСК ПО КЛЮЧЕВЫМ СЛОВАМ
КОМБИНИРОВАННЫЙ ПОИСК

Заставка <F1>					Помощь <F9>
------------------	--	--	--	--	----------------

Рис. 4.43. Основное меню АРМ читателя (конечного пользователя)

Подсказка: Операции – Перемещайте стрелку вверх и вниз – нажмите <Enter>

ПОИСК ПО КЛЮЧЕВЫМ СЛОВАМ

==>

** : Просмотр по ключевым словам всех заглавий
** : Поиск по ключевым словам книг/ЖУ/обычных носителей
** : Поиск по ключевым словам статей
** : Поиск по ключевым словам Период./Серияльн. изданий

Осн. Меню <F1>					Помощь <F9>
-------------------	--	--	--	--	----------------

Рис. 4.44. Меню выбора режима поиска по ключевым словам

Список записей есть перечень данных, которые могут быть запрошены и найдены (например, заглавия книг, изданий, имена читателей, рубрики классификатора, термины тезауруса и пр.). На рис. 4.45 и 4.46 в качестве примеров приводятся соответственно списки ключевых слов и заглавий книг.

Подсказка:Список-Перемещ. стрелку и нажмите <Enter> или <F10> для поиска скач

Area	*Кн :1	*Ст :0	*Сер :0
Art	*Кн :5	*Ст :1	*Сер :0
Artists	*Кн :1	*Ст :0	*Сер :0
==> Arts	*Кн :5	*Ст :1	*Сер :0
ASIS	*Кн :0	*Ст :0	*Сер :1
ASLIB	*Кн :0	*Ст :0	*Сер :2
Assesment	*Кн :0	*Ст :1	*Сер :0
Athena	*Кн :0	*Ст :1	*Сер :0
Athenaeum	*Кн :1	*Ст :0	*Сер :0
Automate	*Кн :1	*Ст :0	*Сер :0

Операции <F1>	Осн. Меню <F2>			Помощь <F9>	<F10>
------------------	-------------------	--	--	----------------	-------

Рис. 4.45. Просмотр частотно-видового словаря по ключевым словам

Подсказка: Запись—Укажите стрелкой на элемент — нажмите <ENTER> для навигации

Ключевое слово :Arts	Заглавие книги :Arts festivals in the UK	Заглавие книги :Battersea Arts Centre : review and analysis of marketing and publicity	⇒ Заглавие книги :Beyond mass media : channel for the arts / challenge for the arts	Заглавие книги :The modern arts : an outline for the librarian educator, and general reader	Заглавие статьи :The Impact of Thatcherism on the Arts Council
Список <F1>	Обратно <Alt/F1>/<F1>	Осн. Меню <F2>			Помощь <F9>

Рис. 4.46. Экран просмотра списка заглавий для выбранного ключевого слова ARTS

Выбор одной из записей для более подробного просмотра осуществляется путем подвода к ней курсора, что приводит к появлению *Содержания записи* (библиографическое описание, данные о читателе и пр.), которое может занимать несколько последовательных экранов (рис. 4.47—4.49).

Рис. 4.47—4.49 иллюстрируют также возможности Tinlib по ассоциативному доступу к смежным записям: переход от записи (рис. 4.47) по полю *рубрика* (*BROADCASTING*) через экран для рубрики (рис. 4.48) к одному из документов той же рубрики (рис. 4.49).

Заглавие :Beyond mass media : channel for the arts / challenge for the arts	Статус выдачи	Выдано. Ожидается возврат :30/10/96			
Номер издания :A29	Расположено :MEI 120.1	*****			
Редактор :Meijers, Christel	Конференция:Channels for the Arts/Challenge for the Arts, Amsterdam, 2nd-4th May 1991				
Издательство :Felix Meritis Foundation; Amsterdam; 1993	ISBN : 9074706010				
Классификационный индекс :120.1	⇒ Предметная рубрика :BROADCASTING				
	Предметная рубрика :CULTURAL POLICY				
	Предметная рубрика :ECONOMICS OF THE ARTS				
Список <F1>	Обратно <Alt/F1>/<F1>	Осн. Меню <F2>			Помощь <F9>

Рис. 4.47. Экран просмотра регистрационной карточки выбранного издания

В базе данных Tinlib определены следующие типы полей:

- поля данных (содержат произвольную информацию, не проверяемую системой, поэтому наличие или отсутствие поля, а также формат данных определяются пользователем);

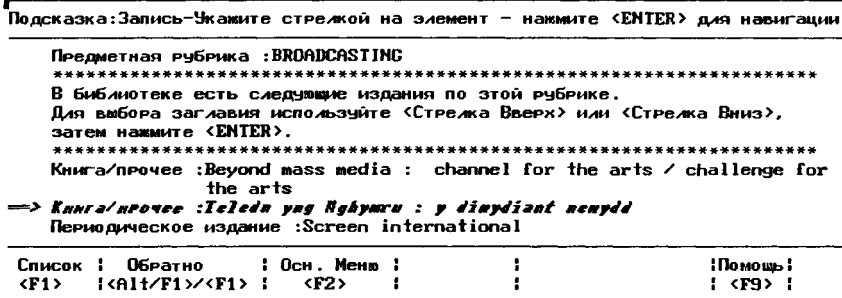


Рис. 4.48. Экран просмотра заглавий изданий, соответствующих выбранной выше предметной рубрике BROADCASTING

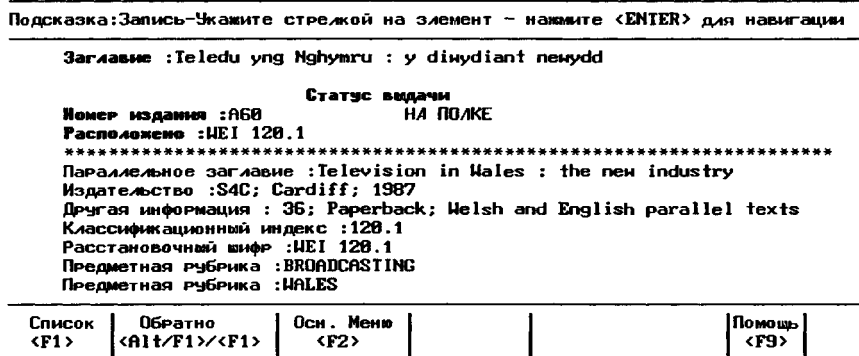


Рис. 4.49. Карта издания, ассоциативно связанного через предметную рубрику с изданием (рис. 4.47)

- обязательные поля (должны содержать данные, в противном случае запись не будет сохранена) используются в системе для установления связей между записями;
- проверяемые поля (содержимое проверяется при вводе по словарям или логическим выражениям — НОМЕР ЧИТАТЕЛЯ, НОМЕР ИЗДАНИЯ, КЛАСС ЧИТАТЕЛЯ);
- поисковые (индексируемые) поля содержат данные, используемые для прямого доступа (АВТОР, ISSN и пр.);
- подполя (подчиняются какому-либо полю — НОМЕР В СЕРИИ и пр.).

Важным типом обязательных полей являются классы:

- *класс читателей* — определяет максимальное число выдач, разрешенных читателю, ограничения на период выдачи, код

штрафа, дату окончания действия читательского билета. Классы читателей разграничивают группы пользователей (например, дети могут иметь меньшие штрафы по сравнению со взрослыми);

- *классы изданий* — используются для того, чтобы определить периоды выдачи книг и статей, предназначенных для выдачи. Возможные классы — *выдаваемые, выдаваемые на короткий период и справочные издания*.

Ограничения, установленные в классе издания, имеют приоритет по отношению к ограничениям, связанным с классом читателя.

Основные меню и подменю АРМ TinLib

Ниже приводятся распечатки основных меню, некоторых меню, и подменю при работе с АРМ Tinlib.

I. Интерфейс конечного пользователя (читателя):

1. АВТОРЫ
2. ЗАГЛАВИЯ
 - 2.1. Поиск по Всем Заглавиям
 - 2.2. Поиск по Книгам/AV/Обычным носителям
 - 2.3. Поиск по Статьям
 - 2.4. Поиск по Периодическим/Сериальным изданиям
3. ПРЕДМЕТНЫЕ РУБРИКИ
 - 3.1. Поиск по Предметным рубрикам
 - 3.2. Поиск по Терминам тезауруса
4. ПОИСК ПО КЛЮЧЕВЫМ СЛОВАМ
 - 4.1. Просмотр по ключевым словам всех заглавий
 - 4.2. Поиск по ключевым словам книг/AV/обычных носителей
 - 4.3. Поиск по ключевым словам статей
 - 4.3. Поиск по ключевым словам Период./Сериальн. изданий
5. КОМБИНИРОВАННЫЙ ПОИСК
 - 5.1. Просмотр по ключевым словам всех заглавий
 - 5.2. Поиск по ключевым словам книг/AV/обычных носителей
 - 5.3. Поиск по ключевым словам статей
 - 5.4. Поиск по ключевым словам Период./Сериальн. изданий

II. Интерфейс оператора:

1. КАТАЛОГИЗАЦИЯ
 - 1.1. Заглавия документов
 - 1.1.1. Поиск/редактирование книг
 - 1.1.2. Поиск/редактирование листового материала/статей
 - 1.1.3. Поиск/редактирование периодич./сериальных изданий
 - 1.1.4. Редактирование серийных изданий
 - 1.1.5. Редактирование подсерий издания
 - 1.1.6. Поиск по всем видам документов

- 1.1.7. Поиск книг
- 1.1.8. Поиск статей
- 1.1.9. Поиск периодических/серийных изданий
- 1.1.10. Поиск серий (книги)
- 1.1.11. Поиск серий (периодические издания)
- 1.1.12. Поиск по терминам из заглавия

1.2. Авторы/Редакторы и др.

- 1.2.1. Поиск/редактирование любого автора/редактора и др.
- 1.2.2. Поиск по авторам/редакторам
- 1.2.3. Поиск книг по автору
- 1.2.4. Поиск статей по автору
- 1.2.5. Поиск книг по коллективному автору
- 1.2.6. Поиск статей по коллективному автору
- 1.2.7. Поиск книг по конференции/встрече
- 1.2.8. Поиск статей по конференции/встрече
- 1.2.9. Поиск периодических изданий по конференции/встрече

1.3. Предметные рубрики/Тезаурус

- 1.3.1. Термины тезауруса — построение/редактирование
- 1.3.2. Предметные рубрики — редактирование
- 1.3.3. Поиск по всем терминам тезауруса
- 1.3.4. Поиск по всем предметным рубрикам
- 1.3.5. Поиск по всем терминам заглавия

1.4. Издательство

- 1.4.1. Поиск/редактирование информации об издательствах
- 1.4.2. Поиск по всем издательствам

1.5. Классификация

- 5.2. Поиск по расстановочному шифру

1.6. ISBN и др.

- 1.6.1. Поиск по ISBN/ISSN
- 1.6.2. Поиск по ISBN
- 1.6.3. Поиск по ISSN
- 1.6.4. Поиск по справочному номеру издателя

1.7. Обмен данными TINLIB

- 1.7.1. Выходной формат для книг
- 1.7.2. Выходной формат для статей
- 1.7.3. Выходной формат для периодических изданий

2. ОБСЛУЖИВАНИЕ ЧИТАТЕЛЕЙ

2.1. Кафедра выдачи

- 2.1.1. Выдача/редактирование по номеру читателя
- 2.1.2. — по имени читателя
- 2.1.3. — по номеру издания
- 2.1.4. — по заглавию — книги
- 2.1.5. — по заглавию — статьи
- 2.1.6. Возврат по имени читателя
- 2.1.7. — по заглавию — книги
- 2.1.8. — по заглавию — статьи
- 2.1.9. Продление по имени читателя
- 2.1.10. — по заглавию — книги
- 2.1.11. — по заглавию — статьи

- 2.1.12. Регистрация изданий — книг по заглавию
- 2.1.13. Регистрация изданий — статей по заглавию
- 2.1.14. Процедура : Возврат всех изданий, выданных читателю
- 2.1.15. Процедура : Продление всех выданных изданий

2.2. Уведомления

- 2.2.1. Издания, на которые отослано более трех уведомлений
- 2.2.2. Утерянные издания, требующие возмещения — по читателю
- 2.2.3. Статус — все ПЕРВЫЕ уведомления о просроченных изданиях
- 2.2.4. Статус — все ВТОРЫЕ уведомления о просроченных изданиях
- 2.2.5. Статус — все ТРЕТЬИ уведомления о просроченных изданиях
- 2.2.6. Создание и печать полугодовых напоминаний
- 2.2.7. Создание и печать уведомлений о просроченных изданиях
- 2.2.8. Печать напоминаний о заказанных изданиях
- 2.2.9. Печать уведомлений о продлении

2.3. Заказы

- 2.3.1. Произвести заказ по номеру читателя
- 2.3.2. — по имени читателя
- 2.3.3. — по номеру издания
- 2.3.4. — по заглавию — книги
- 2.3.5. — по заглавию — статьи
- 2.3.6. Освобождение заказов по заглавию — книги
- 2.3.7. — по заглавию — статьи
- 2.3.8. Отмена заказов по имени читателя
- 2.3.9. Отмена всех просроченных заказов
- 2.3.10. Печать всех уведомлений о доступности издания
- 2.3.11. Печать всех уведомлений об отмененных заказах
- 2.3.12. Печать всех уведомлений о сделанных заказах

2.4. Отчеты

- 2.4.1. Список читателей по номерам
- 2.4.2. Список читателей по именам
- 2.4.3. Статус читателей по именам
- 2.4.4. Статус изданий по номерам
- 2.4.5. Список всех выданных изданий (по номерам)
- 2.4.6. Список всех заказанных изданий (по номерам)
- 2.4.7. Все невыполненные заказы (по дате)
- 2.4.8. Отправленные напоминания
- 2.4.9. Издания по классам изданий
- 2.4.10. Читатели по классам читателей
- 2.4.11. Издания со статусом «возвращено по требованию»
- 2.4.12. Издания со статусом «утеряно»
- 2.4.13. Архивный файл — выдачи
- 2.4.14. Архивный файл — заказы

2.5. Поддержка

- 2.5.1. Регистрация читателя (по имени)
- 2.5.2. Регистрация книг (по заглавию)
- 2.5.3. Регистрация статей (по заглавию)
- 2.5.4. Установка классов читателей
- 2.5.5. Установка классов изданий
- 2.5.6. Установка текста и сообщений для уведомлений

2.5.7. Установка Таблицы Управления Обслуживанием

2.5.8. Список читателей (по номерам)

2.5.9. Список изданий (по номерам)

3. КОМПЛЕКТОВАНИЕ КНИГ

3.1. Книги — Заказ

3.1.1. Запрос на издание

3.1.2. Подтверждение запросов

3.1.3. Создание единицы заказа

3.1.4. Создание единичного или множественного заказа на покупку

3.1.5. Исправление и печать заказов на покупку

3.1.6. Исправление и печать заказов на покупку по поставщику

3.2. Книги — Получение/Счета/Задолженности

3.2.1. Получение изданий — по заглавию

3.2.2. — по номеру заказа на покупку

3.2.3. — по поставщику

3.2.4. — по запросившему

3.2.5. Ведение счетов

3.2.6. Задолженности

3.3. Книги — Рекламации/Отмены заказов/Возвраты

3.3.1. Рекламации или Отмены — по заглавию

3.3.2. — по номеру единицы заказа

3.3.3. — по номеру заказа на покупку

3.3.4. Изменение и печать писем на рекламацию

3.3.5. Изменение и печать писем на отмену

3.3.6. Авторизация возвратов

3.3.7. Печать набора писем на рекламацию

3.3.8. Процедура : Печать набора писем на отмену

3.4. Книги — Поддержка

3.4.1. Коды поставщиков

3.4.2. Коды доставки

3.4.3. Коды счетов

3.4.4. Источники финансирования

3.4.5. Сведения о персонале

3.4.6. Сообщения

3.5. Книги — Статус

3.5.1. Запросы/Заказы по заглавию

3.5.2. Заказанные издания по заглавию

3.5.3. Заказы на покупку

3.5.4. Рекламации/Отмены

3.5.5. Источники финансирования

3.5.6. Счета

3.5.7. Поставщики

3.5.8. Коды адресов доставки

3.5.9. Книги по источникам финансирования

4. КОМПЛЕКТОВАНИЕ ПЕРИОДИЧЕСКИХ ИЗДАНИЙ

4.1. Заказ

4.1.1. Запрос на название периодического издания

4.1.2. Подтверждение и создание заказа на покупку периодического издания ИЛИ Подтверждение запросов для их обработки

- 4.1.3. Подготовка заказа на покупку одной или нескольких единиц заказа
- 4.1.4. Изменение и печать заказов на покупку
- 4.1.5. Изменение и печать заказов на покупку по поставщикам
- 4.1.6. Модификация записей выпусков
- 4.1.7. Создание записей *Дополнительных выпусков* или удаление записей

выпусков

- 4.1.8. Обновление заказов на периодические издания
- 4.1.9. Генерация записей выпусков изданий
- 4.1.10. Печать набора заказов на покупку

4.2. Регистрация получения

- 4.2.1. Регистрация — быстрый просмотр
- 4.2.2. Регистрация — краткий список
- 4.2.3. Регистрация — краткий список по названиям
- 4.2.4. Регистрация — краткий список по ISSN
- 4.2.5. Регистрация — полный список
- 4.2.6. Регистрация — полный список по названиям
- 4.2.7. Регистрация — полный список для ежедневных выпусков
- 4.2.8. Изменение записей выпусков
- 4.2.9. Создание записей *Дополнительных выпусков* или удаление записей

выпусков

4.3. Маршрутизация

- 4.3.1. Выбор и печать листов маршрутизации
- 4.3.2. Выбор и печать листов маршрутизации — краткий формат
- 4.3.3. Печать набора листов маршрутизации — краткий формат
- 4.3.4. Печать набора листов маршрутизации

4.4. Счета/Рекламации/Отмены заказов/Возвраты

- 4.4.1. Отмена заказов на периодику (кроме ежедневных)
- 4.4.2. Отмена заказов на периодику (Прямая доставка)
- 4.4.3. Отмена заказов на периодику — для ежедневных
- 4.4.4. Ведение счетов
- 4.4.5. Уведомления о задолженности
- 4.4.6. Визы на возврат
- 4.4.7. Рекламация выпусков прямой доставки
- 4.4.8. Рекламация для ежедневных выпусков по названию
- 4.4.9. Рекламация для ежедневных выпусков
- 4.4.10. Изменение и печать писем о рекламации
- 4.4.11. Изменение распечатанных рекламаций
- 4.4.12. Изменение и печать писем об отмене
- 4.4.13. Печать набора писем о рекламации
- 4.4.14. Печать набора писем об отмене

4.5. Поддержка

- 4.5.1. Источники финансирования
- 4.5.2. Коды поставщиков
- 4.5.3. Коды адресов доставки
- 4.5.4. Коды адресов счетов
- 4.5.5. Сведения о персонале
- 4.5.6. Листы маршрутизации
- 4.5.7. Сообщения
- 4.5.8. Информация о наличии единиц изданий

4.6. Статус

- 4.6.1. Запросы/заказы на периодические издания
- 4.6.2. Источники финансирования
- 4.6.3. Заказы по поставщикам
- 4.6.4. Адреса доставки (заказы)
- 4.6.5. Счета
- 4.6.6. Периодические издания в листах маршрутизации
- 4.6.7. Наличие периодических изданий (кроме ежедневных)
- 4.6.8. Наличие периодических изданий (ежедневные)
- 4.6.9. Периодические издания по источникам финансирования

III. Интерфейс администратора системы:

1. АДМИНИСТРИРОВАНИЕ

1.2. Административные операции

- 1.2.1. Поддержка многопользовательского файла 010
- 1.2.2. Расширение (grow) всех файлов базы данных
- 1.2.3. Реиндексация (reindex) всех файлов базы данных
- 1.2.4. Изменение размера (resize) всех файлов базы данных
- 1.2.5. Реорганизация (Reorg) всех файлов базы данных
- 1.2.6. Модификация (modify) всех файлов базы данных
- 1.2.7. Расширение (grow) всех файлов словаря базы данных
- 1.2.8. Реиндексация (raindex) всех файлов словаря базы данных
- 1.2.9. Изменение размера (resize) всех файлов словаря базы данных
- 1.2.10. Реорганизация (reorg) всех файлов словаря базы данных
- 1.2.11. Модификация (Modify) всех файлов словаря базы данных

1.3. Поддержка — Частота поступлений/Список стоп—слов

1.4. Поддержка — Местонахождение и т. д.

