

Министерство Образования Российской Федерации  
Ульяновский Государственный Технический Университет  
Факультет Информационных Систем и Технологий  
Кафедра Вычислительной техники

# Информационное обеспечение автоматизированных систем

Составил:  
профессор кафедры ВТ,  
д.т.н. Токмаков Г. П.

г. Ульяновск - 2002

## **СОДЕРЖАНИЕ**

<b>ВВЕДЕНИЕ.....</b>	<b>4</b>
<b>ГЛАВА 1 ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ АС .....</b>	<b>7</b>
1.1 ПРОИЗВОДСТВЕННЫЕ СИСТЕМЫ И ПРОБЛЕМЫ УПРАВЛЕНИЯ...	7
1.1.1 Описание производства.....	16
1.1.2 Понятие информационной системы.....	7
1.1.3 Функциональные подсистемы ИнфС.....	10
1.1.4 Структурные подсистемы ИнфС.....	11
1.1.5 Обеспечивающие подсистемы ИнфС .....	14
1.2 ПОНЯТИЕ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ИНФС.....	15
1.2.1 Реквизиты.....	16
1.2.2 Составные единицы информации .....	21
1.2.3 Показатели .....	22
1.2.4 Документы .....	24
1.2.5 Понятия и термины.....	27
<b>ГЛАВА 2 КЛАССИФИКАЦИЯ И КОДИРОВАНИЕ ИНФОРМАЦИИ ..</b>	<b>27</b>
2.1.1 Основные понятия кодирования .....	30
2.1.2 Основные понятия классификации .....	33
2.1.3 Проектирование классификаторов.....	35
2.1.4 Раздельная классификация и кодирование .....	36
<b>ГЛАВА 3 ВНЕМАШИННОЕ ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ АС .....</b>	<b>30</b>
3.1 ДОКУМЕНТАЦИЯ КАК ЗНАКОВАЯ СИСТЕМА .....	30
3.1.1 Первичные документы .....	37
3.1.2 Понятие унифицированной системы документации.....	39
3.2 НОРМАТИВНО-СПРАВОЧНАЯ ИНФОРМАЦИЯ.....	41
3.2.1 Основные понятия .....	41
3.2.2 Организация и ведение фонда НСИ.....	41
<b>ГЛАВА 4 ВНУТРИМАШИННАЯ ФАКТОГРАФИЧЕСКАЯ ИНФОРМАЦИОННАЯ БАЗА .....</b>	<b>43</b>
4.1 СИСТЕМЫ OLTP.....	44
4.1.1 Обработка транзакций в OLTP-системах .....	44
4.1.2 Выполнение транзакций в распределенных БД.....	46

4.1.3	Мониторы транзакций.....	47
4.2	СИСТЕМЫ, ОРИЕНТИРОВАННЫЕ НА АНАЛИЗ ДАННЫХ .....	49
4.2.1	Хранилища данных.....	49
4.2.2	Модели данных, используемые для построения хранилищ .....	53
<b>ГЛАВА 5 ВНУТРИМАШИННАЯ ДОКУМЕНТАЛЬНАЯ ИНФОРМАЦИОННАЯ БАЗА .....</b>		<b>60</b>
5.1	НАЗНАЧЕНИЕ И ОСНОВНЫЕ ПОНЯТИЯ.....	61
5.2	ОБЩАЯ ФУНКЦИОНАЛЬНАЯ СТРУКТУРА ДОКУМЕНТАЛЬНЫХ ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМ.....	63
5.3	АНАЛИЗ ТЕКСТА И ИНДЕКСИРОВАНИЕ.....	64
5.3.1	Структура индекса .....	65
5.3.2	Инвертирование текста .....	68
5.4	ПОИСК.....	68
5.4.1	Поиск на точное совпадение терминов. ....	68
5.4.2	Поиск по образцу в полном тексте.....	69
5.4.3	Поиск с использованием КА.....	71
<b>ГЛАВА 6 ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ИНФС .....</b>		<b>74</b>
6.1	ПОНЯТИЕ ЛИНГВИСТИЧЕСКОГО ОБЕСПЕЧЕНИЯ ИНФС .....	75
6.2	ПОНЯТИЕ ЯЗЫКА «ДЕЛОВОЙ ПРОЗЫ» .....	76
6.3	СРЕДСТВА ОПИСАНИЯ .....	77
<b>ГЛАВА 7 ЭТАПЫ РАЗРАБОТКИ ИО АС .....</b>		<b>78</b>
7.1	РАЗРАБОТКА ИО АС НА ПРЕДПРОЕКТНОЙ СТАДИИ.....	78
7.1.1	Основные методы создания ИО АС.....	78
7.1.2	Сбор и Анализ данных. ....	79
7.1.3	Разработка требований к ИО АС.....	82
7.2	РАЗРАБОТКА ИО АС НА СТАДИЯХ ПРОЕКТИРОВАНИЯ .....	84
7.2.1	Работы по созданию ИО АС на стадии технического проекта .....	84
7.2.2	Работы по ИО АС на стадии рабочей документации.....	85
7.2.3	Работы по созданию ИО АС на стадии ввода в действие АС.....	85
<b>СПИСОК ЛИТЕРАТУРЫ .....</b>		<b>87</b>

## **ВВЕДЕНИЕ**

Одним из эффективных направлений научно-технического прогресса и совершенствования управления являются автоматизированные системы управления (АС). Эффективность применения АС для управления такими объектами как промышленные предприятия, банки, торговые организации, государственные учреждения и т. д., зависит от степени интеграции на основе функции управления, от способности оперативно подготавливать управленческие решения и способности адаптироваться к изменениям внешней среды. АС с момента появления первых ЭВМ претерпели существенное изменение в своем развитии:

- ✧ В 50-е годы на ЭВМ в основном решались отдельные экономические задачи, связанные с необходимостью переработки больших информационных массивов.
- ✧ В 60-е годы возникает идея комплексной автоматизации управления предприятиями и интеграции информационного обеспечения (ИнфоО) на основе баз данных (БД).
- ✧ Реальностью АС стали в 70-ые годы на базе ЭВМ 3-го поколения, которые позволили создать вычислительные системы с распределенной терминальной сетью. Однако недостаточное быстродействие и надежность вычислительных машин, отсутствие гибких средств реализации информационных потребностей пользователей не смогли превратить АС в инструмент коренного повышения эффективности управления предприятиями.
- ✧ 80-е годы отмечены широким использованием персональных компьютеров управленческими работниками, созданием большого набора автоматизированных рабочих мест (АРМ) на базе языков 4-го поколения (4GL), позволяющих с помощью генераторов запросов, отчетов, экранных форм, диалога быстро разрабатывать удобные для пользователей приложения. Однако расщепление АС в виде АРМов, локальная автоматизация не способствова-

ли интеграции управленческих функций и, как следствие, существенному повышению эффективности управления предприятием.

- ✧ Для 90-х годов характерно развитие телекоммуникационных средств, которое привело к образованию гибких локальных и глобальных вычислительных сетей, предоопределившие возможность разработки и внедрения корпоративных АС, которые называются также интегрированными АС (ИАС). ИАС объединяют возможности систем комплексной автоматизации управления 70-х годов и локальной автоматизации 80-х годов. Наличие гибких средств связи управленческих работников в процессе хозяйственной деятельности, возможность коллективной работы как непосредственных исполнителей, так и менеджеров, принимающих управленческие решения, позволяет во многом пересмотреть принципы управления предприятиями.

Но эти возможности пока не стали повсеместной реальностью. Как мы уже отметили, еще в 60-е годы была предложена идея интеграции на основе общих данных для решения различных задач. В рамках отдельных предприятий создание общей БД вполне разрешимая задача, но создание такой базы данных на национальном и международном уровне практически нереализуемая задача, так как глобальная схема должна определять полный универсум элементов данных предметной области (ПрДО). Поэтому вместо одной глобальной АС создаются множество АС на разных уровнях управления, связанные между собой информационными каналами. Но при разделении единой системы на ряд взаимосвязанных подсистем и задач утрачиваются некоторые связи между ними, что приводит к потере целостности АС, целостности ее Инфо. Необходимы качественные изменения в разработке АС, общесистемная увязка всех частей и элементов единой системы управления, распределенной на разных информационных системах (ИнфС), путем создания единого информационного пространства (ЕИП), как в рамках одной страны, так и во всем мире. ЕИП представляет собой совокупность АС и технологий их ведения и использования, информационно-телекоммуникационных систем, функционирующих на основе единых принципов и по единым правилам, обеспечивающим информационное взаимодействие предприятий и организаций.

Этот новый этап совершенствования рассматривается как интеграция, под которой понимается процесс установления и формализации объективно существующих взаимосвязей в многоуровневой иерархической системе АС. Отличительной чертой ИАС является связанность, т.е. соединение в единое целое отдельных, обладающих известной автономностью, частей системы. ИАС представляют собой многоуровневые иерархические АС, обеспечивающие комплексную автоматизацию управления на всех уровнях. Причем эта интеграция обеспечивает системе новое качество, расширяет ее возможности и повышает эффективность.

Важным фактором при построении ИАС является разработка основных направлений интеграции, среди которых выделяется Инфо АС. Информационная интеграция заключается в едином комплексном подходе к созданию едином комплексном подходе к созданию информационной базы системы и ее управлению на основе единого технологического процесса автоматизированного сбора и обра-

ботки информации. При этом возрастает роль ИнфО и требования к нему и появились условия построения единой централизованной информационной системы.

При проектировании ИнфО АС необходимо особое внимание уделить ИнфО руководителей. Информация – основа процесса принятия управленческих решений, которые являются продуктами деятельности руководителя, возглавляющего тот или иной орган управления в целом или его структурное подразделение. Чем выше уровень руководства, тем важнее проблема обеспечения информацией. Поскольку весь комплекс управленческих вопросов реализуется через решение руководителей, то, в конечном счете, эффективность АС определяется степенью удовлетворения их информационных потребностей. С этой целью создаются специальные подсистемы для линейных руководителей – подсистемы «Руководитель» в составе ИАС. *Процесс управления – это процесс распознавания и разрешения проблемных ситуаций.* В подсистеме «Руководитель» необходимо иметь модели проблемных ситуаций (ПрбС), где содержится постановка задач.

Таким образом, в процессе разработки ИнфО АС необходимо выявить и определенным образом организовать информацию объекта управления, на основе которой решается комплекс задач, спроектировать средства ее подготовки и ведения как во внешнем, так и во внутримашинном мире, а также средства выдачи информации для решения той или иной задачи. Состав и содержание ИнфБ, а также технология и организация основных работ во многом являются типичными. Это позволяет обобщить рекомендации по разработке ИнфБ АС.

## **ГЛАВА 1 ОСНОВНЫЕ ПОНЯТИЯ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ АС**

### **1.1 ПРОИЗВОДСТВЕННЫЕ СИСТЕМЫ И ПРОБЛЕМЫ УПРАВЛЕНИЯ**

Методологическую основу проектирования ИАС составляет системный подход, в соответствии с которым любой проектируемый объект представляет собой совокупность взаимосвязанных компонентов (элементов) функционирующих совместно для достижения общей цели.

#### **1.1.1 ПОНЯТИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ**

Понятие «ИнфС объекта управления» широко применяется при изучении действующего на объекте документооборота и состава форм документов, информационных связей между подразделениями, организации формирования, приема, передачи данных на всех уровнях управления, состояния нормативно-справочной информации и систем классификации и кодирования, а также других аспектов обеспечения персонала информацией, необходимой для планирования производственных процессов, учета состояния объекта и принятия решения по управлению им. Так, ИнфС предприятия как объекта управления рассматривается как многоуровневая коммуникационная сеть системы управления, в рамках которой реализуются процессы по поступлению, передаче, обработке, накоплению и выдаче информации.

При сложившихся традиционных методах обработки информации (без использования АС) для целей управления объектом каждое функциональное подразделение (плановый отдел, бухгалтерия, отдел материально-технического снабжения и т. д.) имеет свои информационные связи с производственными подразде-

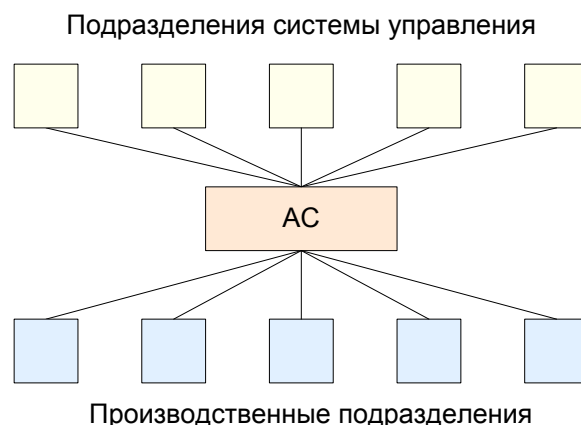
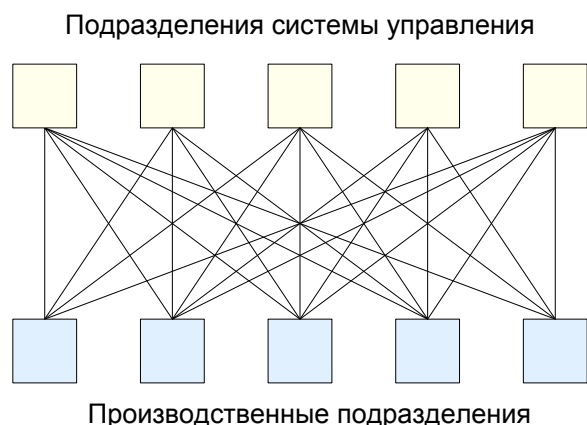
лениями (цехами, складами и т. д.). При этом информационные потоки во многом дублируются (см. Рис.1). Однородные данные, полученные одним функциональным подразделением не могут быть реализованы другим подразделением ввиду разной формы представления. Это приводит к большим затратам на обработку информации.

Создание и функционирование АС, базирующейся на интеграции обработки данных с применением новейших средств и методов, вот основное назначение использования автоматизированных методов обработки информации. При организации интегрированной системы обработки существенно изменяются традиционные информационные потоки (см. Рис.2). Здесь данные направляются от производственных подразделений в центр обработки, который связан с памятью маши-

*Рис.1. Структура ИнфС при традиционной обработке данных*

*Рис.2. Структура ИнфС при интегрированной обработке данных*

ны. И ведет необходимые расчеты для снабжения функциональных служб соответствующей информацией.



Характер производственных процессов оказывает влияние на организационную структуру предприятия и на функциональные обязанности управленческого персонала отдельных подразделений. Эти особенности оказывают влияние на перечень показателей, которыми оперирует управленческий персонал данной функции на конкретном предприятии. Перечень показателей специфичен по составу и вполне определен для каждой функции, так как обуславливает существенные, стабильные для данной системы действия. Все это должно быть отражено в структуре АС, в которой фактически фиксируется модель производственной и управленческой деятельности предприятия..

Любая система управления может функционировать только в том случае, если в ней циркулирует информация. Поэтому процесс управления связан с хранением, преобразованием и использованием информации, отражающей всю производственно-хозяйственную деятельность управляемого объекта в различных показателях. Любой процесс управления – это, прежде всего, информационный процесс, который предполагает выполнение функций по сбору, передаче, обра-



ботке, анализу информации и тию соответствующих решений. В общем, схематическом виде ренные информационные потоки приведены на Рис.3, и включает в себя следующие основные элементы:

- ✧ на вход системы поступают трудовые и материальные ресурсы, необходимые для производства продукции;
- ✧ производственно-технологический блок, или управляемый объект, отражает непосредственный процесс производства, в ходе которого осуществляется переработка ресурсов;
- ✧ на выход системы поступает готовая продукция, что характеризует достижение поставленной цели;
- ✧ блок управления осуществляет информационную связь (прямую и обратную) компонентов производственной системы, обеспечивая их целенаправленное и согласованное функционирование.

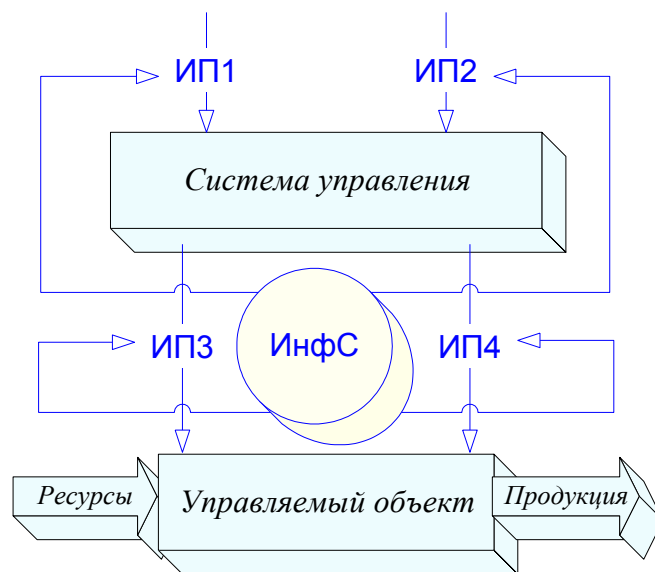


Рис.3. Структура управляемой системы

Основные информационные потоки между внешней средой, управляемым объектом и системой управления помечены метками ИП1, ИП2, ИП3, ИП4 и связаны с поддерживающей их *ИнфС*. *ИнфС* представляет собой совокупность технических, программных и информационных средств, объединенных в единую систему с помощью информационных потоков с целью сбора, хранения, обработки и выдачи необходимой информации, предназначенной для целей управления, где:

- ✧ ИП1 – информационный поток из внешней среды в систему управления, который, с одной стороны, представляет задание на изготовление продукции, поток нормативной информации, создаваемый государственными учреждениями в части законодательства, а, с другой стороны, – поток информации о состоянии внешней среды;
- ✧ ИП2 – информационный поток из системы управления во внешнюю среду, а именно: отчетная информация, всевозможные запросы и т. д.;
- ✧ ИП3 – информационный поток из системы управления на управляемый объект (прямая связь), представляющая собой совокупность плановой, нормативной и распорядительной информации для осуществления процессов по изготовлению продукции;
- ✧ ИП4 – информационный поток от управляемого объекта в систему управления (обратная связь), который отражает учетную информацию о состоянии управляемого объекта (наличия сырья, материальных, денежных, энер-

гетических, трудовых ресурсов) в результате выполнения процессов по изготовлению продукции.

### 1.1.2 ФУНКЦИОНАЛЬНЫЕ ПОДСИСТЕМЫ ИнфС

С позиции кибернетики процесс управления системой как направленное воздействие на компоненты системы для достижения цели можно представить в виде информационного процесса, связывающего внешнюю среду, управляемый объект и систему управления. При этом внешняя среда и управляемый объект формируют систему управления о своем состоянии, система управления анализирует эту информацию, вырабатывает управляющее воздействие на объект управления, отвечает на возмущение внешней среды и при необходимости модифицирует цель и структуру всей системы.

В кибернетическом аспекте цель – это будущий результат деятельности системы, достигаемый с помощью принципа обратной связи. Достижение цели происходит в результате решения определенной задачи функционирования. Средствами достижения поставленных целей (выполнение возложенных функций) служат алгоритмы функционирования. Поэтому иерархии целей и задач порождают иерархию алгоритмов и узлов управления, т. е. определенную материальную структуру, в которой эти алгоритмы реализуются.

Узлы управления, соответствующие иерархии алгоритмов функционирования предприятия (или организации), соответствуют его функциональной структуре. Делимость управляющей системы на подсистемы имеет ряд достоинств с точки зрения разработки и эксплуатации ИнфС, к которым относятся:

- ✧ упрощение разработки и модернизации ИнфС в результате специализации разработчиков по подсистемам;
- ✧ упрощение внедрения и постановки готовых подсистем в соответствии с очередностью выполнения работ;
- ✧ упрощение эксплуатации ИнфС вследствие специализации работников предметной области.

Подсистемы, обслуживающие определенные виды деятельности, характерные для структурных подразделений предприятия или функций управления называются *функциональными подсистемами (ФП)*. ФП ИнфС представляет собой комплекс задач с высокой степенью информационных обменов (связей) между задачами. При этом под *задачей* будем понимать некоторый процесс обработки информации с четко определенным множеством входной и выходной информации (например, начисление зарплаты, учет прихода материалов и т. д.). Состав функциональных подсистем во многом определяется особенностями производственной системы, ее отраслевой принадлежностью и характером деятельности.

На практике чаще всего применяется предметно-функциональный подход разбиения системы управления, согласно которому построение функциональной структуры ИнфС – это разделение ее на подсистемы по характеру производственной деятельности, которое должно соответствовать структуре объекта и системе



Рис.4. Решение задач функциональных подсистем

управления, а также характеру выполняемых функций управления (см. Рис.4). Подсистемы, построенные по функциональному принципу, охватывают все виды деятельности предприятия (производство, снабжение, сбыт, финансы). Подсистемы, построенные по предметному принципу, относятся в основном к оперативному уровню управления ресурсами.

### 1.1.3 СТРУКТУРНЫЕ ПОДСИСТЕМЫ ИнФС

Процесс управления производством основан на анализе данных, поступающих от структурных подразделений предприятия. Эффективность работы структурного подразделения зависит от своевременности поступления информации и от ее качества.

Само по себе наличие таких видов действительности, как сбор, передача, обработка информации еще не определяет понятия ИнФС. Только когда определены методы и способы использования имеющейся информации, можно говорить о понятии ИнФС. ИнФС функционируют, прежде всего, для обеспечения информацией процессов управления, а это связано с методами и средствами выделения информации из совокупности зафиксированных данных. Другими словами, ИнФС является системой, направленной на «производство» информации для управления.

Любая сфера управленческой действительности базируется на определенной системе информации. Однако в качестве системы она может рассматриваться только тогда, когда четко регламентировано информационное взаимодействие между людьми, подразделениями, организациями, когда выделены их информационные потоки и имеется возможность их регулирования.