1.5. Поддержка — Проверка импорта/конвертирования

1.6. Процедуры изменения/обновления

1.7. Процедуры изменения/обновления — Периодика (Удаление временных отметок)

1.8. Операции изменения

1.9. Поддержка — Изменение комплектования книг

1.10. Поддержка — Изменение комплектования периодики

2. ОТЧЕТЫ

2.1. Каталогизация

- 2.1.1. Последние поступления — монографии
- 2.1.2. Последние поступления — статьи
- 2.1.3. Последние поступления — периодические издания
- 2.1.4. Список содержимого полок — Монографии по расстановочному

шифру

- 2.1.5. Список содержимого полок — Периодические издания по расстановочному шифру

- 2.1.6. Тезаурус — все термины

- 2.1.7. Метки корешка переплета — монографии

- 2.1.8. Метки корешка переплета — статьи

- 2.1.9. Метки корешка переплета — периодические издания (TINGEN)

2.2. Комплектование

2.3. Периодические издания

2.4. ИРИ

2.5. Обслуживание читателей

2.6. Источники финансирования

3. АРХИВИРОВАНИЕ

3.1. Архивирование при Комплектовании книг

3.1.1. Архивирование уведомлений о задолженности

3.1.2. Архивирование рекламаций/отмен заказов

3.1.3. Архивирование выборочных счетов

3.1.4. Архивирование выборочных единиц заказа (с заказами на покупку — если таковые имеются)

3.1.5. Архивирование всех счетов, прошедших на оплату

3.1.6. Архивирование всех полностью полученных единиц заказа и содвержания заказа на покупку

3.2. Архивирование при Комплектовании Периодических изданий

3.3. Архивирование при Выдаче

3.3.1. Архивирование сведений о выдачах

3.3.2. Очистка архива и статистики

3.3.3. Обнуление выдач читателя

3.3.4. Обнуление выдач издания

4.5. Географические информационные системы (ГИС)

Важным специфическим классом АИС, специализирующихся на определенном виде информации (в основном — графической), являются географические системы (ГИС) [14].

Согласно некоторым оценкам, 80 % всех существующих данных являются географическими (пространственными) — достаточно упомянуть терабайты информации, получаемой со спутников при фотографировании земной поверхности. Эти данные широко используются многочисленными группами пользователей для самых разнообразных целей — в исследованиях, принятии решений, в быту; их изучением занимаются ученые из разных областей — биологи используют их при описании распространения растений, метеорологи — при изучении закономерностей выпадения осадков, экологи — для уточнения зон риска, где есть угроза исчезновения каких-либо видов животных, и т. п.

В число организаций, работающих с привязанными к месту данными, входят промышленные предприятия, использующие природные ресурсы, предприятия коммунальной сферы, органы местного управления, занимающиеся учетом землепользования и сбором на-

логов, органы образования, пожарной охраны и охраны общественного порядка и т. д. Ранее основным видом информационных продуктов с использованием географических данных были карты. Сейчас автоматизированные ГИС способны обеспечить глобальный доступ к накопленным данным по информационным сетям.

По определению ГИС — это система, обеспечивающая сбор, хранение, обработку, доступ, отображение и распространение пространственно-координированных данных. По территориальному охвату различают глобальные, региональные, локальные или местные ГИС. Проблемная ориентация ГИС определяется решаемыми в ней задачами (научными и прикладными), среди них — инвентаризация ресурсов (в том числе кадастр), анализ, оценка, мониторинг, управление и планирование, поддержка принятия решений.

Могут быть выдвинуты следующие принципы функционирования ГИС:

- географическая информация должна собираться и обрабатываться признанными публичными агентствами, имеющими опыт и длительный стаж работы, а не специально создаваемыми группами или организациями;
- ГИС должна удовлетворять децентрализованной структуре, отвечающей независимым требованиям, которые выдвигаются всеми агентствами-участниками, причем наблюдение за деятельностью агентств должно осуществляться в рамках действующих законодательных установок и обеспечивать стандарты качества;
- ГИС должна базироваться на равноправном и справедливом участии всех заинтересованных в использовании информации, а также тех, интересы которых оно затрагивает.

ГИС можно рассматривать как модель изучаемого объекта и промежуточное звено между объектом и исследователем. Базовым компонентом любой ГИС являются географическая или пространственная информация, представленная в виде цифровых данных о пространственных объектах и включающая сведения об их местоположении и свойствах, пространственных и непространственных атрибутах. Полное описание пространственных данных складывается из взаимосвязанных описаний топологии, геометрии и атрибутики объектов.

Важной компонентой ГИС являются цифровые модели: ЦММ — цифровая модель местности; ЦМР — цифровая модель рельефа; ЦМКМ — цифровая модель карты местности; ЦМКД — цифровая модель кадастрового деления и др.

Характеристики цифровых моделей местности (ЦММ)

Цифровые модели оперируют с различными типами информации.

Метрическая информация передает измерительные характеристики объекта, т. е. координаты, размеры. Она относительно проста и однородна по структуре, в силу чего является сильно типизированной. Метрическая информация в ГИС содержит координатные и некоторые (числовые) атрибутивные данные.

Качественным отличием цифровых моделей, полученных по реальным измерениям, является точностная характеристика модели. Она обусловлена ошибками измерений и последующих вычислений при геометрическом моделировании. Этот параметр определяет применимость цифровой модели, в частности, при получении графических реализаций в разных масштабах.

Атрибутивная информация — информация о свойствах и связях объектов. В ГИС она включает атрибутивные данные и метаданные и может подразделяться на семантическую, технологическую и другие виды. Все эти виды информации являются семантическими, тем не менее более правильно называть ее атрибутивной.

Этот тип информации определяет принадлежность точек или объектов к определенному классу или объекту (сложный или простой объект), описывает свойства объектов и их частей, задает взаимосвязи и условия обработки, условия воспроизведения и т. п.

В исходном виде эта информация неоднородна, сложна по структуре и является слабо типизированной, поэтому для создания ЦММ требуются анализ, классификация и типизация.

Для хранения в базах данных эта информация должна быть типизирована, т. е. разбита на более мелкие группы, имеющие достаточное число сходных признаков.

Синтаксическая информация — определяет последовательность работы при корректировке или обновлении ЦММ, правила построения и представления ЦММ. Эта информация типизирована.

Первые два типа информации определяются логической структурой ЦММ и не зависят от выбора СУБД. В силу этого их можно назвать внутренними по отношению к ЦММ.

Синтаксическая информация зависит от технологии использования ЦММ с учетом конкретных технических средств и их возможностей. Например, она определяется разрешающей способностью монитора и его цветовой палитрой, разрешением принтера или плоттера, погрешностью оцифровывания или фотограмметрической обработки и т. д. Синтаксическая информация является внешней по

отношению к ЦММ, зависит от выбора технических и программных средств.

Логическая и физическая структура ЦММ

Понятия логической и физической структуры ЦММ [14] являются развитием и расширением соответствующих моделей данных.

Логическая структура ЦММ определяется как совокупность схем и логических записей, описывающих данную ЦММ. Такая характеристика относится к описательным. Схемы, составляющие логическую структуру ЦММ, могут быть различными в зависимости от назначения и принципов построения.

Логическая структура обуславливается концепцией и методологией моделирования. Она может включать схемы взаимосвязи частей ЦММ, соответствующих реальным связям объекта на местности, в базе данных, схемы взаимосвязи свойств ЦММ и построения ЦММ. Она содержит логические записи, составляющие информационную основу. Элементом логической структуры ЦММ является логическая запись.

Физическая структура ЦММ определяется способом реализации логической ЦММ на конкретной технической основе. В частности, она задается форматом записи данных, хранимых на носителях информации. Элементом физической структуры ЦММ является физическая запись.

Целостность ЦММ. Недостаточно того, чтобы ЦММ просто содержала набор данных. Необходимо, чтобы она отражала объекты реального мира. Важно, чтобы такое отражение было однозначным и непротиворечивым. В этом случае говорят, что ЦММ удовлетворяет условию целостности. Целостность ЦММ имеет два аспекта: как объекта БД и как модели реального объекта.

Целостность ЦММ как объекта базы данных определяется требованиями СУБД, соответствует понятию целостности информации в БД, и достигается путем типизации и структурирования исходной информации. Такая целостность позволяет осуществлять работу с ЦММ как с БД, направлять к ней запросы, проводить фильтрацию, получать справки или отчеты.

Целостность ЦММ как модели реального объекта определяется требованием получения проекта карты или картографической композиции средствами ГИС. Для достижения такой целостности информация должна быть полной, актуальной и отвечать требованиям точности при получении данного проекта карты. Информация, таким образом, должна включать не только собранные на местности

данные, но и библиотеки условных знаков, которые хранятся в БД независимо от ЦММ. В данном случае целостность ЦММ как модели объекта обуславливается полнотой информации БД.

Дискретность. ЦММ относится к классу дискретных моделей. Отсюда следует, что ЦММ необходимо хранить как объект дискретной базы данных. Геометрическая часть ЦММ содержит отдельные точки поверхности объектов. Несмотря на это ЦММ позволяет строить непрерывные линии и поверхности, т. е. получать аналоговые модели (аналоговые карты) за счет совместного использования метрической и семантической информации.

Многофункциональность. ЦММ создают с целью адаптации при решении различных задач. Графическое отображение ЦММ не должно зависеть от средств воспроизведения графической информации. Например, одна и та же ЦММ может использоваться для получения карт масштабного ряда.

Для многократного использования ЦММ нужны дополнительные данные: описатели, классификаторы, нормативные данные, правила применения и т. д. Обычно их называют метаданными и они составляют содержание словаря данных.

Виды моделирования

Рассмотрим различные виды моделирования при сборе и первичной обработке информации, хранении и обновлении, представлении (отображении).

При сборе информации для построения цифровых моделей используются автоматизированные средства регистрации. Источниками информации служат карты, таблицы, спецификации, геодезические координаты точек и объектов местности, координаты точек на аэрокосмических и наземных фотоснимках, данные, получаемые по телевизионным и/или радиолокационным снимкам, телеметрические данные, информация, считываемая с планов и карт, данные о допусках и погрешностях, дополнительная информация текстового характера.

После сбора первичных данных на уровне хранения и обновления информации осуществляются унификация, коррекция информации, содержащей ошибки и дополнения к ней. Таким образом, формируется унифицированная совокупность данных, одинаковая для различных средств и технологий сбора, позволяющая в дальнейшем применять ее для получения чертежей и планов не одного, а нескольких смежных масштабов.

На уровне представления ЦММ цифровая информация отображается в удобном для пользования виде. ЦММ может генерироваться из разных моделей. Визуальное представление реализуется на современных устройствах вывода информации.

Технологически можно выделить следующие виды моделирования: семантическое, инвариантное, геометрическое, эвристическое, информационное. Они проявляются в разной степени на разных системных уровнях обработки информации.

Семантическое моделирование связано с задачами кодирования и лингвистического обеспечения, поэтому оно используется в основном на уровне сбора первичных данных. Это обусловлено также большим объемом и разнообразием входной информации, сложностью ее структуры, возможным наличием ошибок.

Чем более разнородна входная информация по структуре и содержанию, чем менее она унифицирована, тем больший объем семантического моделирования применяется в подсистеме сбора.

Инвариантное моделирование основано на работе с полностью или частично унифицированными информационными элементами или структурами. Его эффективность доказана опытом применения прежде всего САПР и других АС (в таких системах эти элементы известны как графические примитивы). Этот вид моделирования предполагает использование групповых операций, что обеспечивает более высокую производительность труда по сравнению с индивидуальным моделированием. Кроме того, оно способствует повышению производительности обработки информации, особенно при моделировании (обработке) графических объектов.

Примером инвариантного моделирования является работа со слоями в ГИС. Изменение характеристик слоев влечет за собой изменение всех объектов данного слоя. Одна операция со слоем эквивалентна множеству операций с каждым его объектом.

Геометрическое моделирование заключается в изменении геометрических характеристик объектов или в построении последних. Оно применяется там, где требуется обработка метрических данных.

Эвристическое моделирование заключается в интерактивной обработке. Оно основано на экспертном оценивании результатов каждого шага обработки и последующем принятии решения о выборе очередного метода и этапа обработки; позволяет учитывать индивидуальные свойства нетиповых объектов и при решении специальных нетиповых задач.

Эвристическое моделирование базируется на реализации общения пользователя с ЭВМ по сценарию, учитывающему, с одной стороны, технологические особенности программного обеспечения,

с другой — особенности и опыт обработки данной категории объектов.

Информационное моделирование связано с созданием и построением информационных моделей. Оно включает преобразование различных форм информации, например графической или текстовой, в вид, задаваемый пользователем. Это моделирование эффективно только при предварительной разработке интегрированной информационной основы и применении баз данных.

Следует отметить, что степень использования перечисленных видов моделирования в геоинформационных технологиях различна и зависит от этапа обработки информации. Перечисленные виды моделирования являются типичными и могут использоваться в любых информационных системах. Это обусловлено интеграцией в геоинформационные технологии технологий других автоматизированных систем.

Однако в геоинформатике есть специфическое моделирование, которое применяется только в геоинформационных системах. Это геоинформационное моделирование.

Геоинформационное моделирование — это преобразование моделей пространственных объектов, при котором каждый графический объект связывается с одной или несколькими таблицами БД, а изменение графического объекта вызывает корректировку табличных данных и наоборот.

При геоинформационном моделировании реальное пространственное явление или объект описываются с помощью специального аппарата. Такое описание называют формализованным. Оно служит для представления исследуемых элементов явлений и их взаимосвязей.

Геоинформационное моделирование определяют также как класс моделирования графических объектов, взаимосвязанных с базами данных и включающими пять основных типов моделирования:

- преобразование графической информации, которое приводит к изменению графических и табличных данных;
- преобразование табличных данных, что приводит к изменению графических и табличных данных;
- преобразование графических объектов из одного типа в другой;
- построение цифровых моделей явлений;
- построение, редактирование или модификация графических объектов на основе отношений между пространственными объектами (без использования графических редакторов).

Основу геоинформационного моделирования как специализированной технологии составляют преобразования, основанные на теоретико-множественных отношениях, законах формальной логики, алгоритмах обработки изображений, технологиях работы с компьютерной графикой, технологиях СУБД и многом другом.

Объектами геоинформационного моделирования являются пространственные графические объекты и объекты базы данных ГИС.

Геоинформационное моделирование включает следующие специальные технологии:

- геогруппировку — построение временной динамической модели путем объединения графических объектов в более крупные;
- буферизацию — процедуру построения полигональных объектов по заданным линейным (и точечным) объектам и параметрам буферизации;
- генерализацию — процедуру обобщения графических объектов и изменения их видимости при изменении масштаба;
- комбинирование — процедуру композиции или декомпозиции графических объектов на основе отношений между ними;
- геокодирование — процедуру позиционирования (координатной привязки) данных одной к данным другой позиционно определенной таблицы;
- обобщение данных — процедуру создания атрибутов новых объектов на основе отношений атрибутов исходных;
- построение тематических карт на основе анализа и обработки атрибутивных данных;
- проведение автоматической классификации признаков графических объектов (включая растровые) по заданным критериям.

Процесс геоинформационного моделирования оператор начинается с того, что выделяет объект или объекты моделирования среди множества других, не участвующих в процессе моделирования.

Данная процедура называется активизацией объекта (объектов) и осуществляется по аналогии с процедурой активизации объектов во всех существующих графических редакторах.

Цифровые модели рельефа

Цифровая модель рельефа — это дискретная модель рельефа местности (поверхности), предназначенная для компьютерного моделирования и отображения средствами компьютерной графики.

Каждую точку поверхности Земли невозможно передать в модель. Поэтому используют цифровые модели рельефа, которые пе-

редают отдельные точки. В тоже время наборы этих точек позволяют восстанавливать всю поверхность при использовании методов компьютерного моделирования.

Главная цель ЦМР — при минимальном числе точек модели обеспечить максимально возможную точность построения рельефа.

В простейшем случае ЦМР — это набор трехмерных координат точек рельефа, а также информация о связях между ними и способах восстановления поверхности по данным точкам.

Точки для ЦМР могут собираться по горизонталям, по профилям и по заданной регулярной или нерегулярной сетке.

Сетку задают плановые координаты. Регулярная сетка предпочтительнее нерегулярной тем, что она существенно (на порядки) снижает размерность матриц, необходимых при аналитическом описании и восстановлении любой точки рельефа.

Основная проблема при построении ЦМР заключается в том, что реальная поверхность является нерегулярной, т. е. имеет разрывы. Математические модели, как правило, описывают гладкую, регулярную поверхность.

Поэтому довольно часто цифровую модель рельефа строят из совокупностей разных математических моделей, стыкующихся в определенных точках. Одним из таких подходов является *метод Вороного-Делоне*.

Пространственные данные могут быть охарактеризованы посредством трех основных категорий элементов: признаки, время, задачи пользователей. Нанесение (наложение) разнородных пространственных данных на единую координатную сетку всегда было мощным инструментом географического анализа. Принципы проектирования ГИС должны обеспечивать хранение и представление данных для реализации этого процесса. Точки, линии и участки могут быть представлены в топологическом (графовом) или растровом формате. Растровый формат более удобен для наложения данных и более пригоден для восприятия в силу своей наглядности, однако области неправильной формы легче представляются в топологическом формате. Основными видами аналитической обработки данных являются наложение и разделение карт и данных для вычисления и измерения участков и поиска по критерию близости.

Расстояния между объектами могут измеряться в прямых координатах или с учетом отмеченных на карте препятствий, причем положение последних может изменяться во времени (например, метеорологические фронты). Создание географической БД складывается из шести этапов: разработка проекта, сбор и подготовка данных, составление карт, интеграция карт.

С целью обеспечения доступа к ГИС разрабатываются специализированные структуры данных (БД пространственной информации) и языки запросов, включающие пространственные операторы (ограничения).

Современные ГИС широко используют не только картографическую и векторную информацию, но и данные фотосъемок и соответственно растровые изображения. Векторные данные называют объектными, растровые — полевыми. Последние широко используются для изучения и выявления новых объектов.

Одним из методов геоинформационного моделирования, позволяющим выявлять новые объекты, является процедура классификации данных. Она основана на том, что все пиксели изображения, имеющие сходные спектральные сигнатуры, объединяются в класс.

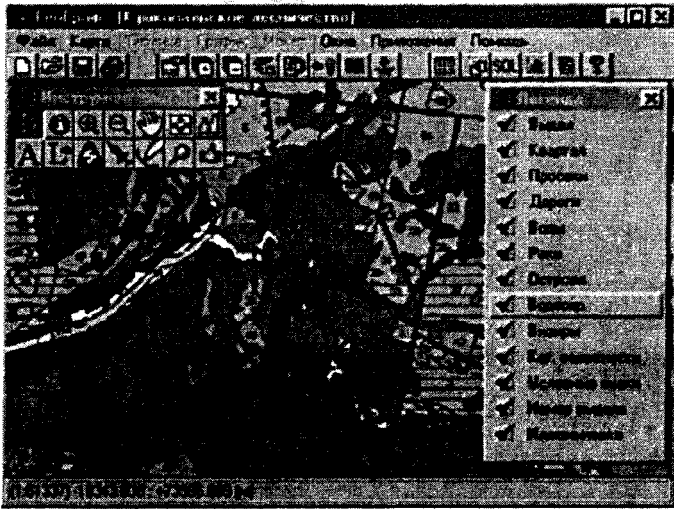
Эта процедура полностью заимствована из методов распознавания образов и является результатом интеграции технологий обработки данных дистанционного зондирования (ДДЗ) в ГИС-технологии.

Классификация заключается в определении принадлежности исследуемого объекта к известному классу. Она производится автоматически по набору признаков, которые задает оператор.

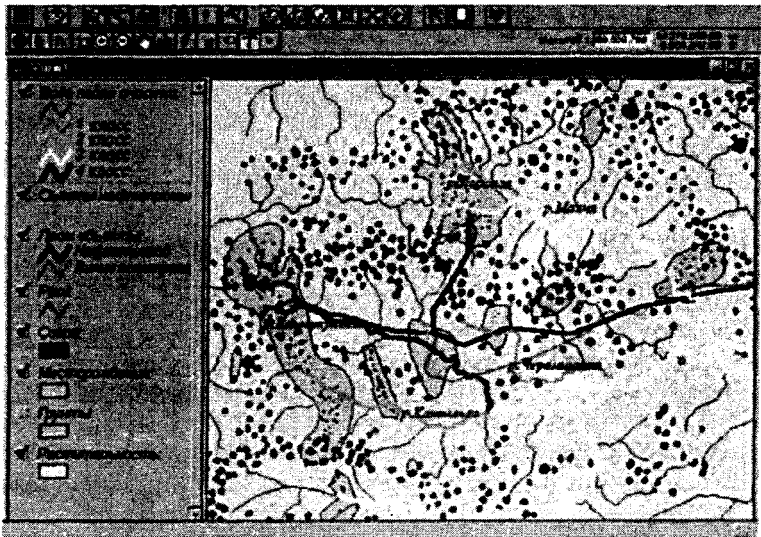
Классификацию можно проводить по одному или нескольким признакам, с обучением или без. Она может быть использована для последующей векторизации изображений или улучшения их качества.

Технология комплекса GeoDraw/ГеоГраф, созданного и развиваемого в Центре Геоинформационных исследований Института географии РАН (ЦГИ ИГ РАН). Технология позволяет получать материалы в соответствии с требованиями действующих нормативных документов (рис. 4.50). Она состоит из следующих основных этапов:

- подготовка единой топоосновы на весь объект;
- оцифровка полигона (границы, просеки) и вычисление геометрических площадей кварталов;
- оцифровка внутренней ситуации кварталов путем сканирования АФС и последующей оцифровки по растру (ранее оцифровка выполнялась на дигитайзерах);
- выдача рабочих планшетов, их считка и редактирование;
- вычисление и приемка площадей;
- подключение таксационных баз данных;
- оформление и печать планшетов, планов лесонасаждений и других выходных материалов;
- оформление электронных карт для передачи заказчику.



а)



б)

Рис. 4.50. Примеры отображения пространственных данных с помощью пакета ГеоГраф:
а — БД по лесному хозяйству; б — БД по нефтехимическим ресурсам

Технология ГИС обещает быстрое развитие всех областей, в которых обрабатываются географические данные. Однако в разработке ГИС между отдельными европейскими странами существуют большой разрыв. В то время как в Великобритании и Нидерландах ГИС уже прочно вошли в практику различных областей (картография, местное управление землепользованием, планирование), в других европейских странах внедрение этой новой технологии проходит чрезвычайно медленно. Существует предположение, что национальные традиции информации о землепользовании определяют организационные условия и процедуры интеграции новой технологии и что это ограничивает возможность использования программного обеспечения универсальной ГИС, разработанной в других национальных условиях.

4.6. Корпоративные интегральные АИС (MRP и ERP-системы)

Системы MRP (Material Requirements Planning)

В начале 60-х гг. в связи с ростом популярности вычислительных систем возникла идея использовать их возможности для планирования деятельности предприятия, в том числе производственных процессов. Необходимость планирования обусловлена тем, что основная масса задержек в процессе производства связана с запаздыванием поступления отдельных комплектующих, в результате чего, как правило, параллельно с уменьшением эффективности производства на складах возникает избыток материалов, поступивших в срок или раньше. Кроме того, вследствие нарушения баланса поставок комплектующих возникают дополнительные осложнения с учетом и отслеживанием их состояния в процессе производства, т. е. фактически невозможно было определить, например, к какой партии принадлежит данный составляющий элемент в уже собранном готовом продукте.

С целью предотвращения подобных проблем, была разработана методология планирования потребности в материалах MRP (Material Requirements Planning). Реализация системы, работающей по этой методологии, представляет собой компьютерную программу, позволяющую оптимально регулировать поставки комплектующих в производственный процесс, контролируя запасы на складе и

саму технологию производства. Главной задачей MRP является обеспечение гарантии наличия необходимого количества требуемых материалов-комплектующих в любой момент времени в рамках срока планирования наряду с возможным уменьшением постоянных запасов. Прежде чем описывать саму структуру MRP, следует перечислить основные понятия.

Материалами являются сырье и отдельные комплектующие, составляющие конечный продукт. В дальнейшем мы не будем делать различий между понятиями «материал» и «комплектующий».

MRP-система, MRP-программа — компьютерная программа, работающая по алгоритму, регламентированному MRP-методологией. Она обрабатывает файлы данных (входные элементы) и формирует на их основе файлы-результаты.

Статус материала является основным указателем на текущее состояние материала: имеется ли данный материал в наличии на складе, зарезервирован ли он для других целей, присутствует ли в текущих заказах или заказ на него только планируется. Таким образом, статус материала однозначно описывает степень готовности каждого материала быть пущенным в производственный процесс.

Страховой запас материала необходим для поддержания процесса производства в случае возникновения непредвиденных и неустраиваемых задержек в его поставках.

Потребность в материале в MRP-программе представляет собой определенную количественную единицу, отображающую возникшую в некоторый момент времени в течение периода планирования необходимость в заказе данного материала. Различают понятия полной потребности в материале, которая отображает то количество, которое требуется отпустить в производство, и чистой потребности, при вычислении которой учитывается наличие всех страховых и зарезервированных запасов данного материала. Заказ в системе автоматически создается по возникновению отличной от нуля чистой потребности.

Процесс планирования включает в себя функции автоматического создания проектов заказов на закупку и/или внутреннее производство необходимых материалов-комплектующих. MRP-система оптимизирует время поставки комплектующих, тем самым уменьшая затраты на производство и повышая его эффективность.

Основными преимуществами использования подобной системы в производстве являются:

- гарантия наличия требуемых комплектующих и уменьшение временных задержек при их поставке, и, следовательно, уве-

личение выпуска готовых изделий без увеличения числа рабочих мест и нагрузки на производственное оборудование;

- уменьшение производственного брака в процессе сборки готовой продукции, возникающего из-за использования несоответствующих технологии комплектующих;
- упорядочивание производства ввиду контроля статуса материалов, позволяющего однозначно отслеживать весь конвейерный путь, начиная от создания заказа на данный материал до его положения в уже собранном готовом изделии. Достигается полная достоверность и эффективность производственного учета.

Основная цель MRP-системы формировать, контролировать и при необходимости изменять моменты заказов таким образом, чтобы все материалы, требуемые для производства, поступали одновременно.

На практике MRP является информационной системой, которую логически можно представить с помощью диаграммы (рис. 4.51). Здесь отображены основные информационные элементы MRP-системы.

Описание состояния материалов (Inventory Status File) является основным входным элементом MRP-программы. В нем должна быть отражена максимально полная информация обо всех материалах, необходимых для производства конечного продукта. В этом

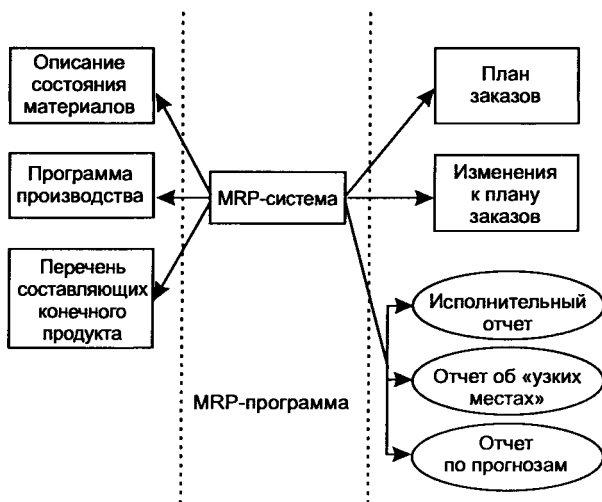


Рис. 4.51. Входные элементы и результаты работы MRP-системы

элементе указывается *статус* каждого материала, участвующего в производственном процессе.

Программа производства (Master Production Schedule) представляет собой оптимизированный график распределения времени для производства необходимой партии готовой продукции за планируемый период или диапазон периодов. Сначала создается пробная программа производства, впоследствии тестируемая на выполнимость дополнительным прогоном через CRP-систему (Capacity Requirements Planning), которая определяет достаточно ли производственных мощностей для ее осуществления. Если производственная программа признана выполнимой, то она автоматически формируется в основную и становится входным элементом MRP-системы.

Это необходимо, потому как рамки требований по производственным ресурсам являются прозрачными для MRP-системы, которая формирует на основе производственной программы график возникновения потребностей в материалах. Однако, в случае недоступности ряда материалов или невозможности выполнить план заказов для поддержания реализуемой с точки зрения CPR производственной программы, MRP-система в свою очередь указывает о необходимости внести в нее корректировки.

Перечень составляющих конечного продукта (Bills of Material File) — это список материалов и их количество, требуемое для производства конечного продукта. Таким образом, каждый конечный продукт имеет свой перечень составляющих. Кроме того, здесь содержится описание структуры конечного продукта, т. е. полная информация по технологии сборки. Чрезвычайно важно поддерживать точность всех записей в этом элементе и соответственно корректировать их всякий раз при внесении изменений в структуру и/или технологию производства конечного продукта.

Напомним, что каждый из вышеуказанных входных элементов представляет собой компьютерный файл данных, используемый MRP-программой. В настоящий момент MRP-системы реализованы на самых разнообразных аппаратных платформах и включены в качестве модулей в большинство финансово-экономических систем.

Цикл работы MRP состоит из следующих основных этапов:

1. Система, анализируя принятую программу, определяет оптимальный график производства на планируемый период.

2. Материалы, не включенные в производственную программу, но присутствующие в текущих заказах, включаются в планирование как отдельный пункт.

На этом шаге, на основе утвержденной программы производства и заказов на комплектующие, не входящие в нее, для каждого от-

дельно взятого материала вычисляется полная потребность в соответствии с перечнем составляющих конечного продукта.

3. На основе полной потребности, учитывая текущий статус материала, для каждого периода времени и для каждого материала вычисляется чистая потребность по указанной формуле. Если чистая потребность в материале больше нуля, то системой автоматически создается заказ на него.

4. Все заказы, созданные ранее текущего периода планирования, рассматриваются и в них при необходимости вносятся изменения, чтобы предотвратить преждевременные поставки и их задержки.

Таким образом, в результате работы MRP-программы производится ряд изменений в имеющихся заказах и при необходимости, создаются новые для обеспечения оптимальной динамики хода производственного процесса. Эти изменения автоматически модифицируют *описание состояния материалов*, так как создание, отмена или модификация заказа, соответственно влияют на статус материала, к которому они относятся. В результате работы MRP-программы создается план заказов на каждый отдельный материал на весь срок планирования, обеспечение выполнения которого необходимо для поддержки программы производства.

Основными результатами MRP-системы являются следующие:

1. *План заказов (Planned order schedule)* — определяет, какое количество каждого материала должно быть заказано в каждый рассматриваемый период времени в течение срока планирования. План заказов является руководством для дальнейшей работы с поставщиками и, в частности, определяет производственную программу для внутреннего производства комплектующих при наличии такового.

2. *Изменения к плану заказов (Changes in planned orders)* — представляют модификации к ранее спланированным заказам. Ряд заказов может быть отменен, изменен или задержан, а также перенесен на другой период.

Также MRP-система формирует некоторые второстепенные результаты в виде отчетов, целью которых является обратить внимание на «узкие места» в течение планируемого периода, т. е. те промежутки времени, когда требуется дополнительный контроль за текущими заказами, а также чтобы вовремя известить о возможных системных ошибках, возникших при работе программы. Итак, MRP-система формирует следующие дополнительные результаты-отчеты:

3. *Отчет об «узких местах» планирования (Exception report)* — предназначен для того, чтобы заблаговременно проинформировать пользователя о промежутках времени в течение срока планирова-

ния, которые требуют особого внимания, и в которые может возникнуть необходимость внешнего управленческого вмешательства. Типичными примерами ситуаций, которые должны быть отражены в этом отчете, могут быть непредвиденно запоздавшие заказы на комплектующие, избытки комплектующих на складах и т. п.

4. *Исполнительный отчет (Performance report)* — является основным индикатором правильности работы MRP-системы и имеет целью оповещать пользователя о возникших критических ситуациях в процессе планирования, таких, как, например, полное израсходование страховых запасов по отдельным комплектующим, а также о всех возникающих системных ошибках в процессе работы MRP-программы.

5. *Отчет о прогнозах (Planning report)* — представляет собой информацию, используемую для составления прогнозов о возможном будущем изменении объемов и характеристик выпускаемой продукции, полученную в результате анализа текущего хода производственного процесса и отчетах о продажах. Отчет о прогнозах также может использоваться для долгосрочного планирования потребностей в материалах.

Таким образом, использование MRP-системы для планирования производственных потребностей позволяет оптимизировать время поступления каждого материала, тем самым значительно снижая складские издержки и облегчая ведение производственного учета.

Эволюция MRP. Переход от MRP к MRP-II

Системы планирования производства постоянно развиваются. Первоначально MRP-системы фактически просто формировали на основе утвержденной производственной программы план заказов на определенный период, что не вполне удовлетворяло возрастающие потребности.

С целью увеличить эффективность планирования в конце 70-х гг. Оливер Уайт и Джордж Плосл предложили идею воспроизведения замкнутого цикла (closed loop) в MRP-системах. Идея заключалась в рассмотрении более широкого спектра факторов при проведении планирования путем введения дополнительных функций. К базовым функциям планирования производственных мощностей и потребностей в материалах было предложено добавить ряд дополнительных, таких, как контроль соответствия количества произведенной продукции количеству использованных в процессе сборки комплектующих, составление регулярных отчетов о задержках заказов, об объемах и динамике продаж продукции, о поставщиках и т. д.

Термин «замкнутый цикл» отражает основную особенность модифицированной системы, заключающуюся в том, что созданные в процессе ее работы отчеты анализируются и учитываются на дальнейших этапах планирования, изменяя при необходимости программу производства, а следовательно, и план заказов. Другими словами, дополнительные функции осуществляют обратную связь в системе, обеспечивающую гибкость планирования по отношению к внешним факторам, таким, как уровень спроса, состояние дел у поставщиков и т. п.

В дальнейшем, усовершенствование системы привело к трансформации системы MRP с замкнутым циклом в расширенную модификацию, которую впоследствии назвали MRP-II (Manufactory Resource Planning) ввиду идентичности аббревиатур. Эта система была создана для эффективного планирования всех ресурсов производственного предприятия, в том числе финансовых и кадровых.

Одним из наиболее распространенных методов управления производством и дистрибуции в мире является стандарт MRP-II, разработанный в США и поддерживаемый американским обществом по контролю за производством и запасами — American Production and Inventory Control Society (APICS). APICS регулярно издает документ «MRP-II Standart System», в котором описываются основные требования к информационным производственным системам.

MRP-II — это набор принципов, моделей и процедур управления и контроля, служащих повышению показателей экономической деятельности предприятия.

MRP II Standart System содержит описание шестнадцати групп функций системы:

- 1) планирование продаж и производства (Sales and operation planning);
- 2) управление спросом (Demand management);
- 3) составление плана производства (Master production scheduling);
- 4) планирование материальных потребностей (Material requirement planning);
- 5) спецификации продуктов (Bill of materials);
- 6) управление складом (Inventory transaction subsystem);
- 7) плановые поставки (Scheduled receipts subsystem);
- 8) управление на уровне производственного цеха (Shop flow control);
- 9) планирование производственных мощностей (Capacity requirement planning);
- 10) Input/output control (Контроль входа/выхода);

- 11) материально-техническое снабжение (Purchasing);
- 12) планирование распределения ресурсов (Distribution resource planning);
- 13) планирование и контроль производственных операций (Tooling planning and control);
- 14) управление финансами (Financial planning);
- 15) моделирование (Simulation);
- 16) оценка результатов деятельности (Performance measurement).

С накоплением опыта моделирования производственных и непроизводственных операций эти понятия постоянно уточняются, постепенно охватывая все больше функций.

Задачей информационных систем класса MRP-II является оптимальное формирование потока материалов (сырья), полуфабрикатов (в том числе находящихся в производстве) и готовых изделий. Система класса MRP-II имеет целью интеграцию всех основных процессов, реализуемых предприятием, таких, как снабжение, запасы, производство, продажа и дистрибуция, планирование, контроль за выполнением плана, затраты, финансы, основные средства и т. д.

Стандарт MRP-II делит сферы отдельных функций (процедур) на два уровня: необходимый и опциональный. Для того чтобы программное обеспечение было отнесено к классу MRP-II, оно должно выполнять определенный объем необходимых (основных) функций (процедур). Некоторые поставщики ПО приняли различный диапазон реализаций опциональной части процедур этого стандарта.

Результаты использования интегрированных систем стандарта MRP-II:

- получение оперативной информации о текущих результатах деятельности предприятия как в целом, так и с полной детализацией по отдельным заказам, видам ресурсов, выполнению планов;
- долгосрочное, оперативное и детальное планирование деятельности предприятия с возможностью корректировки плановых данных на основе оперативной информации;
- решение задач оптимизации производственных и материальных потоков;
- реальное сокращение материальных ресурсов на складах;
- планирование и контроль за всем циклом производства с возможностью влияния на него в целях достижения оптимальной эффективности в использовании производственных мощностей, всех видов ресурсов и удовлетворения потребностей заказчиков;

- автоматизация работ договорного отдела с полным контролем за платежами, отгрузкой продукции и сроками выполнения договорных обязательств;
- финансовое отражение деятельности предприятия в целом;
- значительное сокращение непроизводительных затрат;
- защита инвестиций, произведенных в информационные технологии;
- возможность поэтапного внедрения системы с учетом инвестиционной политики конкретного предприятия.

В основу MRP-II положена иерархия планов (рис. 4.52). Планы нижних уровней зависят от планов более высоких, т. е. план высшего уровня предоставляет входные данные, намечаемые показатели и/или какие-то ограничительные рамки для планов низшего уровня. Кроме того, эти планы связаны между собой таким образом, что результаты планов нижнего уровня оказывают обратное воздействие на планы высшего уровня.



Рис. 4.52. Иерархическая последовательность планов в MRP-II

Стратегическое планирование — это долгосрочное планирование. Оно обычно составляется на срок от одного до пяти лет и основано на макроэкономических показателях, таких, как тенденции развития экономики, изменение технологий, состояние рынка и конкуренции. Стратегическое планирование обычно распространяется на каждый год пятилетки и представляет собой плановые показатели (цели) высшего уровня.