В соответствии с характером обработки информации в ИнфС на различных уровнях управления системой (оперативном, тактическом и стратегическом) выделяются следующие компоненты ИнфС:

- ✧ *Системы обработки данных (СОД)* – предназначены для учета и оперативного регулирования управляемых процессов, подготовки стандартных документов для внешней среды. Для СОД характерны такие операции как регистрация и обработка событий, оформление и выполнение заказов, приход и расход материальных ценностей и т. д. Эти задачи имеют итеративный характер, выполняются непосредственными исполнителями процессов и связаны с оформлением и пересылкой документов, в соответствии с четко определенными алгоритмами. Результаты выполнения операций вводятся через экранные формы в БД.
- ✧ *Система управления (СУ)* – ориентирована на тактический уровень управления: среднесрочное планирование, анализ и организацию работ. Для данного класса задач характерны периодическая повторяемость формирования документов и четко определенный алгоритм решения задач. Решение подобных задач предназначено для руководителей различных служб предприятия. Задачи решаются на базе оперативных данных.
- ✧ *Система поддержки принятия решений (СППР)* – используются на верхнем уровне управления, имеющего стратегическое долгосрочное значение. Задачи СППР имеют нерегулярный характер и для них характерна недостаточность имеющейся информации, ее противоречивость и нечеткость, слабая формализованность алгоритмов ее решения. В качестве инструментов обобщения используются средства составления аналитических отчетов, методы статистического анализа, экспертных оценок, математического и имитационного моделирования. При этом используются базы обобщенной информации, информационные хранилища, базы знаний о правилах и моделях принятия решений.

Идеальной считается ИнфС, которая включает все три типа перечисленных систем. В зависимости от охвата функций и уровней управления различают корпоративные (интегрированные) и локальные ИнфС. *Корпоративная ИнфС* автоматизирует все функции управления на всех уровнях управления, является многопользовательской и функционирует в распределенной вычислительной сети. *Локальная ИнфС* автоматизирует отдельные функции управления на отдельных уровнях управления и функционирует в отдельных подразделениях системы управления.

В состав ИнфО включаются два комплекса: компоненты немашинного ИнфО (ЕСККИ, УСД, система входных и выходных сообщений) и внутримашин-

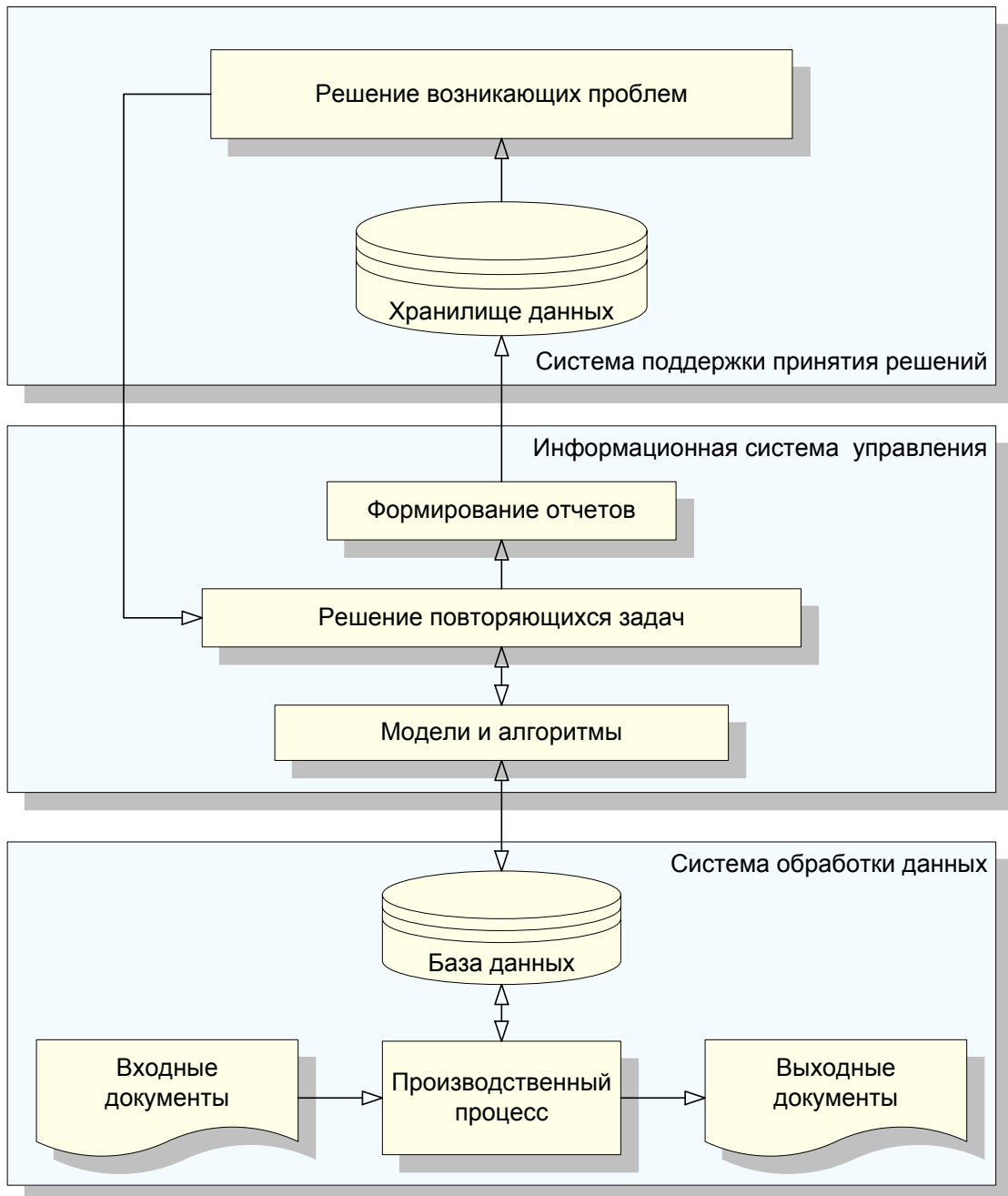


Рис.5. Типы ИнфС

ного ИнфО (экранные формы для ввода данных или вывода результатов, структура ИБ, БД). Центральным компонентом ИнфО является БД, через которую осуществляется обмен данными различных задач. БД обеспечивает интегрированное использование различных информационных объектов в функциональных подсистемах.

*Весь комплекс данных, отражающих течение процессов производства и необходимый для эффективного управления технологическим оборудованием, определяется как ИнфО, является основой для построения любых систем автоматизированного управления (АС) и во многом определяют их экономическую эффективность.*

ИнфО включает также системы классификации и кодирования информации. Вся перечисленная информация представляет собой один из наиболее важных ресурсов предприятия, необходимая для принятия эффективных и своевременных управленческих решений. Для того чтобы приспособить этот ресурс для эффективного поиска, обработки на ЭВМ и передачи по каналам связи, ее необходимо представить в цифровом виде. С этой целью ее нужно сначала упорядочить (классифицировать), а затем формализовать (закодировать) с использованием классификатора. *Классификатор* – это документ, с помощью которого осуществляется формализованное описание информации в ИнфС, содержащей наименования объектов, наименования классификационных группировок и их кодовые обозначения.

Итак, для изучения систем (предприятий) целесообразно выделить три аспекта:

- ✧ вещественно-энергетический аспект охватывает различные виды вещественных и энергетических связей в исследуемой системе, а также законы и формы преобразования энергетических связей в исследуемой системе, а также законы и формы преобразования энергии и вещества в процессе функционирования системы. Этот аспект непосредственно касается технологических процессов в системе и связанных с ним явлений;
- ✧ экономический аспект исследования систем затрагивает вопросы эффективности функционирования систем;
- ✧ информационный аспект соответствует рассмотрению системы с точки зрения передачи, преобразования и хранения информации.

Нас, в первую очередь, будет интересовать последний аспект, ибо процесс управления предприятием связан с хранением, преобразованием и использованием информации, отражающей всю производственную деятельность предприятия в различных показателях (определение показателя см. ниже).

Подсистема «*Лингвистическое обеспечение*» (ЛО) включает совокупность научно-технических терминов и других языковых средств, используемых ИнфС, а также правил формализации естественного языка (ЕЯ), включающих методы сжатия и раскрытия текстовой информации с целью повышения эффективности автоматизированной обработки информации и облегчающих общение человека с ИнфС. Языковые средства, включенные в подсистему ЛО, делятся на две группы: традиционные языки (естественные, математические, алгоритмические языки, языки моделирования) и языки, предназначенные для диалога с ЭВМ (ИПЯ, языки СУБД, языки ОС, входные языки ППП).

#### 1.1.4 ОБЕСПЕЧИВАЮЩИЕ ПОДСИСТЕМЫ ИнфС

Обеспечивающие подсистемы ИнфС являются общими для всей ИнфС независимо от конкретных функциональных подсистем, в которых применяются те или иные виды обеспечения. В состав обеспечивающих подсистем входят подсистемы технического, математического, программного, информационного и лингвистического обеспечения.

Подсистема «*Техническое обеспечение*» (*ТО*) представляет комплекс технических средств, предназначенных для обработки данных в ИнфС. В состав комплекса входят компьютеры, осуществляющие обработку информации, средства подготовки данных на машинных носителях, средства сбора и регистрации информации, средства передачи данных по каналам связи, средства накопления и хранения данных и выдачи результирующей информации.

Подсистема «*Математическое обеспечение*» (*МО*) – это совокупность математических моделей и алгоритмов для решения задач и обработки информации с применением вычислительной техники, а также комплекс средств и методов, позволяющих строить математические модели задач управления. В состав МО входят: средства моделирования типовых задач управления, методы многокритериальной оптимизации, теории массового обслуживания, описание задач, алгоритмы решения задач.

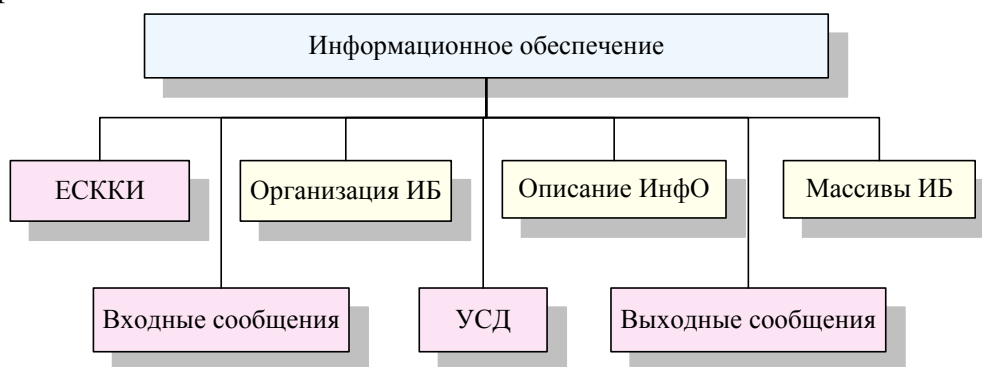


Рис.6. Состав подсистемы Инфо ИнфС

Подсистема «*Программное обеспечение*» (*ПО*) включает совокупность компьютерных программ, описаний и инструкций по их применению. ПО делится на два комплекса: общее (ОС, операционные оболочки, компиляторы, интерпретаторы, программные среды для разработки прикладных программ, СУБД, сетевые программы и т. д.) и специальное (совокупность прикладных программ, разработанных для конкретных задач в рамках функциональных подсистем).