Бизнес-план — это обычно план на год, который также составляется на ежегодной основе. Иногда он неоднократно пересматривается в течение года. Как правило, он является результатом совещания управленческого состава, на котором сводятся планы продаж, инвестиций, развития основных средств и потребности в капитале и бюджетирование. Эта информация подается в денежном выражении. Бизнес-план определяет плановые показатели по объемам продаж и производства, а также другие планы низшего уровня.

Планирование объемов продаж и производства. Если бизнес-план предоставляет итоговые данные по объемам продаж ежемесячно (в денежном выражении), то план объемов продаж и производства разбивает эту информацию на 10—15 ассортиментных групп. В результате получают план производства, который ежемесячно пересматривается, принимая во внимание план предыдущего месяца, реальные результаты и данные бизнес-плана.

План объемов продаж и производства обычно включает следующие элементы: объем продаж, производства, запасы, незавершенный объем производства, отгрузка.

Из этих элементов *объем продаж* и *отгрузка* — это прогнозы, так как это внешние данные, которые прямому контролю не поддаются. *Объем производства* планируется, это внутренний показатель, поддающийся прямому контролю. *Планы по объемам запасов и незавершенного производства* контролируются косвенно, манипулируя данными прогнозов объема продаж, отгрузки и/или плана объемов производства.

Планирование ресурсов. План производства будет нереален, если не будет обеспечено наличие необходимого объема ресурсов. Планирование ресурсов — это долгосрочное планирование, которое позволяет оценить необходимый (для выполнения плана производства) и наличный объем ключевых ресурсов, таких, как люди, оборудование, здания и сооружения. Если возникнет потребность в наличии дополнительных ресурсов, то, возможно, потребуется пересмотреть бизнес-план.

Главный план-график производства (ГППГ) — план производства, наложенный на шкалу времени. ГППГ показывает, что будет производиться, когда и в каких объемах.

Ввиду того, что производственный план выражен в таких единицах, как рубли, часы, тонны, то для получения ГППГ необходимо произвести некоторые шаги по трансформации производственного плана. Плановые объемные показатели по ассортиментной группе необходимо перевести в плановые объемы и сроки по каждому продукту в отдельности. В зависимости от типа и объема выпускаемой продукции ГППГ можно разбить на недельные, дневные и даже сменные планы.

Одна из основных целей ГППГ — это обеспечение буфера: ГППГ отличают прогнозы и потребности отдела сбыта от MRP (планирование потребностей в материалах). Прогнозы и заказы на продажу (заказы клиентов) выражают спрос (или отгрузку), в то время как ГППГ — то, что реально будет произведено. В соответствии с ГППГ возможно производство продукции в период, когда спрос невысок, и наоборот, что может иметь место при производстве продукции, спрос на которую сезонен.

Общее планирование мощностей. Как и планирование ресурсов, общее планирование мощностей является долгосрочным и ведется по ключевым ресурсам. Этот процесс использует данные ГППГ, а не производственного плана. Так если ГППГ выражен в объемных и временных характеристиках, то общее планирование мощностей используется для создания более детализированного плана, который может быть очень полезен при оценке средних потребностей компании в целом.

MRP или планирование потребностей в материалах. Исторически MRP предназначалось для контроля за запасами и их пополнения (см. выше). В рамках MRP-II (планирование ресурсов предприятия) его использование было расширено до планирования потребностей в мощностях и замыкания всей цепочки планирования.

CRP или планирование потребностей в мощностях. Наличие необходимого объема определенных материалов ничего не значит при отсутствии достаточного объема свободного рабочего времени. CRP — это планирование среднего уровня, которое использует данные MRP и заказов на производство для определения необходимого объема рабочего времени (как по трудовым, так и по техническим ресурсам).

Планирование ресурсов и общее планирование мощностей — это планирование высшего уровня, используемое для управления такими ресурсами, как физическое оборудование. CRP является бо-

лее детализированным. Загрузка рабочих мест рассчитывается на основе технологического маршрута, который определяет, каким именно образом производится данный вид продукта. Технологический маршрут похож на инструкцию к применению — набор шагов (или техопераций), которые необходимо совершить для изготовления чего-либо. Каждая техоперация совершается на отдельном рабочем месте, которое может требовать наличия одного или нескольких человек и/или единиц оборудования.

DRP или планирование потребностей в распределении. Когда материалы передвигаются от поставщика к потребителю, они проходят по цепи поставок (или рыночному каналу). Если представить это графически, то цепь поставок представляет собой потоки спроса и предложения между поставщиками и подразделениями компании-заказчика, между этими подразделениями и клиентами или различными подразделениями одной компании. DRP координирует спрос, предложение и ресурсы между подразделениями одной или нескольких компаний.

Системы ERP (Enterprise requirements planning)

В дальнейшем процессе развития корпоративных АИС, системы планирования MRP-II в интеграции с модулем финансового планирования FRP (Finance requirements planning) получили название систем бизнес-планирования ERP (Enterprise requirements planning), которые позволяют наиболее эффективно планировать всю коммерческую деятельность современного предприятия, в том числе финансовые затраты на проекты обновления оборудования и инвестиции в производство новой линейки изделий. В российской практике целесообразность применения систем подобного класса обуславливается, кроме того, необходимостью управлять бизнес-процессами в условиях инфляции, а также жесткого налогового пресса, поэтому системы ERP необходимы не только для крупных предприятий, но и для небольших фирм, ведущих активный бизнес. На рис. 4.53 представлена логическая схема системы планирования ресурсов производственного предприятия.

Хотя благодаря автоматизации и интеграции бизнес-операций ERP-системы и могут повлиять на практические результаты работы, они мало отражаются на самом важном — расширении возможностей деловой активности, росте доли на рынке, увеличении продаж и эффективности бизнеса, а также на повышении ценности бизнеса в целом. Комбинация этих процессов, объединенных в одну систе-

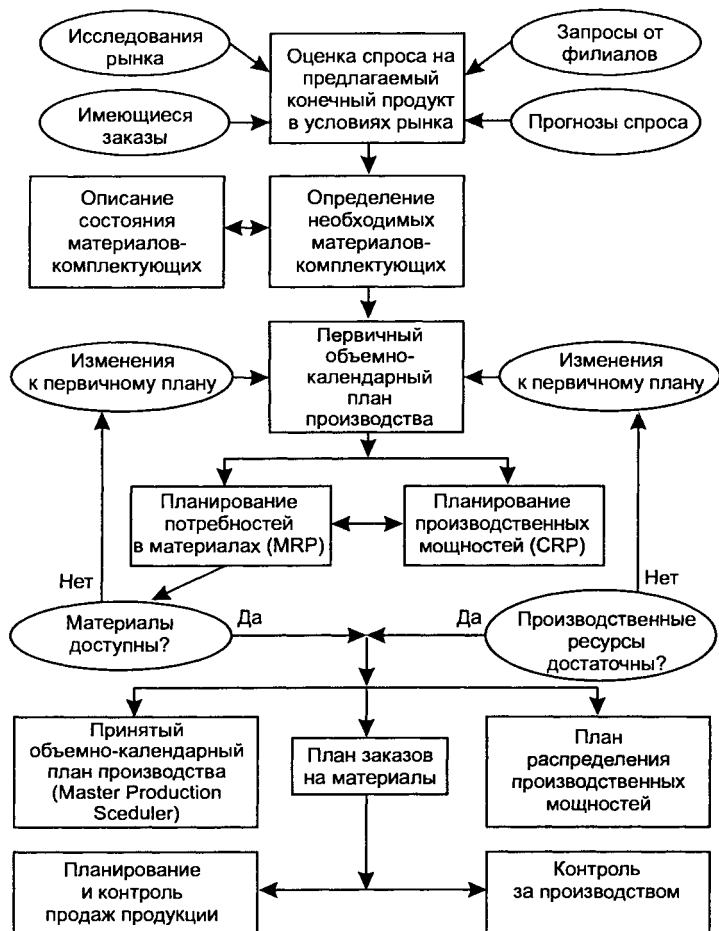


Рис. 4.53. Логическая структура системы планирования ресурсов производственного предприятия

му, называется планированием и контролем менеджмента (Management Planning and Control, MPC).

Системы MPC (Management Planning and Control)

Чтобы использовать MPC-систему для разрешения проблемных процессов головного офиса, необходимо применить такой же подход к управлению данными, как и для ERP-пакетов: он должен га-

рантировать более точное, своевременное и детальное представление о состоянии бизнеса.

Решением является применение интегрированного пакета приложений, предназначенных для поддержки оптимальных методов стратегического планирования, бюджетирования, прогнозирования, финансовой консолидации, управленческой отчетности и анализа. Пакет МРС-приложений позволяет предприятию создавать планы и бюджеты, управлять их исполнением, сравнивать фактическое исполнение с запланированным, выделять исключения и моделировать решения на будущее.

Архитектура, лежащая в основе МРС-решения, позволяющего более эффективно осуществлять управление и реализовывать стратегии, представляет собой:

- многоуровневый проект, где приложение функционально распределено на клиентский, серверный компонент и компонент базы данных;
- централизованную базу для управления планируемыми, актуальными, консолидированными и прогнозируемыми данными и результатами;
- общую бизнес-модель со встроенной финансовой и временной логикой, предназначенную для обработки финансовых отчетов и данных анализа, а также различных временных показателей, без дополнительного программирования со стороны пользователя;
- Интернет-среду, поддерживающую и направляющую пользователя в рамках процессов бюджетирования, отчетности и анализа;
- расширенные возможности нерегламентируемого анализа для быстрой идентификации, обнаружения и проверки аномалий, проблем и благоприятных возможностей;
- всеобъемлющую систему защиты от несанкционированного доступа или изменения в различных частях базы данных;
- централизованное администрирование, обеспечивающее контроль и мониторинг процессов планирования, бюджетирования, консолидации и отчетности.

В основе МРС-системы лежит централизованная реляционная база, поддерживающая полный пакет приложений. Здесь хранятся не только данные и результаты бюджетирования, но также ключевые бизнес-факторы, полученные на основе стратегического планирования, а также реальных и прогнозируемых данных и результатов. И если в приложении используются отдельные файлы, хранящиеся на различных компьютерах, то данные, вводимые в МРС-базу, и ре-

зультаты, рассчитываемые по ним, сразу же становятся доступными для дальнейшего анализа, поскольку их уже не нужно перемещать из одного приложения в другое. Реляционность МРС-системы позволяет масштабировать пакет для поддержки больших объемов данных и более широкого круга пользователей, а также легко интегрировать его с существующими ERP-системами и другими операционными приложениями.

В целом, это означает, что лежащая в основе реляционная схема обеспечивает пользователей приложения многомерным представлением данных и результатов. Оно позволяет осуществлять планирование, бюджетирование, отслеживание и прогнозирование по разным измерениям. Например, пользователь может составлять бюджет продаж, расходов и доходов в конкретной валюте, для некоторого периода времени по определенным продуктам, местам расположения и каналам распространения. Бюджетные показатели можно сравнивать с фактическими результатами по тем же самым измерениям.

На рынке программных средств установились критерии оценки МРС-продуктов, поскольку последние коренным образом отличаются от любых других финансовых аналитических приложений. Приведем девять критериев, которые помогут пользователю отличить настоящие МРС-системы от просто *интегрированных приложений*.

1. *Интегрированные процессы.* Необходимо убедиться, что в системе поддерживается планирование, бюджетирование, консолидация, административная отчетность и анализ в одном замкнутом приложении. Пользователь должен иметь возможность переключаться между процессами, не меняя рабочей среды и не перемещая данные, а руководство — быстро реагировать на все изменения, оценивая альтернативы, регулируя планы и информируя пользователей без дополнительного изучения и поддержки множества технологий и приложений.

2. *Общие бизнес-правила.* Система должна иметь общий набор измерений, бизнес-элементов и правил, хотя последние можно ограничить специфическими процессами. В результате при изменении структуры или элемента связанные с ними отчеты и результаты анализа обновляются автоматически. По тем же причинам всегда будет обрабатываться только один набор данных, даже если требуется несколько версий. Это означает, что любое число, стоящее в плане, бюджете или в наличности, хранится и обрабатывается только в одном месте.

3. *Встроенная финансовая логика.* Система должна автоматически генерировать финансовые отчеты и выполнять связанный с ними анализ, не требуя вмешательства программиста. По мере того как пользователь манипулирует данными, система должна корректно обрабатывать дебиты/кредиты, доходы/убытки и бухгалтерские счета.

4. *Моделирование будущего.* Все бизнес-правила должны свободно использовать любые элементы в БД, а ПО — выполнять консолидацию неоднократно. Используя подобные правила, система сможет работать с рядом задач (как, например, распределение и расчет долей участия дочерних компаний и пр.), которые иным способом решить не удастся.

5. *Временная логика.* В системе должны обрабатываться различные временные показатели, относящиеся к процессам, отчетности и другим функциям, таким, как периодичность учета, скользящие бюджеты (rolling budgets) и показатели за текущий год. Применение этой логики упрощает настройку систем и их перевод на непрерывное планирование (continuous planning).

6. *Стимуляция взаимодействия.* В любой момент времени без предварительного уведомления пользователь должен иметь онлайн-доступ к данным, которые он без всяких ограничений может просматривать и анализировать по всем допустимым измерениям, как-то: проект, продукт, направление деятельности и т. п. Кроме того, необходимо располагать возможностью вращать и представлять измерения во вложенном виде, а также выполнять детализацию («углубление в данные») внутри модели.

7. *Управляемый анализ.* МРС-системы должны обращать внимание пользователя на исключения и предоставлять ему специальные инструменты для детального анализа отклонений. Уведомления об исключениях должны приходить автоматически, иначе последние могут остаться незамеченными.

8. *Интернет-архитектура.* Для эффективного внедрения на предприятии архитектура МРС-решения должна быть основана на Интернет-технологиях. Это означает, что пользователи вне зависимости от уровня подготовки могут обратиться к системе через Web-браузер с любого удаленного компьютера.

9. *Централизованная база данных.* МРС-приложения работают на централизованной базе данных, не используя «фирменные» («специальные») файловые структуры, которые часто применяются в современных системах. Эта база должна основываться на широко распространенной реляционной технологии, что обеспечит масштабируемость и надежность. Применяя существующую технологию

можно максимально использовать вложенные в инфраструктуру инвестиции и минимизировать расходы на реализацию проекта и обучение персонала.

МРС-системы открывают новую веху на рынке финансовых продуктов и наверняка окажутся не менее важными, чем ERP, когда финансовое управление компаний перейдет от установки на высокую производительность к установке на высокую результативность.

4.7. Другие типы систем и прикладных программных продуктов

Информационные системы вузов (АИС ВУЗ)

Университетские ИС (АИС ВУЗ) ориентированы прежде всего на информационное обеспечение основных функций университета: научные исследования, обучение, профессиональную подготовку, финансово-бухгалтерский учет, коммерческую и административно-управленческую деятельность. Она обеспечивает как внутреннюю коммуникацию между подсистемами, так и связь с внешними ИС, и рассчитана на удовлетворение информационных потребностей внутренних и внешних пользователей. В структуре АИС ВУЗ выделяются следующие подсистемы:

1) членов университета (вузы, факультеты, отделения, библиотека, ВЦ и т. п.);

2) пользователей учебно-исследовательских услуг (предприятия, организации, общества);

3) учебно-исследовательских обществ (федеральные, специальные, региональные);

4) общественных организаций (федеральные, региональные, городские), а также другие подсистемы (статистики, найма на работу, здравоохранения и т. п.).

Содержание АИС ВУЗ определяется информационными потребностями пользователей; выделяются индивидуальные (специфические) и общие (интегральные) потребности. Отдельные подсистемы АИС ВУЗ могут передавать информацию из своих БД другим подсистемам с помощью различных средств: ЭВМ (интерактивный режим, электронная почта); массовых средств (печать, радио, телевидение); бюллетеней и статей; периодических изданий; отчетов и др.

Традиционный состав подсистем АИС ВУЗ:

- администрация учебного заведения;
- сведения о студентах (личные и академические данные студентов, семестровая и годовая успеваемость, контроль посещаемости занятий);
- информация о преподавателях университета;
- финансовые службы;
- библиотечное и информационное обслуживание.

Для информационного обеспечения планирования и управления в вузах прежде всего необходимо выделить различные виды статистической отчетности. В качестве модели документа (БД), обеспечивающей необходимую полноту отчетности, в США предложена Единая карта учета деятельности учебного заведения, требованиями к которой являются:

- информативность и полезность для всех потребителей, — студентов, родителей, попечителей учебного заведения, абитуриентов;
- роль дополнительного источника сведений при оценке ресурсов, достижений и планов каждого учебного заведения;
- содержание сравнительной информации, учет не только текущего состояния, но и перспектив;
- простота для заполнения.

Карта включает следующие основные показатели:

- основные демографические данные по контингенту студентов, их динамике;
- данные о достижениях студентов;
- успеваемость и ее динамика;
- степень соответствия тестам;
- информация о дополнительной профессиональной подготовке студентов или местах трудоустройства;
- численность студентов и ее динамика, отсев и пр.;
- расходы и финансирование, с выделением статей по заработной плате, социальным выплатам, стипендиям; в том числе на одного человека;
- размеры студенческих групп и преподавательская нагрузка, их динамика;
- штатное расписание преподавателей, их должности и аргументы в пользу их избрания, читаемые дисциплины.
- учебники и учебные пособия, степень обеспеченности ими студенческих работ, характеристика библиотеки, объем фонда, интенсивность обслуживания;

- оборудование учебных и лабораторных площадей, количество единиц оборудования, в т. ч. на одного человека;
- уровень квалификации преподавателей и возможности их профессионального роста;
- усовершенствование учебных программ, планы и программы развития.

Данная карта предоставляет исчерпывающую информацию об организации и функционировании учебного заведения.

На аналогичном уровне содержат информацию базы данных международной системы TRACE, включающей в себя три подсистемы, которые характеризуют:

- 1) национальную систему образования, в том числе структуру, персоналии, статистику и тенденции;
- 2) отдельные вузы (в тех же аспектах, как и вышеуказанная Единая карта);
- 3) документы, удостоверяющие уровень и направление подготовки, квалификационные требования, порядок выпускной и/или промежуточной аттестации.

Необходимо отметить, что накопление значительного объема информационных машиночитаемых ресурсов, включающих:

- программы курсов;
- образовательные стандарты на специальности и описания специализаций;
- информационные профили и научные интересы преподавателей,

позволяет поставить задачу разработки автоматизированных систем, ориентированных на построение оптимальных учебных планов и программ подготовки, или навигации в сети взаимосвязанных учебных дисциплин, обеспечивающей необходимый уровень подготовки в рамках заданных временных и ресурсных ограничений, с учетом диффузного характера учебных дисциплин и определенной взаимозаменяемости преподавательского контингента.

Информационные системы медицинских учреждений (Hospital Information System — HIS)

Усложнение современной медицины привело к интеграции систем автоматизации медицинского обслуживания. Больницам нужны системы, способные одновременно помогать врачу в постановке диагноза и выборе курса лечения, контролировать ведение последнего, изменяя его ход в соответствии с состоянием пациента, планировать

использование ресурсов госпиталя, собирать информацию для медицинской статистики и страховых компаний, рассчитывать расходы на лечение, формировать необходимые счета и выполнять массу других медицинских и административных задач. Разработка систем, отвечающих всем предъявляемым условиям, требует физического разделения данных при сохранении равноправного доступа к ним всех заинтересованных участников, множественности логических представлений единых физических данных, языка запросов с возможностью определения пользовательских типов данных и функций, развитых средств сбора и обобщения информации и создания прототипов.