Подсистема «*Информационное обеспечение*» (*Инфо*) – это совокупность единой системы классификации и кодирования информации (ЕСККИ), унифицированной системы документации (УСД) и информационной базы (ИБ). Состав Инфо приведен на Рис.6.

## 1.2 ПОНЯТИЕ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ИНФС

Как наука, так и практика управления неотделимы от теории и практики информации. Для повышения эффективности управления необходимо теоретическое и практическое овладение процессами сбора, хранения, передачи, переработки и выдачи информации, т. е. информационными процессами. Кибернетика показала огромное значение информации в управлении, неразрывное

единство процессов управления и информации. Достоверная информация нужна для выработки и принятия управленческого решения. Информация нужна и для регулирования системы. Поэтому информационные процессы являются необходимым атрибутом управления. АС (или ИнфС) прежде всего, обслуживают функции системы управления и их главная цель – организация информационного обеспечения всех основных функций управления.

Специалисты в области разработки АС часто соотносят понятие ИнфО с понятием «служебной информации»: системами словарей, кодовых таблиц, классификаторов. Но в последнее время среди специалистов, работающих в данной области, преобладает другое представление об ИнфО, которое представляется как *определенная совокупность элементов информации: реквизитов, составных единиц информации, показателей, классификаторов, языков записей данных, правил структурной организации массивов, документов, обеспечивающих структурную организацию информации в системе.*

### 1.2.1 ОПИСАНИЕ ПРОИЗВОДСТВА

О такой системе, как предприятие или организация, говорят, что она состоит из ряда взаимоувязанных подсистем, сама входит в состав подсистем типа объединение, отрасль и вместе с ними составляет единую систему – производственный комплекс страны. В процессе решения задач весь взаимосвязанный комплекс систем и подсистем изменяется под воздействием целого ряда факторов как связанных с решением задач, так и внешних обстоятельств вероятностного характера. Задача управления состоит в поддержании этих изменений в пределах соответствующих заданным параметрам функции управления.

С функциональной точки зрения предприятие или организация – это система, представляющая собой устойчивую формальную социальную структуру, которая берет ресурсы из окружающей среды и обрабатывает их, чтобы произвести продукцию. Чтобы изготовить продукцию, необходимы такие ресурсы как материя, энергия и информация:

- ✧ материальный поток образуется загрузкой сырья, материалов, комплектующих изделий (при интеллектуальной деятельности, например, разработке ПО, входной материальный поток составляют исходные данные, содержащиеся в документах);
- ✧ превращение исходного материала в готовую продукцию осуществляется за счет энергии в виде последовательности действий, каждое из которых вызывает преобразование материального потока. Техническими средствами преобразования материального потока является технологическое оборудование, и каждое воздействие на материальный поток можно рассматривать как элементарную технологическую операцию;
- ✧ для преобразования материи нужна информация, которая определяет форму воздействия на элементы материального потока. Эта информация поступает на технологическое оборудование.



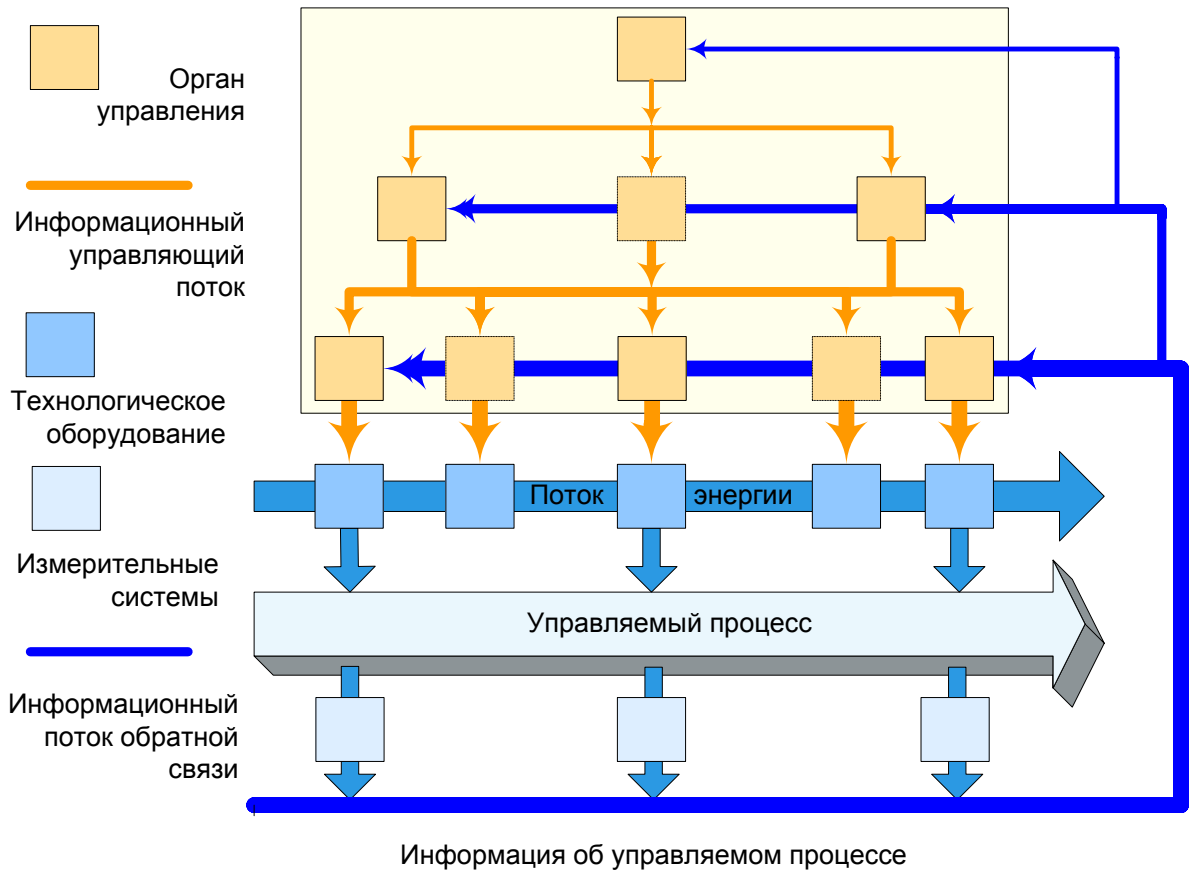


Рис.7. Схема основных потоков при управлении материальным процессом

Первичным источником этой информации является концептуальная модель, т.е. характеристики, которые должно иметь изделие. По существу концептуальная модель является техническим заданием на разработку. Для изготовления изделия необходимо трансформировать эту информацию в алгоритмическую модель, представляющую собой последовательность команд для технологического оборудования, т.е. получить информацию о том, как изготовить данное изделие. После каждой технологической операции надо выполнить обратное преобразование, т.е. изменение состояния потока материи должно быть переведено в информацию. Для этого служат измерительные системы, которые создают поток информации обратной связи. Итак, мы выяснили, что в процессе изготовления изделия на производстве циркулируют два основных принципиально различающихся информационных потока:

- ❖ Управляющий поток – информация, идущая от концептуальной модели к технологическому оборудованию (документы). Этот поток определяет, что должно быть сделано.
- ❖ Поток обратной связи – информация, идущая от измерительных систем (сообщения). Этот поток определяет, что именно и как сделано к определенному моменту времени.

Информационные потоки представляют собой движение информации на бумажных или машинных носителях либо (в случае ИАС) по информационным каналам вычислительной сети.

При сопоставлении информационных потоков выявляется информация согласования, на основе которой принимается решение по управлению. Поэтому, при выполнении определенной работы, решении задачи приходится иметь дело с различного рода информационными совокупностями – массивами данных, документами и недокументированными сообщениями.

- ✧ В процессе решения задачи формируются различные *массивы* для сортировки, обработки и анализа данных. Массивы данных, которые отражают устойчивые свойства объектов и явлений производственно-экономической деятельности предприятия или организации, сохраняются на машинных носителях. Массивы, формируемые в АС, подразделяются на две категории:
  - *постоянные* – включают нормативно-справочную и другую информацию длительного хранения и многократного применения для решения задач;
  - *текущие* – концентрируют оперативную информацию о состоянии управляемого объекта или процесса во времени. Данные, из которых они формируются, поступают или в реальном масштабе времени, или пакетами в определенные промежутки времени;
  - *промежуточные* – содержат информацию, возникающую на стыке различных задач или этапов решения одной как результат предыдущего расчета и исходный материал для следующего. необходимость хранения таких массивов может быть вызвана разрывом во времени между моментом их получения и использования;
  - *накапливаемые* – характеризуются накоплением данных в течение определенного промежутка времени.
- ✧ В *документах*, действующих на предприятии, содержится основная, наиболее устойчивая часть сведений, необходимых управленческому персоналу для выполнения своих функций. Документы выполняют две основные функции:
  - они имеют юридическую силу и могут служить основанием для принятия решений управленческим персоналом на разных уровнях управления;
  - они служат физическими носителями упорядоченных сведений об управляемом объекте, поскольку содержат определенное множество показателей и реквизитов, сгруппированных в графах и строках по определенным признакам.
- ✧ Другая часть информации (сообщения), которую можно назвать недокументированной, передается от измерительных приборов либо письменно в виде первичных документов, либо устно при беседах, телефонных разговорах, на совещаниях, либо фиксируется в массивах данных. Это информация о состоянии производственного процесса. Но недокументированное сообщение

можно представить и в виде документа. Таким образом, принципы изучения сведений, представленных в документах, могут быть распространены и на недокументированную информацию.

## 1.2.2 РЕКВИЗИТЫ

В окружающем нас мире находится огромное разнообразие предметов, объектов, явлений, процессов, отображаемых посредством информации. каждая представляемая информацией сущность (объект, явление) имеет ряд характерных для нее свойств. Свойства физической сущности отображаются с помощью переменных величин, являющихся элементарными единицами информации и называемых *реквизитами*. Реквизит – это логически неделимый элемент любой информационной совокупности, соотносимый с определенным свойством отображаемого информацией объекта или процесса. В обрабатываемой информации реквизиты представляются как бы «атомами», из которых komponуются более сложные по структуре образования информации. И, наоборот, единицы информации любой сложности можно последовательным разложением на составляющие компоненты в конечном итоге расчленить на реквизиты. Другими часто встречающимися синонимами реквизита являются элемент, поле, признак, атрибут.

Один и тот же признак может наблюдаться у разных объектов и явлений. Поэтому каждый реквизит, хотя и проявляется только в конкретных объектах и явлениях, обладает известной самостоятельностью и имеет особые, характерные для него черты. Одним из таких свойств реквизита является то, что он может фигурировать в самых различных составных единицах информации, относящихся к различным сущностям. Это свойство реквизита находит свое отражение в форме, всесторонне характеризующей реквизит вне зависимости от его конкретного вхождения в ту или иную составную единицу информации.

*Форма реквизита* определяет полное наименование реквизита, имена-синонимы (включая сокращенные – идентификаторы), типы и классы значений. *Наименование реквизита* (или просто имя) служит для обращения к нему и обычно представляется словом или словосочетанием, названием определенной графы (строки) документа, номером, условным шифром, адресом на носителе информации.

Каждому реквизиту присуще некоторое конечное множество значений, именуемое *классом значений*, которое информационно отражает данный реквизит. *Значение реквизита*, таким образом, есть в каждый заданный момент времени одна из позиций класса значений данного реквизита, отображающая соответствующее состояние того свойства объекта, которое характеризует реквизит.

Существует ряд типов реквизитов в зависимости от видов значений, которые они могут иметь. Наиболее распространенными типами реквизитов являются числовой и текстовый.

Реквизиты *числового типа* характеризуют количественные свойства сущно-

стей, полученные в результате подсчета натуральных единиц, измерения, взвешивания, вычисления на основе других количественных данных и т.п. Поэтому значениями таких реквизитов служат числа со всеми свойственными им чертами и атрибутами, подробно рассматриваемыми в курсах программирования. В конкретных представлениях фигурирует несколько типов числовых величин в зависимости от ограничений на диапазон чисел, фиксации запятой, форматы их представления. Поскольку все реквизиты данного типа активно участвуют в арифметических операциях необходимо предусмотреть соответствующий аппарат преобразования.

Реквизиты *текстового типа* выражают качественные свойства сущностей и характеризуют обстоятельства, при которых имел место изучаемый процесс, и были получены те или иные числовые значения. Поэтому такие реквизиты называются *признаками*.

Для определения понятия каждого из множества окружающих нас предметов, явлений необходимо найти то особенное, что отделяет его от других предметов или явлений, что выражает его внутреннюю суть. Это «особенное» представляется в виде качественных определенностей, присущих отдельным разновидностям или их группам. Многообразие форм материи обуславливает и многообразие форм качественной определенности.

Качественная определенность проявляется через совокупность всех присущих понятию свойств, каждое из которых конкретно выражает какую-либо его сторону, какой-либо его момент. Вещи и явления, обладающие различными свойствами (признаками), по-разному действуют на органы чувств человека и вызывают различные ощущения, благодаря чему и создается возможность их градации и индивидуализации.

Некоторые свойства (признаки) распространяются на многие экземпляры, являются *общими* для многих предметов, явлений. Общие признаки являются признаками сходства и благодаря этому содержат в себе постулат обобщения. Другие свойства (признаки) присущи только единичным экземплярам и поэтому называются *индивидуальными*. Благодаря индивидуальным признакам, являющимся признаками различия, различаются между собой экземпляры группы предметов, имеющих общие свойства. Всякое понятие включает в себя общие и индивидуальные свойства. Главным назначением признаков, таким образом, является:

- ✧ во-первых, указание тех особенностей, которыми одно явление отличается от других, т. е. индивидуализация сообщений, устранение возможности смешивания фактов;
- ✧ во-вторых, представление таких свойств, которые могут послужить основой для обобщения.

Значениями признаков являются последовательности символов, называемые строками, или текстом. Полный набор символов данной ИнФС составляет его алфавит. Класс значений признаков может быть установлен явным перечнем всех конкретных значений, называемого номенклатурой данного признака и разрабо-

тываемого на практике для классификаторов и при зашифровке значений признака (кодировании). Значение реквизита-признака, следовательно, есть значение одной из позиций его номенклатуры – множества класса значений, обычно именуемого массивом.

Кроме текстового и числового типов величин при обработке информации часто применяют логический тип, которым предполагается, что величина имеет одно из двух значений – истинность или ложность. Над логическими величинами осуществляются операции математической логики; они участвуют в логических выражениях, вычисляемые значения которых присваиваются переменным логического типа.

В некоторых применениях встречаются объявления специальных типов данных вроде географических координат, даты и др.

### 1.2.3 СОСТАВНЫЕ ЕДИНИЦЫ ИНФОРМАЦИИ

Каждый из наблюдаемых объектов, процессов характеризует, как уже говорилось, ряд присущих ему свойств. Но точно так же, как взятое в отдельности любое свойство еще не представляет сущность (объект, процесс) в целом, так и изолированно взятый реквизит не может служить полной информацией о наблюдаемом объекте (процессе). Требуется некоторая взаимосвязанная совокупность реквизитов (признаков, чисел) для того, чтобы воспроизвести некоторое *сообщение* о сущности, определенную информацию о явлении.

Каждое  $j$ -е свойство в сообщении  $C_i$  представлено значением определенного приписанному этому свойству реквизита  $R_j$ , так что

$$C_i = (R_1, R_2, \dots, R_j, \dots, R_m),$$

где реквизиты могут быть и признаками, и числовыми переменными в зависимости от того, отображают ли они свойства качественного или количественного характера. При этом каждый реквизит имеет в сообщении лишь одно значение. Однако, поскольку одна и та же сущность (допустим, факт отпуска изделий покупателям) фиксируется многократно с возникновением в каждый раз нового сообщения, значения одного и того же реквизита  $R_j$  меняются в зависимости от изменения обстоятельств.

Так как каждый из  $m$  реквизитов сообщения  $C_i$  может принимать одно из  $k_j$  значений, где  $k_j$  есть длина номенклатуры для реквизита-признака и диапазон значений для реквизита числового типа, то потенциальное множество сообщений

данного вида равно декартову произведению  $\prod_{i=1}^m k_j$ . В действительности, из-за на-

личия определенной логической взаимосвязи реквизитов множество значений меньше теоретически возможного, но тем не менее весьма существенно. Каждое сообщение в множестве сообщений данного вида отличается от другого значением хотя бы одного из входящих в сообщение реквизитов. Все множество этих со-

общений объединяется в один вид благодаря одинаковому составу свойств, отображаемых реквизитами, или *формой сообщения*.

Формой сообщения объединяется некоторое множество реквизитов разных форм, т. е. в данном случае налицо некоторое более сложное по составу информационное образование, состоящее из элементарных единиц информации – реквизитов. Единицу информации, состоящую из совокупности других единиц информации, ассоциативно связанных между собой, назовем *составной единицей информации* (СЕИ). В рассмотренном выше примере роль составляющих играли реквизиты  $R_1, R_2, \dots, R_j, \dots, R_m$ .

СЕИ, в свою очередь, может быть единицей информации в составе СЕИ более высокого уровня, образуя таким образом иерархическую систему.

#### 1.2.4 ПОКАЗАТЕЛИ

Все экономические данные как знаковые представления информации, обращающейся в сфере управления производством, с известной условностью можно разделить на две части. К одной из них относятся данные, носящие «качественный» характер: их содержание выражается без количественных характеристик отображаемых объектов и процессов. Основная масса такого рода данных пока не используется в АС, если не считать документальных информационно-поисковых систем библиографического типа. Вторую часть экономических данных образуют количественные данные, т. е. данные, характеризующие не только качественную сторону отображаемых объектов, но и меру их свойств, выраженную численно. Количественные данные называются показателями.

Прежде чем приступить к определению показателя рассмотрим процесс общения между людьми. Человек воспринимает информацию на естественном языке (ЕЯ). Его знакам – словам, словосочетаниям, предложениям – в памяти воспринимающего соответствуют понятия, суждения и более сложные модели реального мира. Восприятие смысла сообщения обеспечивается тем, что знак сопоставляется с вызываемым из памяти соответствующим понятием. В памяти помимо понятий фиксируются различные связи между ними, благодаря чему вместе с понятием, соответствующим полученному знаку, вызываются и связанные с ним другие понятия. Это чрезвычайно важное свойство, так как только в сопоставлении с взаимосвязанными понятиями возможна интерпретация сообщения.

Естественно предполагать, что общение между человеком и машиной можно облегчить, если приблизить используемого при этом языка к структуре естественного, а машинную память строить примерно на тех же принципах, на которых построена человеческая, так, чтобы машина реагировала на запрос подобно человеку. Для этого, прежде всего, необходимо перейти к содержательному кодированию и размещению показателей в памяти по содержательным признакам. Структура памяти, ее организация должны быть построены таким образом, чтобы по некоторому набору содержательных признаков в ней можно было бы отыскать необходимый показатель.

*Показатель* – это СЕИ, состоящая из одного реквизита числового типа, именуемого *основанием показателя*, отражающего тот или иной факт в количественной оценке, и ряда характеризующих его и связанных с ним логическими отношениями *реквизитов-признаков* текстового типа (времени, места действия, действующих лиц, предметов и т. д.). Общий вид показателя следующий:

$$П(P_1, P_2, \dots, P_n, Q),$$

где  $P_1, P_2, \dots, P_n$  – реквизиты-признаки,  $Q$  – реквизит-основание показателя.

Одна из причин выделения показателей в особую разновидность СЕИ является то, что показатель, по существу, минимальная по составу информационная совокупность, сохраняющая информативность и поэтому достаточная для образования самостоятельного документа, который дальше может существовать изолированно от ИнфС, имея свою форму и алгоритм вывода. Отдельно взятые реквизиты и группы из одних признаков-реквизитов или одних реквизитов числового типа таким свойством информативности потенциального самостоятельного документообразования не обладают.

В приведенном определении показатель представляет собой *законченное высказывание с законченным смыслом, включающее как название переменной величины, так и ее конкретное количественное значение со всеми качественными признаками, необходимыми для идентификации последнего*. Такая точка зрения естественна для теории автоматизированной обработки данных, поскольку ее интересует идентификация именно каждого конкретного числа, а также возможность выполнения широкого круга логических операций над качественными признаками, характеризующими различные массивы данных. В период ручной обработки данных на эти моменты не обращали внимания, поскольку предполагалось, что человек сможет легко идентифицировать данные, если будет определен их смысл. Он то и выражается в названиях переменных величин.

Но в настоящее время содержательное кодирование в АС используется недостаточно широко, хотя с появлением СУБД и языка SQL в этом направлении был достигнут существенный прогресс. Другое направление решения этой проблемы – это классификация понятий, которая устанавливает определенные отношения между ними, знаки которых входят в описания показателей, и тем самым позволяют распознавать взаимосвязи между показателями. При таком подходе описание показателей не зависит от того, в каких документах они встречаются и каково их положение в документе. Система классификаций, становится, таким образом, частью языка, единого для всей системы экономической информации и позволяющего начать переход к описанию показателей с помощью содержательного кодирования.