АИС в больницах и поликлиниках внедряются в развитых странах с начала 60-х гг. Эти АИС весьма многообразны — от систем на ПЭВМ до терминальных сетей больших ЭВМ (и их комплексов). Известны следующие типы этих систем:

- лабораторные АИС (регистрация результатов анализов, иногда управление их выполнением);
- ИС пациентов (выполняет регистрацию пациентов, ведет истории болезней и архив, планирует и контролирует поток пациентов);
- документальные (медицинская информация и доступ к медицинской библиотеке);
- диагностические (клинические ЭС — БЗ промежуточных гипотез и решений, языковой интерфейс, обработка клинических задач и система объяснений);
- организационные (способствуют выполнению административных и финансовых функций).

В начале 80-х гг. сложилась концепция интегрированных больничных АИС как сетей ПЭВМ (распределенная обработка), поддерживаемых большой ЭВМ. С помощью системы осуществляется ввод и хранение данных об анамнезе и статусе больных при поступлении в клинику, о ранее проведенном и текущем лечении и его результатах, обработка этих данных (сопоставление, статистика), подготовка документации.

АИС медсестринской службы имеют три главных аспекта применения: планирование ухода за больными, регистрация рабочей нагрузки персонала, ведение графиков дежурств. Некоторые из этих ИС обеспечивают структурированный перечень видов деятельности персонала на каждый день недели. При этом для каждой процедуры определены продолжительность затрачиваемого времени и квалификация медсестры. Таким образом, с помощью ЭВМ легко определить количество и качество штата среднего и младшего медицин-

ского персонала для каждого отделения больницы. При учете рабочей нагрузки персонала больные делятся на категории по степени тяжести их состояния. Для каждой категории устанавливается время, затрачиваемое на уход в течение суток, и квалификация медсестры. Такой метод позволяет предусмотреть определенную гибкость графика.

С помощью ЭВМ составляется стандартный график, который автоматически корректируется в зависимости от времени суток, дня недели, категории больных. Важное достоинство ИС в том, что они облегчают и упорядочивают финансовые расчеты — это особенно важно, если учесть, что стоимость медсестринской службы составляет значительную часть бюджета системы здравоохранения. ИС помогают организовать контрактную работу, позволяют систематически оценивать качество и стоимость обслуживания больных, выявлять тенденции и прогнозировать необходимые изменения штата.

Прочие профессионально-ориентированные АИС (рынок прикладного программного обеспечения России)

На сегодняшний день на российском рынке представлено около 800 различных программ автоматизации бухучета. Они могут быть разбиты на следующие группы:

- для крупных предприятий;
- для средних;
- для малых предприятий и индивидуальных предпринимателей;
- для физических лиц.

На этом рынке ведущие позиции занимают продукты компаний «1С», «Инфо-бухгалтер», «Фолио», «Светон», «ДИЦ», «АиТ», «Компьютер-сервис», «Парус», «Инотек», «Интеллект-сервис», «Гринсофт», «Ланкс».

Рынок российского программного обеспечения начал бурно развиваться в конце 80-х — начале 90-х гг. с появлением у нас первых персональных компьютеров и развитием бизнеса. Лучшие программисты заводов и «почтовых ящиков» начали выпускать программы не только для внутренних нужд, но и на заказ.

Ниже перечислены основные направления на рынке российских разработок:

- бухгалтерские программы;
- финансовые программы;
- складские системы;
- программы для строительных организаций;

- программы для делопроизводства;
- редакторы и корректоры;
- графические системы;
- системы распознавания текста;
- системы перевода текстов;
- словари;
- гипертекстовые системы;
- САПР;
- архитектурные программы;
- системы для научно-технических расчетов;
- системы управления проектами;
- юридические и правовые системы;
- голосовые технологии;
- системы коммуникации.

Бухгалтерские системы. Лидером на рынке бухгалтерского программного обеспечения можно по праву назвать ПП «1С:Бухгалтерия» («1С»). Это универсальная бухгалтерская программа, подходящая для ведения как простого, так и сложного учета. Известны версии: «1С:Бухгалтерия базовая 7.7», «1С:Бухгалтерия Проф 7.7». Базовую версию отличает простота в освоении и использовании, сочетающаяся с мощными возможностями по синтетическому и аналитическому учету. Версия «Проф» предназначена для ведения сложного учета (валютные операции, учет операций в торговле, аудит).

Основой программы «1С:Бухгалтерия» является журнал хозяйственных операций. Все операции предприятия вводятся в него в виде бухгалтерских проводок или документов. Проводки могут автоматически формироваться после ввода документа. Помимо классической проводки «дебет—кредит—сумма» в журнале операций хранится информация об объектах аналитического учета, количестве и валюте. При вводе информации программа позволяет получать оперативные данные об остатках товаров на складе, задолженности организаций, остатках средств на счетах, в кассе и др. Используя режим формирования произвольных отчетов, можно изменять существующие и создавать собственные отчеты и справки как для внутреннего учета, так и для внешней отчетности. Начать работу с программой можно с самых простых операций типа печати платежных поручений, ввода проводок, а более сложные функции осваивать постепенно.

Наиболее распространенной программой для малого и среднего бизнеса является на сегодняшний день программа «Инфо-бухгалтер» фирмы «Инфо-бухгалтер» (ранее известной как ТОО «Информатик»). Имеются следующие версии: минимальная 8.1, профессиональная 8.1. Существует версия для бюджетных организаций, а так-

же блок финансового анализа для Windows-версий. Для ведения учета на нескольких компьютерах разработан сетевой вариант.

Все версии рассчитаны на неподготовленного пользователя, имеют дружелюбный интерфейс, большой набор типовых операций, широкие возможности аналитического учета. Программа выполняет все необходимые расчеты и имеет исчерпывающее описание. В ней можно рассчитать зарплату, налоги, амортизацию основных средств, провести финансовый анализ предприятия.

Рассмотрим другую программу бухучета — «Турбо-бухгалтер» фирмы ДИЦ, которая выходит в следующих версиях: «Турбо-бухгалтер 6.5 Лайт», «6.5 Про».

Принципиальным отличием «Турбо-бухгалтер» является ведение многомерного и многоуровневого аналитического учета: до нескольких десятков аналитических признаков с неограниченной вложенностью. Все данные учета и выходных документов представляются в текстовом виде, что обеспечивает открытость информации и простоту работы для бухгалтера.

«Инотек бухгалтер» — первая Windows-бухгалтерия, созданная российскими разработчиками — фирмой «Инотек». В настоящее время существует версия Inotec Accountant Western, имеющая две бухгалтерские книги, позволяющая вести параллельный учет одновременно по двум стандартам: российскому и любому международному, включая ГААР. «Инотек бухгалтер Професионал 6.0» — модуль, имеющий все функциональные особенности Inotec Accountant Western. Программу отличает лучшее из всех представленных соотношение цена—качество.

Если бухгалтеру не нужны комплексные программы по бухучету, он может выбрать ту, которая автоматизирует тот или иной сектор бухгалтерской отчетности. На рынке представлено много таких программ, например «Платежка+касса» и «Счет-фактура» фирмы Трек, «1С:Платежные документы», «Фолио-Платежные документы» и др.

Складские системы. Одной из наиболее продаваемых программ в розничной сети Москвы является «Win-склад» фирмы «Фолио» и ее модификации, автоматизирующие складской учет от рабочего места продавца до директора. Наиболее мощная система «Win-склад Мастер-класс» полностью учитывает приход—расход—оплату товара, формирует накладные сборки в производстве, работает с комплектами, имеет шестиуровневый классификатор товаров, ведет учет в любой валюте, контролирует задолженность организаций, дает отчеты с реализаторами, анализирует движение товара по любым признакам по любому числу складов, имеет генератор отчетов, форм и прайс-листов. Можно передавать данные в системы «Фолио»,

«1С:Бухгалтерия», «Турбо-бухгалтер», «Инфо-бухгалтер», «Инотек бухгалтер». Имеется контекстная обучающая система и версия «клиент/сервер».

Программа «Фарм-склад» предназначена для производственных и торговых фармацевтических предприятий, «Прод-склад» — для фирм, торгующих продовольственными товарами или использующих учет партий, серий и сроков годности.

Кадровые системы. Это программы, автоматизирующие отделы кадров предприятий и организаций. Одна из наиболее мощных клиент-серверных программ на рынке — «Управление персоналом» (серверный вариант на платформе MS SQL) фирмы «АиТ». Программа представляет собой комплекс совместимых программных продуктов, предназначенных для решения задач управления персоналом на предприятиях различного типа: от небольших коммерческих фирм до крупных административных структур. Она включает в себя несколько функциональных модулей: «Зарплата», «Кадры», «Учет коллективного заработка», «Табельный учет», «Передача данных в Пенсионный фонд». Программа работает как на локальных рабочих местах, так и в масштабе всего предприятия, позволяет вести расчет любых видов оплаты труда на базе настройки расчетных формул, хранение досье на сотрудников, ведение штатного расписания и системы приказов. Она обеспечивает формирование необходимой бухгалтерской и кадровой отчетности по работе с персоналом. Имеется возможность совместного использования с системами Platinum SQL, «Алеф-бухгалтерия», «1С:Бухгалтерия», «Инфо-бухгалтер», «Турбо-бухгалтер», «Суперменеджер», а также передачи сведений о персонале в Госналогслужбу и Пенсионный фонд. Любой модуль может быть приобретен отдельно.

Комплексные системы предприятия. Очень часто фирмы-разработчики предлагают комплексные бухгалтерские, складские, торговые и финансовые программы. Один из таких разработчиков — фирма «Интеллект-сервис». «БЭСТ-4» — комплексная бухгалтерская, торговая и складская система, которая проводит бухучет в полном объеме, учет основных фондов, материалов, рассчитывает зарплату и поддерживает мощную аналитику. В торгово-складском модуле имеется связь с кассовыми аппаратами. Для этого предлагается «БЭСТ-КПМ» — кассово-программный модуль. Для оценки финансово-хозяйственной деятельности предназначена «БЭСТ-Ф», имеющая возможность обмена данными с «БЭСТ-4». Для средних предприятий предлагается «БЭСТ-2+», а для малого бизнеса — «БЭМБИ+», работающая при минимальной конфигурации компьютера.

Фирмой «Инфософт» разработано несколько аналогичных комплексов. Сетевая бухгалтерия «Интегратор» предназначена для автоматизации крупных и средних фирм с выявлением финансовых результатов деятельности и получением всех форм бухгалтерской отчетности. Система «Зарплата» имеет взаимосвязь с «Интегратором»; система «Кадры» предназначена для учета личного состава предприятий любого типа, включая государственные и бюджетные; система «Складской учет» предлагается оптовым и мелкооптовым фирмам и отделам сбыта промышленных предприятий.

«Конструкторы» автоматизации. Все большее значение придается автоматизации складов и торговли. Одной из главных проблем при выборе соответствующей программы является многообразие форм и методов ведения учета в торговых организациях. Всех потенциальных клиентов в этой области можно условно разбить на две категории: «традиционные» и «нестандартные». К категории «традиционных» относятся клиенты, придерживающиеся на практике типовых форм учета. Большинство представленных на рынке программ предназначены именно для таких клиентов. Однако в последнее время все больше и больше организаций используют в своей работе нестандартные методы учета. Для таких клиентов выбрать необходимую систему, которая учитывала бы их специфические требования, очень трудно. Многие решают для себя эту проблему просто: ищут разработчика, который пишет программу под заказ, но это, как правило, долго и болезненно. Лучше всего в данной ситуации приобрести универсальную программу-конструктор, имеющую в своем составе гибкие средства настройки. Стоимость адаптации этих программ к требованиям клиента существенно ниже стоимости создания новых разработок под заказ. Наиболее яркие представители этого класса программ-конструкторов — «1С:Торговля» от «1С» и «Торговый склад» фирмы «Компьютер-сервис».

На рынке представлено несколько версий этой программы: «Торговый склад», «Розничная торговля», «Промышленный склад», «Книжный склад». Все они дают максимальную гибкость настроек: основного меню для различных работников, справочников и рабочих баз данных. Существуют и более сложные настройки, например проектирование собственных печатных форм первичных документов, создание новых отчетов. Архитектурное построение программ позволило разработчикам решить одну из наиболее сложных проблем — контроль товаров на реализации. Программа позволяет делать проводки в «1С:Бухгалтерию», «Инфо-бухгалтер», «Турбо-бухгалтер», «Янус», «Парус».

Персональные финансовые программы. Они предназначены для ведения домашних финансов, например — «1С:Деньги». Этот продукт позволяет рассчитать расходы и поступления, куплю-продажу товаров, ценных бумаг, имущества, валюты, напомнит о долгах и кредитах, поможет не только контролировать, но и планировать домашнюю финансовую деятельность, напечатает отчеты, платежные поручения и бланки.

Компанией «Арсеналь» разработана программа ведения личных финансов «Декарт», которая помогает в учете персональных расходов, операций с пластиковыми картами, при анализе операций по банковским счетам.

Подготовка документов. Для работы с большим количеством документов по многим отраслям хозяйственной деятельности шаблоны договоров и контрактов можно позаимствовать в системах «Бланк-договор-100» для Windows фирмы СВМ, «Договоры» для Windows фирмы «Тит-бит». Все эти системы позволяют вести базу по клиентам и осуществлять поиск по реквизитам документов.

Если потребности пользователя превышают возможности данных ПП, можно обратиться к системам «Деловые бумаги» (ВМИ) или «Формы правовых документов» («Гарант-сервис»). В этих системах содержится более 1500 типовых договоров, контрактов, заявлений, писем по 70 видам деятельности. Все эти системы снабжены универсальными средствами поиска и обработки документов.

Составление налоговой декларации. Программа «Советник», разработанная екатеринбургской компанией «Метатрон», обеспечивает пользователя информацией по всем вопросам налогообложения физических лиц, расчетов и минимизации налогов. Система обеспечивает правильность и своевременность оформления отчетных налоговых документов, конфиденциальность представленной информации, эффективное налоговое планирование.

Программа «Советник» выпускается в трех вариантах: облегченном, стандартном и профессиональном. Облегченный вариант предназначен для пользователей, которые имеют доходы от самостоятельной деятельности, состоят на государственной службе или имеют два места работы и более. Стандартный — для тех, кто имеет систематический доход от операций с активами, ведет частную практику, является частным предпринимателем. Профессиональный — для сотрудников предприятий, налоговых органов и адвокатов. От правовых баз данных «Советник» отличается тем, что не только содержит нормативные документы, а с помощью ЭС использует их для решения стоящих перед пользователем задач расчета, оформления и уменьшения налоговых платежей.

Составление бизнес-плана. В настоящее время наблюдается значительный рост интереса инвесторов, в том числе иностранных, к российским предприятиям. Первым шагом для начала диалога с инвестором является профессионально разработанный бизнес-план. Для его создания могут использоваться программы составления бизнес-плана «Project Expert» («Про-инвест Консалтинг») (версии «Standart», «Prof») и «Success-Успех» («Тор-консалтинг»).

Программа «Success-Успех» предназначена для оценки не только финансовой, но и коммерческой состоятельности инвестиционных проектов и новых планов, включая составление плана маркетинга, оперативных планов и других обязательных аналитических разделов для разработки рыночной стратегии, а также для оформления бизнес-плана в строгом соответствии с международными требованиями. Программа проста в освоении, действует в режиме диалога. «Success-Успех» включает в себя четыре готовых полных образца бизнес-планов на русском языке и один на английском. Консалтинговая часть программы содержит свыше 250 рекомендаций, примеров и образцов для профессионального составления бизнес-плана. Английский модуль предназначен для перевода бизнес-плана на английский язык, а входящий в программу расчетный модуль на основе Excel позволяет автоматизировать расчеты по инвестиционному проекту или при составлении финансового раздела бизнес-плана.

Разработка «Project Expert» является инструментом, позволяющим построить детальную финансовую модель действующего в условиях рынка предприятия независимо от его отраслевой принадлежности и формы собственности. Финансовая модель может быть использована для решения следующих задач:

- разработка детального финансового плана предприятия, расчет бюджета и определение потребности в финансировании;
- определение оптимальной схемы финансирования;
- разработка инвестиционного проекта и бизнес-плана развития как вновь создаваемого, так и действующего предприятия, полностью соответствующего международным стандартам;
- разработка финансовой части проекта эмиссии акций и оценка стоимости акции с учетом перспективы развития предприятия;
- контроль процесса реализации плана развития предприятия (только для версии «6.0 Prof»);
- работа с группами проектов (версия «6.0 Prof»).

Для малого и среднего бизнеса можно использовать облегченную версию «Project Expert Business Plan Guide 4.5» — длительность

проекта до 10 лет (в «Prof-версии» до 50), расчет до пяти наименований товаров и услуг (в «Prof-версии» до 10 000) или «Project Expert Biz Planner» — до 5 лет и до 5 услуг соответственно.

Кроме этих систем для оценки уже сделанных инвестиционных и инновационных проектов и анализа их риска выпущена программа того же разработчика «Questionnaire & Risk», а для прогнозирования в экономике и финансах — «Forecast Expert».

В модуле «Questionnaire and Risk» один или несколько независимых экспертов могут посредством многокритериального анализа провести начальную экспертизу нового проекта, рассчитать интегральный показатель уровня эффективности. При этом экспертный лист адаптируется пользователем в соответствии с целями и задачами инвестора. Предусмотрены средства, которые позволяют с помощью диалога проанализировать риск проекта, выделить факторы наибольшего риска и прокомментировать причины их возникновения и создать необходимый перечень факторов риска.

Таблица 4.12. Некоторые экономические характеристики основных программных продуктов

Наименование продукта	Цена*	Компания-производитель
1С:Бухгалтерия базовая 7.7	70\$	1С
1С:Бухгалтерия Проф 7.7	240\$	1С
Инфо-бухгалтер : минимальная 8.1	320 руб.	Инфо-бухгалтер
Инфо-бухгалтер профессиональная 8.1	4950 руб.	Инфо-бухгалтер
Турбо-бухгалтер 6.5 Лайт	78\$	ДИЦ
Турбо-бухгалтер Проф 6.5 Про	299\$	ДИЦ
Интотек-бухгалтер Профессионал 6.0	590\$	Интотек
Платежка+касса	360 руб.	Трек
Счет-фактура	360 руб.	Трек
1С:Платежные документы	19\$	1С
Wп-склад	165\$	Фолио
Фарм-склад	250\$	Фолио
Прод-склад	180\$	Фолио
БЭСТ-4	610\$	Интеллект-сервис
БЭСТ-КПМ	150\$	Интеллект-сервис
БЭСТ-Ф	150\$	Интеллект-сервис
Управление персоналом (серверный вариант на платформе MS SQL)	8000\$	АиТ

* Цены указаны на 30.04.2000 г.