Однако одной системы классификации, как бы хорошо ни была она разработана, недостаточно для полного решения проблемы. В системе классификации устанавливаются только немногие, вполне определенные отношения между понятиями, а именно так называемые классификационные отношения: «род-вид», «целое-часть». На самом деле гамма всевозможных отношений между понятиями существенно шире. Недостаток системы классификаций не в том, что в ней не от-

ражаются эти отношения – они и не должны в ней отражаться. Дело в том, что язык, который ограничивается одной системой классификаций, был бы по природе закрыт для описания этих отношений. Поэтому для полного решения этой проблемы нужен язык с развитой лексикой и грамматикой, на котором можно было бы выразить смысл показателей. Разрабатываемые классификации и номенклатуры должны органически включаться в такой язык. Только в его структуре они окажутся полностью увязанными в единую систему.

### 1.2.5 ДОКУМЕНТЫ

На основе показателей строятся документы. Под *документом* следует понимать такую информационную совокупность, которая имеет самостоятельное смысловое значение и характеризуется полным набором реквизитов и показателей. Данная информационная совокупность должна быть зафиксирована на материальном носителе с существующими правилами и имеющими юридическую силу. Совокупность однородных документов представляет собой массив документов.

Информация, отражаясь в документах, подразделяется на текущую или переменную и постоянную. *Переменная информация* отражает фактические количественные и качественные характеристики производственно-хозяйственной деятельности объекта. Она может меняться для каждого фиксированного случая как по названиям реквизитов признаков, так и по количественной величине реквизитов оснований. Переменная информация, как правило, участвует в одном цикле обработки, поэтому ее еще называют разовой. *Постоянная информация* остается неизменной в течение длительного времени, многократно используется при обработке переменной информации. Постоянная информация подразделяется на справочную, нормативно-расценочную, плановую.

Рассмотрим, к примеру, информационную совокупность, отражающую информацию, которая содержится в таком распространенном документе, как приказ-накладная на отпуск готовых изделий (см. Рис.8). Форму этого документа условно можно разбить на три части: общую, предметную и оформительскую (подписи). Следовательно, представленную этим документом информационную совокупность  $S$  можно соответственно разделить на три информационные совокупности-компоненты:  $C11$  (общая часть),  $C12$  (предметная часть) и  $C13$  (оформительская часть), что можно представить записью

$$S.(C11,C12,C13),$$

где перед точкой указана составная величина  $S$ , точка – знак иерархического отношения (подчинения), а перечисленные в круглых скобках переменные – составляющие величины по отношению к  $S$ , запятые между ними знаки отношения следования в рамках одного уровня.

Составляющая  $C11$ , представленная общей частью документа, в свою очередь является СЕИ и включает три составляющих:  $C11.(C21,C22,C23)$ .



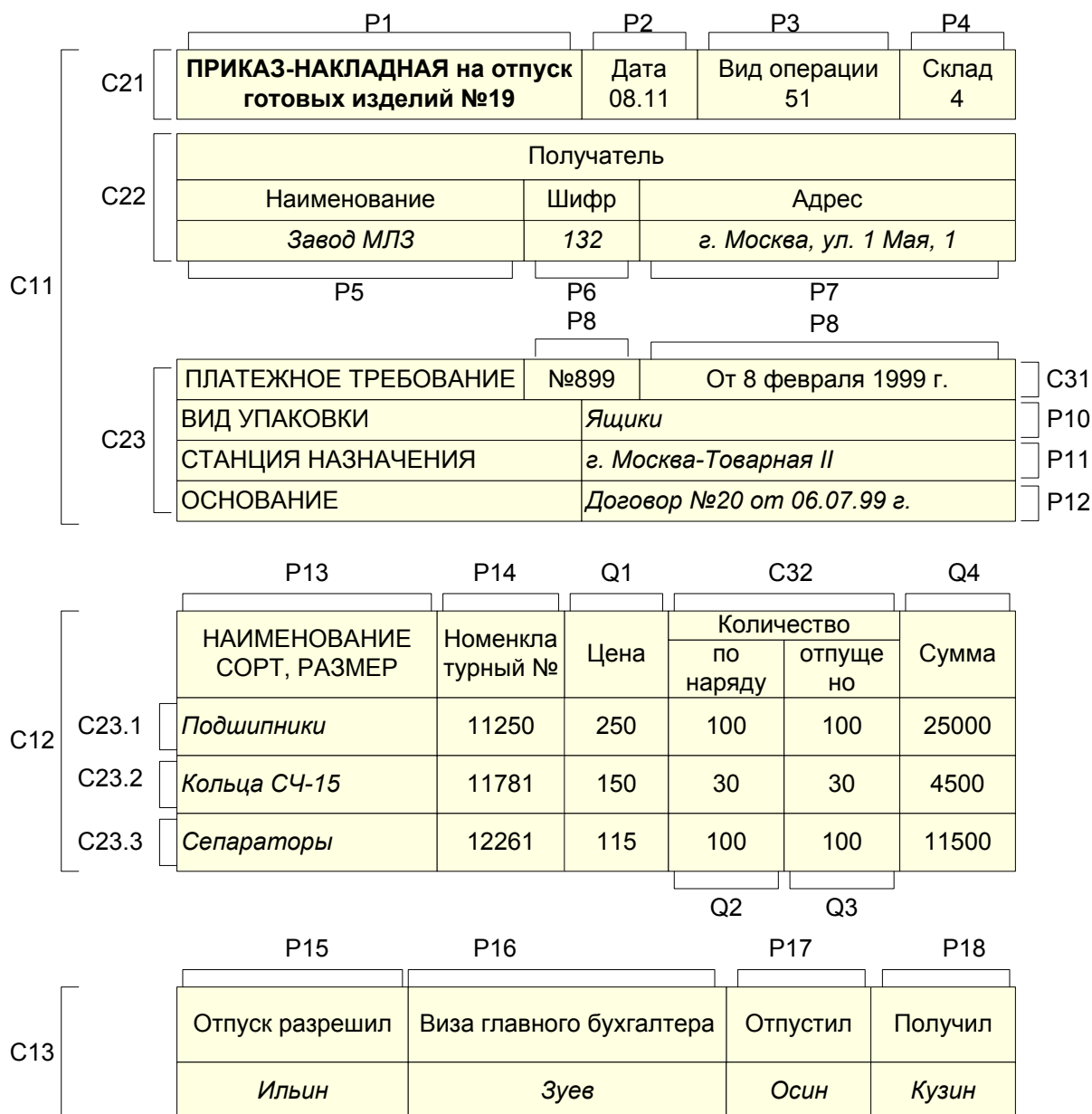


Рис.8. Структура документа

Составляющая *C21* также составная и включает четыре составляющих элемента следующего уровня: *C21.(P1,P2,P3,P4)*. Заметим, что все эти последние составляющие – элементарные в том смысле, что не подвергаются дальнейшему расчленению, и все они реквизиты-признаки.

Несколько сложнее структура информационной совокупности *C23* – общей части документа. Она включает элементы *C31* (данные о платежном требовании), *P10* (вид упаковки), *P11* (станция назначения) и *P12* (основание для сделки). Однако ее составляющий элемент *C31* в свою очередь является СЕИ, содержащий два элемента: номер платежного требования – *P8* и дату выписки платежного требования – *P9*. В целом состав *C23* может быть записан так:

$C23.(P8,P9),P10,P11,P12).$

Рассмотрим теперь предметную часть документа – информационную совокупность  $C12$ . Эта составная единица информации, поскольку содержит пять элементов:  $P13$  – наименование, сорт, размер;  $P14$  – номенклатурный номер;  $Q1$  – цена;  $C23$  – количество и  $Q4$  – сумма. Один из них –  $C23$  – состоит из двух элементов:  $Q2$  – количество по наряду и  $Q3$  – количество отпущенного. В целом для  $C12$  можно написать:  $C12.(P13,P14,Q1,C32.(Q2,Q3), Q4).$

Такая формулировка была бы справедливой для случая, если бы  $C12$  представляла лишь одно сообщение. Но как видно из Рис.8, в рамках совокупности  $C12$  приводятся три сообщения, а в общем случае и больше. Другими словами  $C12$  представляет собой массив. Для учета массива  $C12$  переформулируем предыдущую запись следующим образом

$$C12.(P13,P14,Q1,C32.(Q2,Q3),Q4).$$

$$C12.(P13,P14,Q1,C32.(Q2,Q3),Q4).$$

Оформительская часть документа состоит из четырех элементов, каждый из которых является реквизитом:  $C13.(P15,P16,P17,P18).$

Итак, мы рассмотрели структуру СЕИ одного приказа-накладной на отпуск готовых изделий. Однако этот первичный документ является массовым, и следует рассматривать массив таких документов. Структура СЕИ-массива  $S$  из  $M$  позиций может быть представлен следующей записью

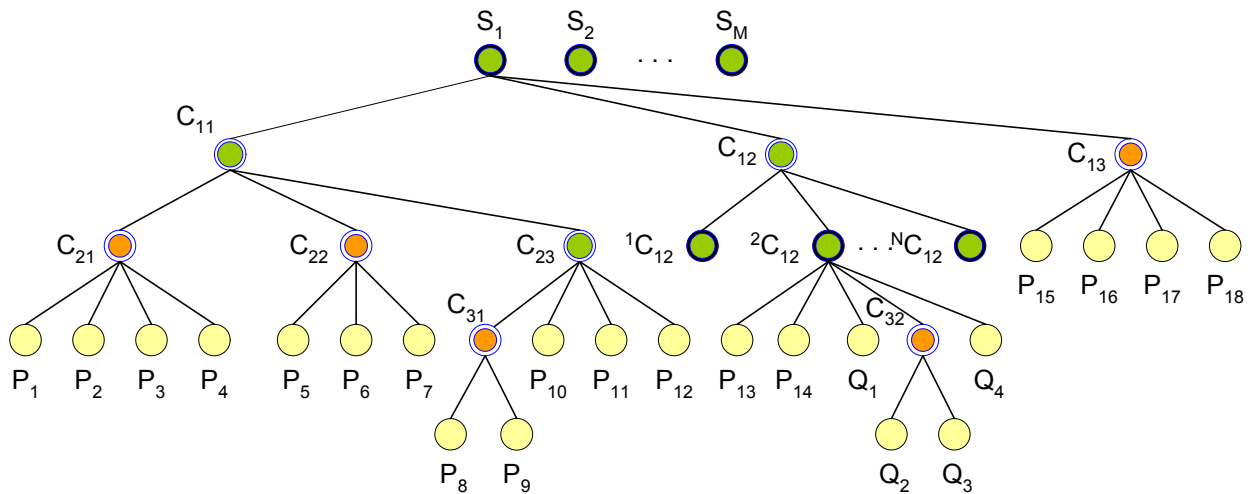


Рис.9. Графическое представление структуры СЕИ документа

$$\begin{aligned} S_{i=1}^M &.(C11.(C21.(P1,P2,P3)P4), \\ &C22.(P5,P6,P7), \\ &C23.(C31.(P8,P9),P10,P11,P12)), \\ C12_{j=1}^N &.(P13,P14,Q1,C32.(Q2,Q3),Q4), \\ &C13.(P15,P16,P17,P18)) \end{aligned}$$

Графическая интерпретация этой формулы дана на Рис.9, где терминальные узлы дерева представляют собой реквизиты, из которых реквизиты-признаки имеют идентификаторы, начинающиеся буквой Р, а реквизиты числового типа – идентификаторы с первой буквой Q. Остальные вершины это СЕИ.