Окончание табл. 4.12

Наименование продукта	Цена	Компания-производитель
Зарплата	400\$	АиТ
Кадры	400\$	АиТ
Учет коллективного заработка	400\$	АиТ
Передача данных в Пенсионный фонд	400\$	АиТ
Интегратор	580\$	Инфософт
Зарплата	280\$	Инфософт
Кадры	350\$	Инфософт
Складской учет	420\$	Инфософт
1С:Торговля	280\$	1С
Торговый склад	1250 руб.	Компьютер-сервис
Розничная торговля	1250 руб.	Компьютер-сервис
Промышленный склад	1250 руб.	Компьютер-сервис
Книжный склад	1250 руб.	Компьютер-сервис
1С:Деньги	18\$	1С
Декарт	40\$	Арсеналь
Бланк-договор-100	30\$	СВМ
Договоры	33\$	Тит-бит
Двловые бумаги	1560 руб.	ВМИ
Формы правовых документов	798 руб.	Гарант-сервис
Гарант-Максимум	36 800 руб.	Гарант-сервис
Консультант-плюс проф. расшир. верс. законод. России	15787 руб.	ВМИ
Референт-сервис стандарт	9408 руб.	Референт
Законодательство России	300 руб.	Элекс
Юридический советник	360 руб.	ЦКР
Юридический справочник для всех	42 руб.	ЦКР
Советник	115\$	Метатрон
Project Expert Standart	893\$	Про-инвест консалтинг
Project Expert Prof	1720\$	Про-инвест консалтинг
Success-Успех	3000 руб.	Топ-консалтинг
Forecast Expert	884\$	Про-инвест консалтинг
Audit Expert Stand	456\$	Про-инвест консалтинг
Audit Expert Prof	884\$	Про-инвест консалтинг

Для комплексного анализа финансового состояния и результатов деятельности предприятия предназначена программа «Audit Expert for Windows» («Standart», «Prof»), которая позволяет привести финансовую отчетность прошлых лет к единому международному стандарту, причем итоговая информация представлена в виде трех основных финансовых таблиц: отчет о прибылях, убытках и движении денежных средств. Пакет поможет провести анализ прошлой деятельности и текущего финансового состояния, рассчитать показатели ликвидности, провести горизонтальный анализ финансовых данных. В результате всех действий программа сделает предварительную переоценку баланса по данным детального описания структуры активов и пассивов предприятия. Данные из «Audit Expert for Windows» могут использоваться для автоматического описания стартового баланса компании в программе Project Expert 6.0 и для построения графиков и диаграмм.

Контрольные вопросы

1. Приведите классификацию АИСЗ, систем электронной коммерции и библиотечных.
2. Какова логическая и физическая структура БД ЮРИУС?
3. Сопоставьте структуру и состав словаря данных ЮРИУС (табл. 4.2) и STAIRS (гл. 2).
4. Охарактеризуйте функциональные возможности продукта Lotus Notes.
5. Какие виды информации поддерживаются в БД LN?
6. Перечислите и поясните основные функции и команды LN.
7. Охарактеризуйте структуру и динамику мирового рынка электронной коммерции.
8. Каковы функциональные возможности АРМ РТС?
9. Каковы типовая функциональная структура и состав базы данных АБИС?
10. Что такое пространственные данные в ГИС?
11. Охарактеризуйте различные типы цифровых моделей в ГИС.
12. Опишите процесс навигации в БД АБИС Tiplib при работе пользователя.
13. Кто такие ERP-системы? В чем их отличие от MRP и MRP-II систем?
14. Охарактеризуйте состав и структуру отечественного рынка программных продуктов.

Заключение

Информационные системы позволяют объективно оценить достигнутый уровень развития экономики, выявить резервы и обеспечить успех деятельности на основе применения правильных решений. Для успешной реализации проекта (ИС) прежде всего должен быть адекватно описан объект проектирования, построены полные и непротиворечивые функциональные и информационные модели ИС. Накопленный к настоящему времени опыт проектирования ИС показывает, что это логически сложная, трудоемкая и длительная по времени работа, требующая высокой квалификации участвующих в ней специалистов. В недавнем времени проектирование ИС выполнялось в основном на интуитивном уровне с применением неформализованных методов, основанных на искусстве, практическом опыте, экспертных оценках и дорогостоящих экспериментальных проверках качества функционирования ИС.

Перечисленные факторы способствовали появлению программно-технологических средств специального класса — CASE-средств, реализующих CASE-технологии создания и сопровождения ИС. CASE-технология представляет собой методологию проектирования ИС, а также набор инструментальных средств, позволяющих в наглядной форме моделировать предметную область, анализировать эту модель на всех этапах разработки и сопровождения ИС и создавать приложения в соответствии с информационными потребностями пользователей.

Полный жизненный цикл ИС включает в себя, как правило, стратегическое планирование, анализ, проектирование, реализацию, внедрение и эксплуатацию. Он должен поддерживаться «сквозной технологической цепочкой средств разработчика (средств ведения общей информации о проекте в целом, средств описания предметной области с возможностью трансформации в схемы баз данных и код приложений, средств сопровождения проекта группой разработчиков с контролем версий, встроенными средствами тестирования и т. д.).

В условиях рынка все большее число компаний осознают преимущества использования информационных систем. Во многих слу-

чаях ИС — это не только набор услуг, но и важнейший компонент бизнеса.

В учебном пособии кратко рассмотрены как типовые составные компоненты и структуры АИС, так и некоторые наборы отличий конкретных систем, определяемые факторами разнообразия областей применения АИС, — предметной областью БД, типом входной информации, преобразованиями данных, типом пользователей и пр. В рассмотренных примерах особенности конкретных приложений заключены, реализованы, как правило, *в логических и физических структурах БД, модульном составе программного обеспечения систем и пользовательских интерфейсах АИС*, что образует как бы «полную совокупность координат», в которых «закодированы» специфические системы.

Однако существуют определенные «степени свободы» (вариативности) перечисленных координат, в рамках которых возможны вариации областей применения тех или иных АИС. Программные средства, разработанные для решения определенного круга задач, имеют тенденцию к развитию «вширь» (при условии наличия дееспособного коллектива разработчиков и/или очевидной для всех продуктивности идей, заложенных в основу ПО), подобно тому, как может расширяться область применения перспективного математического метода.

Здесь возможны два взаимодополняемых подхода: разработка полифункциональных программных оболочек, охватывающих максимальный круг данного класса пользователей (в том числе разработчиков и администраторов систем), при этом настройка оболочки на приложение заключается в *непроцедурной* генерации (заполнение конфигурационных таблиц) либо использование систем программирования прикладных АИС с развитыми средствами управления данными (при этом разработка заключается в *полупроцедурном* процессе, близком к традиционному программированию).

Литература

1. Проектирование АИС / Н. М. Абдикеев, Н. З. Емельянова, Т. Л. Партыка, В. П. Романов; Под ред. К. И. Курбакова. М.: Изд-во Рос. экон. акад., 2004. 145 с.
2. *Василевич А., Зиновьев С., Молотков Л.* Онлайн-доступ к базам данных международной информационной сети STN International // Информационные ресурсы России. 1996. Вып. 4–5. С. 36–38.
3. Информационные ресурсы и документальные базы данных. Создание, использование, анализ: Учеб. пособие / Е. Н. Васина, О. Л. Голицына, Н. В. Максимов, И. И. Попов. М.: РГГУ, 1997. 178 с.
4. *Гайдамакин Н. А.* Автоматизированные информационные системы, базы и банки данных. Вводный курс: Учеб. пособие. М.: Гелиос, 2002. 368 с.
5. *Голицына О. Л., Максимов Н. В., Попов И. И.* Базы данных: Учеб. Пособие. М.: ФОРУМ: ИНФРА-М, 2003. 352 с.
6. *Голицына О. Л., Попов И. И.* Основы алгоритмизации и программирования: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2002. 432 с.
7. *Горностаев Ю. М., Поляк Ю. Е.* (ред). Русский Интернет. 7500 Webs. М.: ЦЭМИ РАН, МЦНТИ, 1998. 155 с.
8. ГОСТ 34.201—89. Виды, комплектность и обозначение документов при создании автоматизированных систем. М.: Изд-во стандартов, 1991. 43 с.
9. ГОСТ 34.601—90. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Стадии создания. М.: Изд-во стандартов, 1991. 37 с.
10. ГОСТ 34.602—89. Техническое задание на создание автоматизированной системы. М.: Изд-во стандартов, 1991. 25 с.
11. ГОСТ Р ИСО/МЭК 9126—93. Информационная технология. Оценка программной продукции. Характеристики качества и руководство по их применению. Государственный стандарт Российской Федерации. Издание официальное. М.: Госстандарт России, 1994. 12 с.
12. *Грейвс М.* Проектирование баз данных на основе XML. М.: Изд-во «Вильямс», 2002. 640 с.

13. *Зиндер Е. З.* Новое Системное Проектирование: Информационные Технологии и Бизнес-реинжиниринг. Часть 3 — методы Нового Системного Проектирования // Системы Управления Базами Данных, № 4, 1995 (часть 1); № 1, 1996 (часть 2) № 2/96. С. 61—76.
14. Геоинформатика / А. Д. Иванников, В. П. Кулагин, А. Н. Тихонов, В. Я. Цветков. М.: МАКС Пресс, 2001. 349 с.
15. Инструкция по работе с программой «Рабочая станция РТС» Версия 1.2 (RTS V1.2). М., 1997, 25 с.
16. Информационные системы в экономике: Учебник / Под ред. В. В. Дика. М.: Финансы и статистика, 1996. 272 с.
17. *Клещев Н. Т., Романов А. А.* Проектирование информационных систем: Учеб. пособие / Под общей ред. К. И. Курбакова. М.: Рос. экон. акад., 2000. 386 с.
18. *Коннолли Т., Бегг К., Страчан А.* Базы данных: проектирование, реализация и сопровождение: Теория и практика, 2-е изд.: Пер. с англ. Учеб. пособие. М.: Изд-во «Вильямс», 2000.
19. *Кренке Д.* Теория и практика построения баз данных. М.: Питер, 2003. 800 с.
20. Основы информатики (учебное пособие для абитуриентов экономических ВУЗов) / К. И. Курбаков, Т. Л. Партыка, И. И. Попов, В. П. Романов. М.: Экзамен, 2004. 320 с.
21. *Курбаков К. И.* Информационный ресурс и национальная система баз данных и баз знаний высшей школы России // Проблемы информатизации. 1993. № 3—4. С. 73—89.
22. *Либерти Д.* Разработка баз данных Microsoft SQL Server 2000 на примерах. М.: Изд-во «Вильямс», 2001. 464 с.
23. *Маклаков С. В.* ВРwin и ERwin. CASE-средства разработки информационных систем. М: Диалог МИФИ, 1999. 256 с.
24. *Максимов Н. В., Попов И. И.* Компьютерные сети: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2003. 336 с.
25. *Мартин Дж.* Организация баз данных в вычислительных системах. М.: Мир, 1980. 662 с.
26. *Мартин Дж.* Планирование развития автоматизированных систем. М.: Финансы и статистика, 1984.
27. *Меллинг В. П.* Корпоративные информационные архитектуры: и все-таки они меняются // СУБД. 1995. № 2.
28. *Партыка Т. Л., Попов И. И.* Информационная безопасность: Учеб. пособие. М.: ФОРУМ: ИНФРА-М, 2002. 368 с.
29. *Партыка Т. Л., Попов И. И.* Операционные системы, среды и оболочки. М.: ФОРУМ: ИНФРА-М, 2003. 400 с.

30. *Попов И. И.* Автоматизированные информационные системы (по областям применения): Учеб. пособие / Под общей редакцией К. И. Курбакова. М.: Изд-во Рос. экон. акад., 1999. 103 с.
31. *Попов И. И.* Информационные ресурсы и системы: реализация, моделирование, управление. М.: ТПК «Альянс», 1996. 408 с.
32. *Попов И. И.* Оценка и оптимизация информационных систем. М.: МИФИ, 1981. 86 с.
33. *Попов И. И.* Моделирование и оптимизация автоматизированных документальных информационных систем (учебное пособие) М.: РГГУ, 1996. 128 с.
34. *Попов И. И., Храмов П. Б. Максимов Н. В.* Введение в сетевые информационные ресурсы и технологии (учебное пособие). М.: РГГУ, 2001. 203 с.
35. *Попов И. И., Храмов П. Б.* Мировые информационные ресурсы и сети (методы доступа к ним): Учебник / Под ред. К. И. Курбакова. М.: Изд-во Рос. экон. акад., 1998. 145 с.
36. *Родионов И. И.* Интернет для российских предпринимателей. М.: МЦНТИ, 1997. 260 с.
37. Информационные системы: Учеб. пособие / А. Г. Романенко, Г. Ю. Максимович, О. Р. Самойлюк и др.; Под общей ред. К. И. Курбакова. М.: Изд-во Рос. экон. акад., 1999. 198 с.
38. *Смирнова Г. Н., Сорокин А. А., Тельнов Ю. Ф.* Проектирование экономических информационных систем. М.: Финансы и статистика, 2001. 512 с.
39. *Фокс Дж.* Программное обеспечение и его разработка. М.: МИР, 1985.
40. *Хомоненко А., Цыганков В., Мальцев М.* Базы данных: Учебник для высших учебных заведений. М.: КОРОНА принт, 2002. 672 с.
41. *Хоуг К.* MCSE/MCSD: SQL Server 7. Проектирование баз данных: Учеб. пособие. М.: Лори, 2000. 426 с.
42. *Храмов П. Б.* Лабиринт Интернет. Практическое руководство. М.: Электронинформ, 1996. 256 с.
43. *Шрайберг Я. Л., Воройский Ф. С.* Автоматизированные библиотечно-информационные системы России: состояние, выбор, внедрение, развитие. М.: Либерия, ГПНТБ РФ, 1996. 273 с.
44. *Popov I. I.* Description of internals of structure of the integrated Data Bank for full-text legal information // 2-nd East-West international on-line information meeting proceedings, Moscow, Russia, (30 sept. — 2 oct. 1992). Moscow. ICSTI. С. 147—152.

Глоссарий

CASE-средства (технологии) — программные средства, поддерживающие процессы создания и сопровождения ИС, включая анализ и формулировку требований, проектирование прикладного ПО (приложений) и баз данных, генерацию кода, тестирование, документирование, обеспечение качества, конфигурационное управление и управление проектом, а также другие процессы.

CRP (Capacity Requirements Planning) — корпоративные АИС, реализующие планирование производственных мощностей.

ERP (Enterprise Requirements Planning) — системы бизнес-планирования, или системы, реализованные на основе интеграции MRP-II с модулем финансового планирования FRP (Finance Requirements Planning). Они позволяют эффективно планировать всю коммерческую деятельность современного предприятия, в том числе финансовые затраты на проекты обновления оборудования и инвестиции в производство новой линейки изделий.

INIS (International Nuclear Information System) — структурное подразделение Международного агентства по атомной энергии (МАГАТЭ), секретариат которого базируется в Вене (Австрия).

MRP (Material Requirement Planning) — интегральные корпоративные информационные системы, реализующие методологию планирования потребности в материалах).

MRP-II (Manufactory Resource Planning) — системы замкнутого цикла планирования. Ориентированы на эффективное планирование всех ресурсов производственного предприятия, в том числе финансовых и кадровых.

STN International — научно-техническая информационная сеть, предоставляющая пользователю доступ к большинству мировых БД по науке и технике; содержит более 170 БД, суммарный объем — более 170 млн записей.

Автоматизированная информационно-поисковая система (АИПС) — совокупное название как для программных оболочек, ориенти-

рованных на ввод, хранение, поиск и выходное представление документов (структур данных сложного или неопределенного формата), так и для конкретных систем определенного наполнения и предметной ориентации, реализованных на основе таких оболочек (или иными программными методами). Примерами программных оболочек АИПС являются STAIRS, DPS, ISIS, IRBIS.

Агрегат данных — именованная совокупность элементов данных, представленных простой (векторной) или иерархической (группы или повторяющиеся группы) структурой. Примеры — массивы, записи, комплексные числа и пр.

Агрегатные функции — вычислительные функции СУБД, реализующие при выводе отчета или чтении таблицы данных вычисление агрегатных данных по группе записей БД (строк таблицы) — MAX (максимальное значение данного поля в обработанной группе записей/строк), MIN (минимальное значение), AVER (среднее значение), COUNT (подсчет числа записей) и пр.

Администратор базы данных (АБД) — лицо или группа, уполномоченные для ведения БД (модификация структуры и содержания БД, активизация доступа пользователей, выполнение других административных функций, которые затрагивают всех пользователей). С этой целью он/она идентифицирует объекты БД и моделирует базу, используя язык описания данных. Администратор решает все вопросы, связанные с размещением БД в памяти, выбором стратегии доступа и ограничением доступа к данным. В функции АБД входят также организация загрузки, ведения и восстановления БД.

Адрес данных — идентификация места расположения данных.

Адресация — способ размещения и выборки данных в памяти.

Атрибут — поле данных, содержащее информацию об объекте.

База данных (БД) — именованная совокупность взаимосвязанных данных, отображающая состояние объектов и их отношений в некоторой предметной области, используемых несколькими пользователями и хранящимися с минимальной избыточностью.

Банк данных — определяется как система программных, языковых, организационных и технических средств, предназначенных для централизованного накопления и коллективного использования

данных. Компонентами банка данных являются базы данных, а также система управления базами данных (СУБД).

Безопасность — защита данных от преднамеренного или непреднамеренного доступа, модификации или разрушения.

Библиографические данные (записи) — выходные данные (включают авторов, заголовков, классификационный индекс, место публикации и пр.), иногда реферат.

Брат — в иерархической структуре данных узел, имеющий того же родителя, что и другой узел.

Ветвь — в иерархической модели данных узел дерева вместе со всеми его отпрысками, отдаленными потомками и родительскими источниками.

Внешняя схема — представление данных с точки зрения пользователя или прикладной программы.

Внутренняя схема — физическая структура данных.

Восстанавливаемость — запроектированная возможность восстановления целостности БД после любого сбоя системы.

Документ — агрегат данных в документальных системах (АИПС), имеющий иерархическую структуру и, кроме форматных полей (элементы или агрегаты данных фиксированной длины), обычно содержащий текстовые поля или символьные последовательности неопределенной длины, логически подразделяющиеся на параграфы (PAR, SEGM), предложения (SENT), слова (WORD).

Запись логическая — идентифицируемая (именованная) совокупность элементов или агрегатов данных, воспринимаемая прикладной программой как единое целое при обмене информацией с внешней памятью. Запись — это упорядоченная в соответствии с характером взаимосвязей совокупность *полей* (элементов) данных, размещаемых в памяти в соответствии с их *типом*.

Запись физическая — совокупность данных, которая может быть считана или записана как единое целое одной командой ввода-вывода.

Иерархическая модель данных — использует представление предметной области БД в форме иерархического дерева, узлы которого

связаны по вертикали отношением «предок—потомок». Навигация в БД представляет собой перемещение по вертикали и горизонтали в данной структуре. Одной из наиболее популярных иерархических СУБД была Information Management System (IMS) компании IBM, появившаяся в 1968 г.

Импорт (загрузка, download) — утилита (функция, команда) СУБД, служащая для чтения файлов операционной системы, которые содержат данные из базы данных, представленные в некотором коммуникативном формате.

Инвертированный файл (список) — файл, предназначенный для быстрого произвольного поиска записей по значениям ключей, организованный в виде независимых упорядоченных списков (индексов) ключей — значений определенных полей записей основного файла.

Индекс — таблица, используемая для определения адреса записи.

Информационная база — данные, отражающие состояние определенной предметной области и используемые информационной системой. Информационная база состоит из двух компонент: коллекций записей *собственно данных* и описаний этих данных — *метаданных*. Данные отделены от описаний, но в то же время данные не могут использоваться без обращения к соответствующим описаниям.

Клиент — программы, написанные как пользователями, так и поставщиками СУБД, внешние или «встроенные» по отношению к СУБД. Программа-клиент организована в виде приложения, работающего «поверх» СУБД и обращающегося для выполнения операций с данными к компонентам СУБД через интерфейс внешнего уровня.

Ключ — значение (элемент данных), используемое для идентификации или определения адреса записи.

Коммуникативные (обменные) форматы данных — соглашения о представлении агрегатов информации при передаче.