### 1.3 ПОНЯТИЯ И ТЕРМИНЫ

А теперь вкратце сформулируем основные понятия и термины, использующиеся при составлении ИнфО ИнфС. Итак, при организации обработки информации в ходе управления производственным процессом используется следующая терминология:

#### 1.3.1 ИНФОРМАЦИОННАЯ СИСТЕМА И ИНФОРМАЦИОННЫЕ СОВОКУПНОСТИ

Экономическая информация является одной из разновидностей многих видов информации. Она содержит количественную и качественные характеристики производственных процессов. К ней относятся сообщения, которые циркулируют в экономической системе и которыми она обменивается с внешней средой. Объект отражения экономической информации – производственно-хозяйственная деятельность, а ее назначение – совершенствование процесса управления с целью улучшения технико-экономических показателей системы. Объектом информации являются различные производственные подразделения, изделия, детали, оборудование, работники, т.е. все элементы, сведения о которых передаются в систему. Источником информации является такой элемент, который может формировать или вырабатывать входное сообщение в систему. Им может быть должностное лицо, отдел, «датчик» и т.п.

Каждый объект управления не может функционировать без связанной с ним ИнфС. ИнфС может расчленяться на составляющие элементы, каждый из которых будет обладать свойствами этой системы. В результате полного расчленения получается система взаимосвязанных единиц информации, которые оказываются на разных уровнях по отношению друг к другу. Таким образом, *информационную совокупность* ИнфС можно рассматривать как многоуровневую систему.

Из информационных совокупностей состоит информация любого объема. Информационные совокупности подразделяются на номенклатуры и позиции.

*Номенклатура* – это универсальное множество данной информационной совокупности, например, перечень всех материалов, потребляемых данным предприятием. *Позиция* – это отдельный элемент номенклатуры, например, номенклатурный номер материала.

**Информационные ресурсы (ИР)** – это совокупность данных, имеющих смысловую нагрузку, отражающих всю производственно-хозяйственную деятельность предприятия и документально зафиксированных на носителях различного вида. По содержанию ИР представляют собой отображение реальных процессов производства, зафиксированных в проектно-конструкторской документации, плановых и отчетных документах, нормативных и инструктивных материалах и т.д.

**Реквизиты.** Информационные системы  $k$ -го уровня являются информационной совокупностью, представляющей информацию массива документов некоторой формы. Такая информационная совокупность принимается за единицу информации, которая называется составной единицей информации (СЕИ). В свою очередь, каждую СЕИ можно последовательно расчленять на информационные совокупности более низкого уровня до тех пор, пока дальнейшее расчленение станет практически не возможно из-за «неделимости». Путем такого последовательного деления СЕИ можно свести к реквизитам. *Реквизитом называется информационная совокупность, которая не поддается дальнейшему расчленению и характеризует отдельное свойство сущности.* Реквизиты являются минимальными единицами информации, на которых образуются все СЕИ.

**Показатели, основания и признаки.** Реквизиты подразделяются на основания и признаки. *Основания* характеризуют количественные свойства сущностей, полученные в результате вычислений или измерений. *Признаки* характеризуют качественные свойства сущностей (время и место событий, обстоятельства, при которых были получены числа-основания). Например, в наряде признаками являются: ФИО работника, номер чертежа детали, наименование операции, дата и т.д. Основаниями являются: расценка, количество деталей, сумма. При помощи признаков достигается индивидуализация сообщений.

При обработке информации над основаниями выполняются арифметические операции, а при помощи признаков осуществляются поиск информации, ее сортировка, выборка и т.д. (логические операции). Основание с относящимися к нему признаками, образующими информационную совокупность с минимальным составом, достаточного для образования документа, называется *показателем*:

*Основание +  $k$  признаков = показатель*

В конкретной системе показателей отражается вся деятельность предприятий. Элементы показателя, как и показатель в целом, можно рассматривать с двух точек зрения: по форме и по содержанию. Под *формой* понимается наименование граф и строк. Формы показателей имеют постоянный характер. Под *содержанием* понимается записанные в этих графах и строках конкретные числа или другие данные. Содержание показателей имеет переменный характер.

### 1.3.2 ИНФОРМАЦИЯ И СООБЩЕНИЕ

Источники формирования информации – это различные данные, создаваемые при подготовке производства, планировании и учете, которые насыщают информационную модель управления объектом (предприятием или организацией). В АС сырьем для производства информации служат данные. В результате обработки данные приобретают смысл, т.е. становятся информацией. С другой стороны информация раскрывает семантику данных. Традиционно данные всегда играли существенную роль в деятельности любого предприятия. Благодаря своевременному использованию информации, полученной на основе достоверных данных, принимаются правильные решения. В общем случае под информацией понимается определенная совокупность сведений об окружающем мире, событиях, чьей-либо деятельности; это форма связи между управляющими и управляемыми объектами, составляющими любую систему управления.

Форма представления информации – сообщение. *Сообщение* – это совокупность информации, достаточной для выработки какого-либо суждения о конкретном явлении, факте, процессе, и состоит из одного или нескольких реквизитов. Сообщение дает определенную количественную и качественную характеристику описываемого объекта. Исходные данные поступают на обработку сформированными в виде информационных сообщений, в которых имеется необходимый набор реквизитов – признаков, и оснований. При этом основание количественно характеризует описываемый объект, а признак функционально предназначен для его качественной характеристики. С этой точки зрения показатель представляет собой частный случай сообщения.

### 1.3.3 ДОКУМЕНТЫ И МАССИВЫ

Массив – набор однородных показателей или документов по одной форме. Это основная информационная совокупность, подвергающаяся обработке в информационных процедурах.

## **ГЛАВА 2 ВНЕМАШИННОЕ ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ АС**

Основной формой существования экономической информации является документация. Переработка регламентированных и нерегламентированных документов, их составление, заполнение, оформление, пересылка, считывание, сортировка занимает большую часть времени работников, занятых в сфере управления. С прогрессирующим усложнением характера и системы управления народным хозяйством быстро возрастает и число управленческих документов. В настоящее время уже не представляется возможным мириться с таким положением, когда требования переработки увеличивающегося числа документов удовлетворяются за счет роста численности управленческого персонала. Возникает актуальнейшая необходимость в выработке методов сокращения числа, унификации и стандартизации документов, основанных на формализованном описании документальных систем.

### **2.1 КЛАССИФИКАЦИЯ И КОДИРОВАНИЕ ИНФОРМАЦИИ**

#### **2.1.1 ОСНОВНЫЕ ПОНЯТИЯ КОДИРОВАНИЯ**

При машинной обработке информации появляется проблема записи ее на языке удобном для восприятия техническими средствами и подчиняющемся формальным правилам. Перевод информации на формальный язык осуществляется посредством кодирования. Понятие "код" можно определить как систему условных обозначений или сигналов.

В области машинной обработки технико-экономической информации различают два типа кодов: экономические и машинные. Под машинными кодами понимают такие, которые используются для управления машиной, для подачи ко-

манд и т. д., например: служебные коды машины. Под технико-экономическими кодами понимают все коды, используемые для предоставления технико-экономической информации, например, коды вида оплат, профессий, оборудования и т. д.

Термины и определения в области классификации и кодирования технико-экономической информации регламентированы ГОСТ 17369-85. Под *кодом* понимается знак или совокупность знаков, принятых для обозначения классификационной группировки и (или) объекта классификации. *Кодирование* – это образование и присвоение кода классификационной группировке и (или) объекту классификации. Совокупность методов и правил кодирования классификационных группировок и объектов классификации заданного множества называется *системой кодирования*.

Система знаков, принятых для образования кода, представляет собой алфавит кода. Число знаков в алфавите кода называется основанием кода. Позиция знака в коде является разрядом кода, а число знаков без учета пробелов представляет собой длину кода. Структура кода – условное обозначение состава и последовательности расположения знаков в коде.

Основными целями кодирования информации в АС являются:

- ✧ представление информации в виде, удобном для обработки на ЭВМ (например, возможность группировки информации, распознавание признаков, контроль признаков, ...);
- ✧ экономия места, требующегося для записи информации;
- ✧ дополнение информации, имеющейся на исходном языке (например, к коду комплектующего изделия добавляется вид приемки и т. д.);
- ✧ однозначная идентификация объектов кодирования и проведение различия между информационными совокупностями, которые трудно различимы в исходном языке (например, одна и та же деталь может различаться по месту изготовления).

При разработке кодов должны учитываться следующие основные требования:

- ✧ система кодирования должна соответствовать Единой системе классификации и кодирования (ЕСКК), соответствующим регламентирующим материалам и ГОСТ;
- ✧ с целью обеспечения наиболее компактного представления информации алфавит кода должен быть по возможности цифровым;
- ✧ основание кода должно быть по возможности минимальным;
- ✧ кодовое обозначение должно обеспечить резерв, т. е. возможность кодирования дополнительных объектов.

Различают следующие основные виды кодов: порядковые, серийные, позиционные (десятичные) и смешанные.

Порядковый код представляет собой такое обозначение позиций номенклатуры, которое соответствует их порядковым номерам после расположения в какой-либо последовательности. Порядковый код очень прост по построению, но

Табл. 1. Принцип построения серийного кода

<i>Подразделения предприятия</i>	<i>Кол-во разновидностей в группе цехов или подразделений</i>	<i>Выделяемая серия номеров (кодов)</i>	<i>В том числе запасные номера</i>
Основное производство	12	1001 – 1025	13
Вспомогательное производство	5	1026 – 1039	8
Отделы и службы	15	1040 – 1065	10
Склады	7	1066 - 1077	4

позволяет лишь отличить один объект от другого. По коду невозможно судить о свойствах объекта и производить какие-либо группировки по признакам с целью подведения итогов. При кодировании необходимо предусмотреть одинаковое количество разрядов в кодовом обозначении каждого объекта; при появлении новых объектов им присваивается очередной номер в конце списка, тем самым нарушается последовательность расположения позиций, если она была заранее определена. Учитывая недостатки порядковой системы кодирования, применять ее рекомендуется только для кодирования небольших массивов с одним используемым признаком.

Серийный код предполагает разделение всей номенклатуры на группы, для каждой из которых отводятся с учетом резерва серии порядковых номеров. Серийный код обычно применяется для объектов, имеющих небольшое количество признаков, по которым необходимо вести группировку.

Принципы построения серийного кода рассмотрим на примере, приведенном в таблице, в которой кодируются подразделения предприятия.

Преимуществом данного кода является сравнительная простота построения и возможность выделения признаков. В то же время ему присущи серьезные недостатки: невозможность автоматического получения итогов более чем по одному признаку, возникающему при декодировании, т. к. необходимо помнить, с какого номера начинается и каким кончается каждая группа.