Коммуникативный формат — способ представления данных в файле операционной системы, обеспечивающий безошибочное распознавание единиц информации (записей, агрегатов, полей) при импорте (загрузке) информации файла ОС в базу данных (обычно в одну из таблиц БД). Файлы в коммуникативных форматах

(КФ) предназначены для архивного хранения и транспортировки содержимого БД. Примерами КФ могут служить основные методы, используемые в системах семейства FoxPro: (1) *SDF (системный формат данных)*, когда каждая строка таблицы БД выводится в виде записи текстового файла, где каждый элемент данных (поле) имеет длину и тип, соответствующие описанию в БД; (2) *Метод с разделителями*, где каждый элемент данных отделяется от соседнего символом-разделителем («терминатором»). В системах, ориентированных на обработку документов (АИПС), используются более сложные форматы, описанные в стандарте ISO 2709, примером которых может служить МЕКОФ (международный коммуникативный формат), использующий рекурсивное описание агрегатов и элементов данных.

Концепция баз данных — информационная технология интегрированного хранения и обработки данных, в основе которой лежит механизм передачи обрабатывающей программе из всех хранимых данных только необходимых и в требуемой форме.

Лист — для иерархической модели данных узел, у которого нет отпрысков.

Логическая структура БД — определение БД на физически независимом уровне.

Логический файл — файл в представлении прикладной задачи, состоящий из логических записей, структура которых может отличаться от структуры физических записей, представляющих информацию в памяти.

Маркетинговые исследования — в англо-американской литературе различают два понятия. Market research — сбор и анализ информации о конкретном рынке и marketing research — исследование всех элементов комплекса маркетинга (товарной политики, ценовой политики, системы продвижения товара, системы управления персоналом, коммуникационной политики). В российской литературе по маркетингу и рекламе термин употребляют и в том, и в другом значении, проводя различия в основном в зависимости от применяемых методов исследования: количественных и качественных.

Машина баз данных (МБД, DBM — data base machine) — специализированное электронное устройство (интеллектуальный контрол-

лер дискового накопителя), реализующее аппаратным способом функции СУБД.

Метод доступа — метод (способ) организации обмена данными обычно между оперативной и внешней памятью, поддерживаемый ОС, например, прямой, последовательный, индексный метод доступа.

Моделирования парадигма — условности, определяющие способ представления взаимосвязи объектов на уровне *структур данных*. С этой точки зрения различаются реляционные, сетевые, иерархические, объектные, объектно-реляционные, документальные и другие виды моделей.

Модель данных — базовый инструментарий, обеспечивающий на формальном абстрактном уровне конкретные способы представления объектов и связей.

Модель инфологическая (концептуальная) — описание предметной области, выполненное с использованием естественного языка, математических выражений, таблиц, графов и других средств, понятных всем людям, работающих над проектированием базы данных.

Модель физическая — определяющая размещение и способы поиска данных на внешних запоминающих устройствах СУБД.

Независимость данных логическая (физическая) — свойство системы, обеспечивающее возможность изменять логическую (физическую) структуры данных без изменения физической (логической).

Нормализация — представление сложных структур данных (документов) в виде двумерных таблиц (отношений).

Обход дерева — процесс обследования по очереди каждого узла дерева в иерархической модели данных.

Открытость — свойство информационных технологий и систем, предполагающее способность объединять разные информационные системы, аппаратуру и программные продукты различных производителей, что делает возможным обмен между ними данными, распределенный доступ к информационным ресурсам.

Отношение (relation) — агрегат данных, хранящийся в одной из таблиц (строка таблицы) табличной, реляционной БД или созда-

ваемый виртуально в процессе выполнения операций над базой данных при выполнении запросов.

Отпрыск — для иерархической структуры данных узел, который «принадлежит» другому узлу. Узел, которому принадлежит узел-отпрыск, называется *родительским узлом*.

Полнотекстовые документы (записи) — полный (или почти) исходный текст журнальной статьи или другого документа.

Пользователь БД — программа или человек, обращающийся к базе данных с помощью средств управления данными СУБД.

Постреляционная СУБД (ПРСУБД) — система управления базами данных, поддерживающая расширенную реляционную модель данных. Расширения могут носить различный характер, в частности, ADABAS поддерживает повторяющиеся групповые поля (фактически вложенные таблицы) в экземплярах отношений.

Предметная область (ПрО) — набор объектов, представляющих интерес для актуальных или предполагаемых пользователей, когда реальный мир отображается совокупностью конкретных и абстрактных понятий, между которыми фиксируются определенные связи.

Проектирование базы данных — упорядоченный формализованный процесс создания системы взаимосвязанных описаний таких моделей предметной области, которые связывают (фиксируют) хранимые в базе данные с объектами предметной области, ими описываемые.

Распределенная база данных — совокупность баз данных, которые обрабатываются и управляются по отдельности, а также могут разделять информацию.

Реинжиниринг бизнес-процессов — концепция проектирования АИС, которая предполагает системный анализ деятельности предприятия и органическое встраивание в бизнес-процессы новых информационных технологий.

Релевантность — свойство некоторой информации (документ, факт и пр.) удовлетворять информационную потребность пользователя АИС (relevant — относящийся к делу).

Реляционная алгебра — алгебра (язык), включающая набор операций для манипулирования отношениями.

- Реляционная база данных** — база данных, состоящая из отношений. Здесь вся информация, доступная пользователю, организована в виде таблиц, которые имеют уникальные имена и состоят из строк и столбцов, на пересечении которых содержатся элементы данных. Операции над данными сводятся к операциям над этими таблицами.
- Реляционная СУБД (РСУБД)** — система управления базами данных, поддерживающая реляционную модель данных (РМД).
- Реляционный процессор (РП)** — компонента машины баз данных, реализующая основные операции над отношениями (реляционную алгебру), а также ряд агрегатных функций — max, min, aver, count и пр.
- Рунет** — сокращенное название для «русского», т. е. русскоязычного и российско-ориентированного Интернета. Например, робот-индексировщик Yandex автоматически считает серверы в доменах su, ru, am, az, by, ge, kg, kz, md, ua, uz принадлежащими Рунету. Остальные серверы вносятся в базу, если на них найден текст на русском языке или владельцы ресурсов убедят администрацию поисковой машины в том, что их сервер интересен пользователям русскоязычного Интернета.
- Сервер** — программа, реализующая функции СУБД: определение данных, запись-чтение-удаление данных, поддержку схем внешнего-концептуального-внутреннего уровней, диспетчирование и оптимизацию выполнения запросов, защиту данных.
- Сетевая модель данных (модель CODASYL)** — предложенная CODASYL модификация иерархической модели, в которой одна запись может участвовать в нескольких отношениях предок/потомок.
- Словарь данных** — исчерпывающий набор таблиц или файлов, представляющий собой каталог всех описаний данных (имен, типов). Может содержать также информацию о пользователях, привилегиях и т. д., доступную только администратору базы данных. Является центральным источником информации для СУБД, АБД и всех пользователей.
- Согласованность** — свойство БД по отношению к некоторой совокупности пользователей, если в любой момент времени БД реа-

гирует на их запросы одинаковым образом (т. е. все пользователи получают одинаковый ответ).

Справочники (указатели) — списки (перечни, словари, описания продуктов, веществ, проектов, организаций, грантов, персон и другая систематизированная информация об объектах).

Структура данных — атрибутивная форма представления свойств и связей предметной области, ориентированная на выражение описания данных средствами формальных языков, и таким образом учитывающая возможности и ограничения с целью сведения описаний к стандартным типам и регулярным связям. Структура данных с точки зрения программирования — это способ отображения значений в памяти — размер области и порядок ее выделения (который и определит характер процедуры адресации-выборки).

Структура данных линейная — порядок следования элементов данных, который имеет линейный характер и соответствует порядку расположения элементов в памяти.

Структура информации — схематичная форма представления сложных композиционных объектов и связей реальной предметной области, выделяемых как актуально необходимые для решения прикладных задач.

Структуры данных — способы композиции простых данных в агрегаты и операции над ними. Исторически первыми структурами данных были массивы, комплексные числа, записи файлов системы КОБОЛ (т. е. агрегаты разнородных элементарных данных).

СУБД — система управления базами данных — совокупность языковых и программных средств, предназначенных для создания, введения и совместного использования БД многими пользователями. Систему управления базой данных можно также определить как специальный пакет программ, посредством которого реализуется централизованное управление базой данных и обеспечивается доступ к данным.

Таблица — основная единица информации в системе управления реляционной базой данных. Состоит из одной или более единиц информации (строк), каждая из которых содержит значения некоторого вида (столбцы).

Тип данных — характер данных, определяющий способ представления значения в памяти и соответственно множество стандартных операций (примитивов) преобразования этого значения.

Типы данных — совокупность соглашений о программно-аппаратной форме представления и обработки, а также ввода, контроля и вывода элементарных данных. К типам данных прежде всего относятся классические типы — целое, действительное числа, булевское значение.

Топология БД — схема распределения компонент базы данных по физическим носителям, в том числе различным узлам вычислительной сети.

Точка сохранения — момент времени, когда в БД записывается вся работа в транзакции. В транзакции может применяться ряд точек сохранения, выступающих в роли промежуточных точек для работы.

Транзакция — последовательность операций над данными базы, переводящая БД из одного непротиворечивого состояния в другое, которое может быть представлено как одно «событие».

Уникальные документы (страницы, URL) — в Интернет часто одни и те же документы лежат на разных серверах или на одном сервере, но в разных кодировках. Yandex умеет определять совпадения документов и индексирует только один документ (URL) из группы одинаковых.

Уровни представления данных — концептуальный, внутренний и внешний. *Внутренний уровень* — глобальное представление БД, определяет необходимые условия в первую очередь для организации хранения данных на внешних запоминающих устройствах. *Представление на концептуальном уровне* представляет собой обобщенный взгляд на данные с позиций предметной области. *Внешний уровень* — представляет потребности пользователей и прикладных программ.

Утилита СУБД — программа, которая запускается в работу командой операционной системы главного компьютера и выполняет какую-то функцию над базой данных (обычно на физическом уровне данных), либо команда (функция ядра СУБД, доступная только АБД), реализующая аналогичную операцию.

Файл — именуемая единица информации, поддерживаемая операционной системой. Доступ к данным реализуется либо в рамках ОС, либо пользовательскими программами, либо в рамках СУБД, либо комбинированно.

Файл ASCII (ASCII-File) — файл, содержащий символьную информацию, представленную только ASCII-кодами «левой части» (первые 128 символов кодовой таблицы или код Latin-1) и символьную разметку.

Файл базы данных — физический файл ОС, используемый для размещения БД. Управление данными в таком файле производится совместно ОС и СУБД. Крайние варианты размещения БД по файлам — (1) все данные БД — в одном файле (файл DATA, СУБД ADABAS); (2) каждая таблица БД — в отдельном файле ОС (DBF-файлы, системы FoxPro). Промежуточный вариант размещения, например, ORACLE — база данных состоит из одного или более табличных пространств, которые в свою очередь состоят из одного или более файлов базы данных.

Файл бинарный (Binary File) — файл, содержащий произвольную двоичную информацию (текст с бинарной разметкой, программа, графика, архивный файл).

Файл графический (Image file) — бинарный файл, содержащий данные, обычно полученные с помощью растрового сканера и соответствующие двумерному изображению объекта.

Файл текстовый (Text file) — файл, содержащий символьную информацию в одном из соответствующих кодов и коды, управляющие режимом отображения символов на печать и экранные устройства.

Фонд данных — общая совокупность хранящихся в банке данных, состоящая из БД и архивов. Основанием для этого разбиения является режим использования данных БД — только часть фонда данных, находящаяся под оперативным управлением СУБД и размещенная, как правило, на магнитных дисках. Все остальные данные — архивы, обычно располагаются на магнитных лентах. Одни и те же данные в разные моменты времени могут входить как в БД, так и в архивы.

Форматы файлов — представление информации на уровне взаимодействия операционной системы с прикладными программами.

Целостность — свойство БД, при котором она удовлетворяет некоторым определенным ограничениям значений данных и сохраняет это свойство при всех модификациях (замена, добавление или удаление) данных.

Централизованное управление данными — осуществляется средствами, входящими в состав СУБД, обеспечивает: сокращение избыточности в хранимых данных; устранение несовместимости приложений; совместное использование хранимых данных, что достигается необходимой интеграцией данных; целостность, которая обеспечивается с помощью процедур, предотвращающих внесение в БД неверных данных и ее восстановление после отказов системы; лучший учет противоречивых требований к использованию БД в различных приложениях путем создания соответствующих структурированных БД.

Численные данные — записи, которые содержат таблицы статистических, финансовых и других сведений.

Экспорт (выгрузка, upload) — утилита (функция, команда) СУБД, служащая для вывода информации из БД (обычно одной из таблиц) в файл(ы) операционной системы, организованные в некотором коммуникативном формате.

Электронная коммерция — форма поставки продукции, при которой выбор и заказ товаров осуществляется через компьютерные сети, а расчеты между покупателем и поставщиком — с использованием электронных документов и/или средств платежа. При этом в качестве покупателей товаров (или услуг) могут выступать как частные лица, так и организации.

Элемент данных (элементарное данное) — неделимое именованное данное, характеризующееся типом (напр., символьный, числовой, логический, и пр.), длиной (в байтах) и обычно рассчитанное на размещение в одном машинном слове соответствующей разрядности. Это минимальная адресуемая (идентифицируемая) часть памяти, на которую можно ссылаться при обращении к данным. Ранние языки программирования (Алгол, Фортран) были рассчитаны на обработку элементарных данных или их простейших агрегатов — массивов (матрицы, векторы). С возникновением ЯП Кобол появляется возможность представления и обработки агрегатов разнотипных данных (записей). В реляционных БД элементарное данное есть элемент таблицы. Иногда используется термин *поле записи* в качестве синонима.

Язык манипулирования данными (ЯМД) — обычно включает в себя средства запросов и поддержания базы данных (добавление, удаление, обновление данных, создание и уничтожение БД, изменение определений БД, обеспечение запросов к справочнику БД).

Язык описания данных (ЯОД) — средство внутрисистемного определения данных, представляющего обобщение внешних взглядов. Описание представляет собой модель данных и их отношений, т. е. структур, из которых образуется БД.

Язык структурированных запросов (SQL) — основной интерфейс пользователя и АБД для запоминания и поиска информации в базе данных для ряда СУБД (Oracle, MS SQL Server и пр.). Он включает в себя в качестве подмножеств следующие категории операторов: язык описания данных (ЯОД), язык управления данными (ЯУД), язык манипулирования данными (ЯМД).

Обзор основных команд системы программирования FoxPro

APPEND [BLANK]	Добавляет новые записи в конец активного файла базы данных. Это основная команда для полноэкранного ввода данных
APPEND FROM	Копирует записи в конец активного файла базы данных из другого файла
BROWSE	Используется для открытия окна и высвечивания записей из файла базы данных. Имеется возможность редактирования и добавления записей
BROWSE FOR	В окне Browse высвечиваются только те записи, которые удовлетворяют условию. Для лучшей работы в FOR используется оптимизируемое выражение
CLOSE	Используется для закрытия файлов журнала, файлов базы данных, файлов формата экрана, индексов и файлов процедур
CREATE	Создание нового файла БД
CDUNT	Выполняет подсчет числа записей активной БД, удовлетворяющих заданным условиям
DELETE	Помечает для исключения записи активного файла базы данных
DELETE FILE	Удаляет файл
DO	Выполняет программный или процедурный файл
EXPORT	Копирует данные из базы данных FoxPro в файл с другим форматом
GO	Позиционирование указателя записи
GO [RECORD]	Устанавливает указатель записи на запись GO TOP BOTTOM Установка указателя записи на первую и последнюю в выбранной базе данных. Если база используется с индексом, то первой будет запись с минимальным значением ключа, последней – запись с максимальным значением ключа
HELP	Открытие окна помощи
INDEX	Создает индексный файл для упорядочения БД
JOIN	Объединяет два файла БД
LIST	Отображение записей БД
LOCATE	Выполняет в активном файле БД поиск по заданному условию
MODIFI COMMAND MODIFI FILE	Открывает окно редактирования
MODIFY STRUCTURE	Модифицирует структуру файла базы данных

PACK	Физически удаляет записи, помеченные для удаления в базе данных, уменьшает размер текстового файла
QUIT	Закрывает все открытые файлы, завершает сеанс работы системы FoxPro и возвращает управление операционной системе
REINDEX	Перестраивает все активные индексы
RENAME	Выполняет изменение имени файла на диске
RECALL	Снимает маркировку на удаление записей активного файла базы данных, установленную командой DELETE
REPLACE	Обновляет значения заданных полей записей активного файла базы данных
SKIP	Перемещает указатель записи в активном файле базы данных вперед или назад
SELECT – SQL	Получение данных из одной или нескольких таблиц в массив или курсор
SEEK	Выполняет поиск первой записи индексированного файла базы данных с ключом, совпадающим с заданным выражением
SET ALTERNATE	Направление вывода, ведущегося на экран или в окно и в текстовый файл
SET DEFAULT TO	Для смены директории, используемой по умолчанию
SET RELATION	Связывает два открытых файла базы данных по вычисленному значению <выражения> через ключевое поле или по номеру записи
SET INDEX	Открывает заданные индексы текущего файла базы данных
SET INDEX TO	Закрывает все открытые индексы в текущей рабочей области
SET FILTER	Позволяет выводить и обрабатывать только те записи файла БД, которые удовлетворяют заданному условию
USE	Открывает БД в указанной рабочей области

Методические указания по выполнению лабораторной работы «Исследование физической структуры и навигации в БД на примере АИСЗ ЮРИУС»

Часть 1. Взаимосвязь файлов и алгоритмы поиска информации в IX-интерфейсе

Порядок выполнения работы:

1. Установить текущей директорию RZ;
2. Вызвать СУБД FoxPro:



3. Задать имя файла для вывода протокола диалога и установить режим альтернативного вывода.

4. Убрать меню командой

```
SET MENU OFF.
```

5. Ознакомиться со структурой, индексами и форматами вывода файла регистрационных карт:

- отобразить структуру файла;
- отобразить запись файла номер N с помощью команды EDIT (здесь и далее N — порядковый номер варианта задания);
- отобразить эту же запись с использованием файла формата экрана:

```
SET FORM TO SZREG  
EDIT;
```

- вывести три записи начиная с номера N , используя для вывода записей файла форму отчета SZR.FRM.

6. Ознакомиться со структурой, индексом и форматом вывода файла полных текстов:

- отобразить структуру файла и запись файла номер N ;
- отобразить три записи файла, упорядоченные по индексу PAGED.IDX (использовать для вывода записей файла форму отчета SZDOC.FRM).

7. Ознакомиться со структурой и индексом файла частотных словарей:

- отобразить структуру файла и запись номер N ;
- отобразить 10 записей файла, упорядоченных по индексу FRV.IDX.

8. Ознакомиться со структурой и индексом инверсного файла:

- отобразить структуру файла;
- отобразить запись файла номер N ;
- отобразить 10 записей файла, упорядоченных по индексу IND.IDX.

9. Подготовить файлы БД к выполнению поисковых операций:

- в рабочей области номер 1 открыть файл регистрационных карт и установить индекс T1P1;
- в рабочей области номер 2 открыть файл текстов и установить индекс PAGED;
- в рабочей области номер 3 открыть инверсный файл и установить индекс IND;
- в рабочей области номер 4 открыть файл частотных словарей и установить индекс FRV;
- с помощью команды SET EXACT OFF установить принцип сравнения символьных строк, при котором строки считаются равными, если одна из них является подстрокой другой;

10. Выполнить поиск регистрационных карт, удовлетворяющих условию, указанному в варианте задания (табл. ПЗ.1), по следующему алгоритму:

- в файле частотных словарей найти и отобразить запись со значением, указанным в варианте задания;
- в инверсном файле найти и отобразить первую запись, у которой в поле ADRES стоит значение, равное RECNO(4) (т. е. значение, равное физическому номеру текущей записи файла, открытого в рабочей области 4, или файла частотных словарей);
- найти в файле регистрационных карт запись, у которой значения полей T1+P1 равно значению поля PAGEI текущей записи

Таблица П3.1. Варианты задания к поиску по регистрационным картам

№ п/п	Поле	Значение	№ п/п	Поле	Значение
1	1	АКЦИОНЕРНОЕ	16	1	РОССИИ
2	1	БЮРО	17	1	РСФСР
3	1	ВОЕННОСЛУЖАЩИХ	18	1	СОЦИАЛЬНОЙ
4	1	ВЬЕТНАМ	19	1	ТРАНСПОРТА
5	1	КАТАСТРОФЫ	20	1	ФОНДА
6	1	МОСКВА	21	1	ЧЕРНОБЫЛЬСКОЙ
7	1	НАЛОГОВОЙ	22	1	ЧРЕЗВЫЧАЙНЫХ
8	1	НАУКИ	23	1	ЭМИССИИ
9	1	ОБОРОНЫ	24	1	РЕСПУБЛИКИ
10	1	ОБЩЕСТВО	25	2	93.07.07
11	1	ОРУЖИЯ	26	2	93.07.09
12	1	ПРАВО	27	2	93.07.15
13	1	ПРЕЗИДЕНТА	28	4	УКАЗ
14	1	ПРИВАТИЗАЦИИ	29	5	ПРЕЗИДЕНТ РФ
15	1	РЕФОРМЫ	30	5	ЗАМПРЕД СМ РФ

инверсного файла, и вывести найденную запись, используя файл формата отчета SZR.FRM;

- вывести все записи, удовлетворяющие условию, используя связь файлов БД:

- установить связь между инверсным файлом и файлом регистрационных карт:

```
SET RELATION TO SF1->T1+SF1->P1=PAGFI INTO SF1;
```

- вывести регистрационные карты, используя файл формата отчета SZR.FRM:

```
REPORT WHILE ADRES=RECNO(4) FORM SZR;
```

- вывести текст документа, соответствующий Регистрационной карте номер *N* (*N* — номер варианта задания):

- найти регистрационную карту номер *N*;
- перейти в рабочую область номер 2 — файл текстов;
- выполнить команду:

```
SEEK RTRIM(SF1->T1+SF1->P1)
```

(функция RTRIM убирает пробелы в конце текстовой строки);

- вывести текст документа в формате SZDOC:

```
REPORT WHILE PAGED=STRIM(SF1->T1+SF1->P1) FORM SZDOC;
```

- провести контекстный поиск, т. е. выделить из текста страницы документа некоторый контекст (слово или его часть), например [ДЕПУТАТ], и выполнить операцию

```
REPORT WHILE PAGED=STRIM(SF1->T1+SF1->P1) FOR [ДЕПУТАТ] $
TEXTD FORM SZDOC;
```

- выполнить аналогичный поиск под интерфейсом IX:

```
>do IXRZ$;
```

- удостовериться в совпадении результатов.

Отчет о лабораторной работе должен содержать:

- распечатку протокола работы с краткими комментариями к выполненным действиям;
- описание алгоритма поиска регистрационных карт по значению из файла частотных словарей;
- схему взаимосвязи данных файлов БД.

Часть 2. Физическая структура и навигация в БД DC-интерфейса

Порядок выполнения лабораторной работы:

- по аналогии с частью 1 просмотреть и распечатать новые с точки зрения навигации файлы БД (DCIND.DBF, DCFRV.DBF);
- активизировать файлы DCFRV, DCIND, SZDOC с соответствующими индексами и провести поиск и распечатку первой страницы полного текста. Варианты поисковых терминов приведены в табл. ПЗ.2, примеры распечаток, соответствующих навигации в полнотекстовой БД, — в прил. 3;
- выполнить аналогичный поиск под интерфейсом DC:

```
> do DCRZ
```

- удостовериться в совпадении результатов.

Отчет о лабораторной работе должен содержать:

- распечатку протокола работы с краткими комментариями к выполненным действиям;
- описание алгоритма поиска регистрационных карт по значению из файла частотных словарей;
- схему взаимосвязи данных файлов БД.

Таблица П3.2. Варианты задания к поиску по полиым текстам

№ п/п	Термин п./т	№ п/п	Термин п./т
1	АКТОВ	32	КУЛЬТУРЫ
2	АКЦИЙ	33	ЛИНИЙ
3	АКЦИОНЕРНОГО	34	ЛИЦ
4	АППАРАТА	35	МАТЕРИАЛОВ
5	БЮДЖЕТА	36	МЕЖДУНАРОДНЫХ
6	ВЕДОМСТВА	37	МЕРАХ
7	ВЕРХОВНЫЙ	38	МЕРОПРИЯТИЯ
8	ВЛАСТЯМИ	39	МЕСТ
9	ВНУТРЕННИХ	40	МЕСЯЦЕВ
10	ВОЗДУШНЫХ	41	МИНИСТЕРСТВА
11	ВОПРОСОВ	42	НАУКИ
12	ВРЕМЯ	43	НАЧАЛЬНИК
13	ВЫПОЛНЕНИЯ	44	НЕОБХОДИМОСТИ
14	ГЕНЕРАЛЬНОГО	45	НЕОБХОДИМЫЕ
15	ОБЛАСТИ	46	ГОД
16	ГОРОДОВ	47	ОБМЕНА
17	ГОСКОМИМУЩЕСТВУ	48	ОБОРОНЫ
18	ГОСКОМПРОМ	49	ОБОРУДОВАНИЯ
19	ГОСУДАРСТВ	50	ОБРАЗОВАНИЯ
20	ГРАЖДАН	51	ОБЩЕСТВА
21	ГРАНИЦ	52	ОТРАСЛЕЙ
22	ГРУЗА	53	ОХРАНЫ
23	ЗАКОНА	54	ПАССАЖИРОВ
24	ЗАЩИТЫ	55	ПЕРЕГОВОРЫ
25	ИЗМЕНЕНИЙ	56	ПЕРЕЧЕНЬ
26	ИНОСТРАННЫХ	57	ПОДДЕРЖКИ
27	ИНФОРМАЦИИ	58	ПОДПИСАНИЯ
28	ИСПОЛНИТЕЛЬНОЙ	59	ПОЛИТИКИ
29	КОЛЛЕКТИВА	60	ПОЛОЖЕНИЕ
30	КОМИССИИ	61	ПОМОЩИ
31	КОМИТЕТА	62	ПРИВАТИЗАЦИИ

**РОССИЙСКАЯ ЭКОНОМИЧЕСКАЯ АКАДЕМИЯ
ИМ. Г. В. ПЛЕХАНОВА**

ИНСТИТУТ ИНФОРМАТИКИ

**Отчет о лабораторной работе № 1
Исследование физической структуры и навигации в БД
на примере АИСЗ ЮРИУС**

по курсу

АВТОМАТИЗИРОВАННЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Выполнил ст. _____

Проверил преподаватель _____

**Москва
2004**

Вариант 8, поисковое слово — НАУКИ

Часть 1. Протокол навигации в БД ЮРИУС (RZ) при поиске по регистрационным картам (PK)

```

SELE 1
USE SF1 INDE TIP1           (открытие файла PK)
SELE 2
USE IXFRV INDE FRV         (открытие частотного словаря)
SELE 3
USE IXIND INDE IND         (открытие инверсного списка)
SELE 4
USE SZDOC INDE PAGED       (открытие полных текстов)
SELE 2
SEEK [101НАУК]             (поиск по основе слова, находящегося в заголовке PK)
LIST NEXT 10

```

Record# (код слова)	FNMBF	FLDNF	VALUF	FREQF
1082	1	01	НАУК	1
538	1	01	НАУКИ	3
100	1	01	НАУЧНО-ИССЛЕДОВАТЕЛЬ	4
804	1	01	НАУЧНО-ПРОИЗВОДСТВЕН	1
43	1	01	НАУЧНО-ТЕХНИЧЕСКОГО	1
284	1	01	НАУЧНО-ТЕХНИЧЕСКОМУ	4
827	1	01	НАХОДКА	1
447	1	01	НАЧАЛЬНИКЕ	1
881	1	01	НАЧАЛЬСТВУЮЩЕГО	1

```

SELE 3
SEEK 538                   (поиск по коду слова)
LIST NEXT 10

```

Record#	ADRES	PAGFI	
972	538	RZ9659	(слово НАУКИ=538 в заголовке содержат 3 PK)
987	538	RZ9660	
1296	538	RZ9677	
973	539	RZ9659	
978	540	RZ9659	
982	541	RZ9660	
983	542	RZ9660	
1767	542	RZ9708	
984	543	RZ9660	
985	544	RZ9660	

SELE 1
 SEEK [RZ9659] (поиск первой из этих РК)
 REPO NEXT 1 FORM SF1 (печать найденной РК)

PAGE NO. 1
 03/30/04

Распечатка списка регистрационных карт
 базы данных *Новейшее законодательство России*

Заглавие документа в БД	Дата и кем принят	Номер	Вид	Стр. акта
ОБ УТВЕРЖДЕНИИ ПОЛОЖЕНИЯ О МИНИСТЕРСТВЕ НАУКИ И ТЕХНИЧЕСКОЙ ПОЛИТИКИ РОССИЙСКОЙ ФЕДЕРАЦИИ	93.07.12	650	ПОСТАНОВЛЕНИЕ ПРАВИТЕЛЬСТВА РФ	9659

SELE 4
 SEEK [RZ9659] (поиск начала документа для этой РК)
 REPO FORM SZDOC WHILE PAGED=[RZ9659] (выдача этого документа)

PAGE NO. 1
 03/30/04

Распечатка текста БД
 Новейшее законодательство России

** Том, документ, стр: RZ9659-001

СОВЕТ МИНИСТРОВ — ПРАВИТЕЛЬСТВО РОССИЙСКОЙ
 ФЕДЕРАЦИИ

ПОСТАНОВЛЕНИЕ

от 12 июля 1993 г. № 650 г. Москва

Об утверждении Положения о Министерстве науки и технической политики Российской Федерации

Совет Министров — Правительство Российской Федерации постановляет:

1. Утвердить прилагаемое Положение о Министерстве науки и технической политики Российской Федерации.

2. Установить Министерству науки и технической политики Российской Федерации резерв средств на финансирование научно-исследовательских и опытно-конструкторских работ, необходимость выполнения которых

возникает в течение года, в размере 1,5 % государственных бюджетных ассигнований, выделяемых Министерству.

Председатель Совета Министров —
Правительства Российской Федерации В.Черномырдин

** Том, документ, стр: RZ9659-002

УТВЕРЖДЕНО

постановлением Совета Министров —
Правительства Российской Федерации
от 12 июля 1993 г. № 650

ПОЛОЖЕНИЕ

О Министерстве науки и технической политики
Российской Федерации

1. Министерство науки и технической политики Российской Федерации (Миннауки России) является центральным органом федеральной исполнительной власти, обеспечивающим формирование и практическую реализацию государственной научно-технической политики, осуществление мер по сохранению и развитию научно-технического потенциала России.

9. Финансирование центрального аппарата Миннауки России осуществляется за счет средств республиканского бюджета Российской Федерации, предусматриваемых на содержание центральных органов федеральной исполнительной власти Российской Федерации.

10. Миннауки России осуществляет организационно-техническое обеспечение деятельности Межведомственной координационной комиссии по научно-технической политике.

11. Миннауки России является юридическим лицом, имеет печать с изображением Государственного герба Российской Федерации и со своим наименованием, расчетный и иные счета в учреждениях банков.

Часть 2. Протокол навигации в БД ЮРИУС (RZ) при поиске по полным текстам

Структура файла частотного словаря полных текстов.

```
SELE 1
USE DCFRV INDE DCFRV
DISPLAY STRUCTURE
```

Structure for table: d:\foxprow\dcfrv.dbf

Number of data records: 10951

Date of last update: 12.10.02

Field	Field Name	Type	Width	Dec	Index
1	VALUF	Character		20	

```
2 FREQF Numeric      6
** Total **          27
```

Структура файла инверсного списка.

```
SELE 2
USE DCIND INDE DCIND
DISPLAY STRUCTURE
```

Structure for table: d:\foxprow\dcind.dbf

Number of data records: 43264

Date of last update: 18.02.94

Field	Field Name	Type	Width	Dec	Index
1	ADRES	Numeric	6		
2	PAGEN	Character	10		
3	WORFR	Numeric	3		
** Total **			20		

После открытия частотного словаря полных текстов, инверсного списка, файла полных текстов — поиск термина по основе «ПРАВО» и распечатка следующих 17 слов.

```
SEEK [ПРАВО]
LIST NEXT 17
```

Record#	VALUF	FREQF
588	ПРАВО	57
4353	ПРАВОВОГО	4
2077	ПРАВОВОЙ	7
10662	ПРАВОВОМ	1
4266	ПРАВОВОМУ	2
1338	ПРАВОВЫЕ	1
9460	ПРАВОВЫМ	1
4516	ПРАВОВЫМИ	1
3684	ПРАВОГО	2
592	ПРАВОМ	3
7167	ПРАВОМОЧНО	1
5191	ПРАВОМОЧНЫМ	2
631	ПРАВОМОЧНЫМИ	2
10148	ПРАВОНАРУШЕНИЕ	1
10141	ПРАВОНАРУШЕНИЯ1	
9216	ПРАВОНАРУШЕНИЯМИ	1
5382	ПРАВООХРАНИТЕЛЬНЫХ	1

Поиск по коду слова ПРАВО (588) — поиск списка содержащих его страниц полных текстов. Выдача части этого списка страниц.

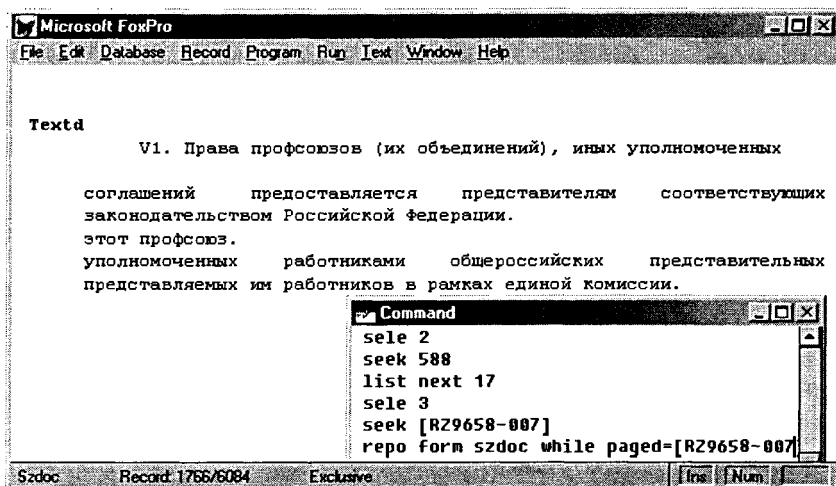
```
SEEK 588
LIST NEXT 17
```

Record#	ADRES	PAGEN	WORFR
809	588	RZ9604-003	
5238	588	RZ9631-022	
6728	588	RZ9634-002	
6840	588	RZ9634-003	
7906	588	RZ9641-001	
10327	588	RZ9658-007	
10408	588	RZ9658-008	
10624	588	RZ9658-010	
11595	588	RZ9659-006	
12353	588	RZ9660-004	
13736	588	RZ9664-005	
15229	588	RZ9669-006	
15356	588	RZ9669-008	
16977	588	RZ9675-004	
17121	588	RZ9675-005	
17604	588	RZ9676-004	
19209	588	RZ9678-002	

Поиск какой-либо из перечисленных страниц (RZ9658-007) и ее распечатка.

```
SEEK [RZ9658-007]
```

```
REPO FORM SZDOC WHILE PAGED=[RZ9658-007]
```



Оглавление

Введение	3
Глава 1. ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ. ОСНОВНЫЕ ПОНЯТИЯ И КЛАССИФИКАЦИЯ	9
1.1. Информационные объекты и процессы. Классификация	9
1.2. Основные классы информационных технологий	14
1.3. Автоматизированные информационные системы (АИС), структура и классификация	20
1.4. Мировые информационные ресурсы, их структура и классификация	27
1.5. Обеспечивающие компоненты (подсистемы) АИС	40
1.6. Состав и структура информационного обеспечения АИС	54
1.7. Структуры баз данных	78
1.8. Организационное обеспечение и пользователи АИС	82
Глава 2. БАЗОВЫЕ ТИПЫ ИНФОРМАЦИОННЫХ СИСТЕМ	86
2.1. Фактографические АИС	88
2.2. Документальные системы	112
2.3. Экспертные системы	155
2.4. Гипертекстовые АИС	170
Глава 3. ТЕХНОЛОГИИ ПРОЕКТИРОВАНИЯ АИС	185
3.1. Классификация методов проектирования систем	187
3.2. Жизненный цикл ИС	191
3.3. Содержание и методы канонического проектирования ИС	199
3.4. Проектирование ИС и РБП — реинжиниринг бизнес-процессов (BPR — business process reengineering)	220

3.5. Автоматизация проектирования АИС (CASE-технологии)	236
3.6. Методы и модели оценки и измерения эффективности АИС	247
Глава 4. АВТОМАТИЗИРОВАННЫЕ ИС ПО ОБЛАСТЯМ ПРИМЕНЕНИЯ	260
4.1. Автоматизированные информационные системы по законодательству	260
4.2. Офисные информационные системы	276
4.3. Автоматизированные информационные системы электронной коммерции	296
4.4. Автоматизированные библиотечные информационные системы (АБИС)	326
4.5. Географические информационные системы (ГИС)	342
4.6. Корпоративные интегральные АИС (MRP и ERP-системы)	353
4.7. Другие типы систем и прикладных программных продуктов	369
Заключение	383
Литература	385
Приложение 1. Глоссарий	388
Приложение 2. Обзор основных команд системы программирования FoxPro	401
Приложение 3. Методические указания по выполнению лабораторной работы «Исследование физической структуры и навигации в БД на примере АИСЗ ЮРИУС»	403
Приложение 4. Отчет о лабораторной работе № 1 «Исследование физической структуры и навигации в БД на примере АИСЗ ЮРИУС»	408

**Емельянова Наталия Захаровна
Партыка Татьяна Леонидовна
Попов Игорь Иванович**

Основы построения автоматизированных информационных систем

Учебное пособие

Редактор *И. С. Холод*
Корректор *А. В. Волковицкая*
Компьютерная верстка *И. В. Кондратьевой*
Оформление серии *Л. Зарецкая*

Сдано в набор 10.09.2006. Подписано в печать 15.11.2006. Формат 60х90/16.
Печать офсетная. Гарнитура «Таймс». Усл. печ. л. 26,0. Уч.-изд. л. 25,6.
Бумага типографская. Доп. тираж 5 000 экз. Заказ № 4772

Издательство «ФОРУМ»
101831, Москва — Центр, Колпачный пер., д. 9а
Тел./факс: (495) 625-32-07, 625-52-43
E-mail: mai@forum-books.ru

ЛР № 070824 от 21.01.93
Издательский Дом «ИНФРА-М»
127282, Москва, ул. Полярная, д.31в
Тел.: (495) 380-05-40
Факс: (095) 363-92-12
E-mail: books@infra-m.ru
Http://www.infra-m.ru

Отпечатано с предоставленных диапозитивов
в ОАО «Тульская типография». 300600, г. Тула, пр. Ленина, 109.