Позиционный код предполагает проведение предварительной классификации объектов с целью выделения определенных признаков, которым отводится соответствующее количество разрядов в кодовом обозначении.

Упрощенная схема позиционного кода, основанная на пятиступенчатой классификации объектов кодирования представлены в Табл. 2. На схеме каждому признаку отводится один разряд в кодовом обозначении, следовательно, полное кодовое обозначение объекта будет пятизначным.

Рассмотрим построение пятизначного кода материалов на основе пятиступенчатой классификации. Первая ступень подразделяет материалы на классы, например: металлы (1), химикаты (2), нефтепродукты (3) и т. д. Вторая ступень вы-



Табл. 2. Упрощенная схема позиционного кода

<u>X</u>	<u>x</u>	<u>X</u>	<u>x</u>	<u>x</u>
1-я ступень (класс)	2-я ступень (подкласс)	3-я ступень (группа)	4-я ступень (подгруппа)	5-я ступень (вид)

деляет в каждом классе соответствующие подклассы, например: класс металлов, подразделяется на черные металлы (1), цветные металлы (2), сплавы (3). Третья ступень выделяет группы в каждом классе, например: для черных металлов: сырье (1), литье (2), прокат (3). Следующая ступень – деление группы на подгруппы – показана на примере проката черных металлов: лист (1), кружок (2), брусок (3), проволока (4). И последняя, пятая ступень делит каждую подгруппу на вид, например: размер листа стали 50×50 (1), диаметр проволоки 7мм (2) и т.д. Итак, кодовое обозначение 11342 соответствует материалу, имеющему следующие признаки: проволока проката черных металлов диаметром 7мм.

Огромным преимуществом данного метода кодирования является то, что он содержит подробную характеристику объекта, выделяет различные признаки и тем самым обеспечивает удобства при машинной обработке: группировку по различным признакам, подведение итогов, выборку и т.д. Недостатком кода является значительная длина кодового обозначения.

Смешанные коды объединяют в себе элементы различных видов кодов. Например: в кодовом обозначении для одних признаков используется позиционный код, для других – серийный.

Работы по проектированию кодов и созданию классификаторов являются весьма трудоемкими и важными при создании АС. Комплекс указанных работ включает следующие этапы:

- ✧ определение перечня позиций всей номенклатуры показателей, подлежащих кодированию;
- ✧ систематизация позиций, их классификация;
- ✧ выбор системы кодирования;
- ✧ построение кода каждой позиции данной номенклатуры (собственно кодирование);
- ✧ составление классификатора.

Классификатор – это официальный документ, представляющий систематизированный свод наименований и кодов классификационных группировок и (или) объектов классификации.

### 2.1.2 ОСНОВНЫЕ ПОНЯТИЯ КЛАССИФИКАЦИИ

При машинной обработке технико-экономической информации с целью выделения определенных признаков кодируемой номенклатуры приходится проводить систематизацию позиций и их классификацию. *Классификация – это разделение множества объектов на подмножества по их сходству или различию в со-*

*ответствии с принятыми методами или системой. Система классификации есть совокупность методов и правил классификации ее результат.* В результате проведения классификации заданное множество разбивается на подмножества. Подмножество объектов, полученное в результате классификации, представляет собой классификационную группировку. Свойство или характеристика объекта, по которому производится классификация, является признаком классификации.

К свойствам классификации предъявляются следующие требования:

- ✧ достаточная емкость, обеспечивающая все объекты классификации;
- ✧ экономически оправданная глубина, т.е. число ступеней классификации;
- ✧ гибкость, позволяющая расширить множества классифицируемых объектов, группировок и признаков, а также вносить необходимые изменения без нарушения структуры классификации;
- ✧ возможность сопряжения с другими классификаторами однородных объектов;
- ✧ простота ведения классификатора.

При классификации технико-экономической информации наиболее широко используют две системы классификации: иерархическую и фасетную.

Иерархическая система предполагает полное разделение множества объектов на подчиненные классификационные группировки. Она строится по следующему принципу: исходное множество объектов классификации делится сначала по некоторому выделенному признаку на крупные группировки, каждая группировка, в свою очередь, в соответствии с выбранным признаком делится на ряд последующих группировок, которые подразделяются на более мелкие, постепенно конкретизируя свойства объекта. Таким образом, классифицируемые понятия строятся в виде иерархического логического дерева путем последовательной конкретизации свойств объектов. В результате создается сложная многоуровневая иерархия понятий.

При построении иерархической системы классификации для отнесения конкретного объекта классификации на каждой ступени только к одной классификационной группировке необходимо соблюдать следующие правила:

- ✧ деление каждой группировки нужно проводить только по одному признаку;
- ✧ получаемые на каждой ступени классификационные группировки не должны повторяться;
- ✧ признаки, по которым проводится классификация, должны определяться в зависимости от решаемых задач.

Фасетная система классификации предполагает параллельное разделение множества объектов на независимые классификационные группировки. При этом исходное множество объектов классификации имеет некоторый набор признаков, сформированных в независимые параллельные фасеты. Рассмотрим данную систему на примере классификации приспособлений для металлорежущих станков. Классификацию проведем по трем независимым признакам: по специализации, способам зажима, виду множительного перемещения.

ПО СПЕЦИФИКАЦИИ	ПО СПОСОБАМ ЗАЖИМА	ПО ВИДУ ОТНОСИТЕЛЬ НОГО ПЕРЕМЕЩЕНИЯ
Специальные Универсальные	Механические Пневматические	Неподвижные С поступательным движением
Для групповой обработки	Гидравлические	Вращающиеся

При построении фасетной системы классификации необходимо соблюдать следующие правила:

- ✧ признаки, используемые в различных фасетах, не должны повторяться;
- ✧ из множества признаков, характеризующих объекты классификации, отбираются такие, которые обеспечивают решение конкретных задач.

### 2.1.3 ПРОЕКТИРОВАНИЕ КЛАССИФИКАТОРОВ

Все классификаторы, разрабатываемые и используемые в ИнфС, имеют эталонную и рабочую формы. *Эталонная форма классификатора* – это официальное издание классификатора на бумажном носителе, удобное для осуществления его ведения. *Рабочая форма классификатора* – это весь классификатор или его раздел, занесенный на машинный носитель и удобный для обработки информации. Весь процесс разработки системы классификаторов для ИнфС можно разбить на четыре этапа.

На первом этапе «*Разработка ТЗ на проектирование*» выполняются работы, связанные с определением состава, назначения и сферы действия классификаторов, используемых в системе.

*Перечень классификаторов* определяется на основе анализа реквизитного состава первичных и результатных документов и выделения из всей совокупности реквизитов-признаков.

Далее определяют *назначение классификаторов*. Каждый классификатор может быть предназначен для однозначной идентификации объекта, передачи информации по каналам связи или для поиска и логической обработки с целью получения и выдачи результатной информации.

По *сфере действия* выделяют следующие виды классификаторов: международные, общегосударственные, отраслевые и локальные. Международные классификаторы входят в состав Системы международных экономических стандартов и обязательны для передачи информации между организациями разных стран мирового сообщества. Общегосударственные классификаторы обязательны для организации процессов передачи и обработки информации между ИнфС государственного уровня внутри страны. Для выполнения процедур обработки информации внутри отрасли используют отраслевые классификаторы. В пределах отдельных организаций используют локальные классификаторы.

Кроме этого, на первом этапе определяется *состав исходных данных и*

**требований** к разрабатываемым классификаторам. К исходным данным, используемым в процессе проектирования классификаторов, относятся:

- ✧ состав задач, для которых разрабатывается классификатор;
- ✧ состав объектов классификации и мощность исходного множества;
- ✧ состав признаков классификации и число значений каждого признака;
- ✧ наименования отдельных группировок и объектов;
- ✧ динамика процесса изменяемости состава задач, объектов и признаков.

К требованиям, которым должны удовлетворять разрабатываемые классификаторы, относятся:

- ✧ полноту охвата объектов и признаков классификации каждым классификатором;
- ✧ согласованность признаков деления множеств объектов с алгоритмами обработки информации;
- ✧ взаимную однозначность наименований объектов и их кодовых обозначений;
- ✧ простоту кодирования и возможность автоматизации классификации и кодирования;
- ✧ возможность увязки с другими классификаторами и системами обозначений;
- ✧ эффективность использования классификаторов при обработке информации.

Содержанием второго этапа проектирования классификаторов является **«Разработка методических материалов проектирования»**, которая включает, прежде всего, разработку *основных критериев и принципов построения* каждого классификатора. К критериям построения классификатора относятся такие как:

- ✧ критерий отнесения того или иного объекта к конкретному классификационному множеству;
- ✧ степень охвата кодируемого множества объектов.

#### 2.1.4 РАЗДЕЛЬНАЯ КЛАССИФИКАЦИЯ И КОДИРОВАНИЕ

В практике разработки классификации задачи классификации и кодирования обычно решаются вместе и настолько тесно связаны, что их вообще не различают. При этом одна задача (обычно классификация) подчиняется другой и в результате ни одна не решается наилучшим образом. Между тем эти задачи существенно различаются по своим целям.

- ✧ Классификация преследует цель установления отношений между понятиями как отображениями объектов или групп объектов с общими свойствами, изучения структуры и упорядочения содержания данных.
- ✧ Кодирование же добивается однозначного обозначения элементов данных, которое имело бы возможно меньшую длину и требовало минимума затрат на перекодирование при обмене данными.

## 2.2 ДОКУМЕНТАЦИЯ КАК ЗНАКОВАЯ СИСТЕМА

Управленческий документ можно рассматривать как специфический знак языка общения между социальными структурами. Как и всякий знак, документ представляет собой единство определенной формы, регламентированного содержания и материального носителя. Основное содержание документа определяется предписанным набором показателей и реквизитов. Модификации содержания документа связаны в первую очередь с изменением значений показателей или с вариациями номенклатурных частей показателей, входящих в состав документа.

Первичные документы строятся на основе определенного множества исходных показателей. Документы каждого вышележащего уровня управления, как правило, строятся на основе документов нижележащего уровня. Эти специфические черты общения с помощью документов позволяют рассматривать документацию как некоторый язык, который может быть формализован на тех же основаниях, что и другие знаковые системы. (Привести аналогию с ЕЯ).

Через экономические показатели документ отображает состояние системы, в частности, регистрирует события, связанные с производством и распределением материальных благ, переработкой информации в системе или в среде. Характерной чертой большинства документов является наличие у них правовой значимости – достоверность данных, заключенных в документе юридически заверяется.

### 2.2.1 ПЕРВИЧНЫЕ ДОКУМЕНТЫ

Состояние объекта и его элементов регистрируется на соответствующих носителях информации – первичных документах. В соответствии с ГОСТ 6.10.2-80 первичный документ – это документ, содержащий исходную информацию. Проектируемые первичные документы должны содержать минимальный, но достаточный набор исходных данных, необходимых для получения максимальной информации, используемой для управления.

Проектирование первичных документов сводится к следующему:

- ✧ Выявляются реквизиты, намеченные для внесения в документы по каждой функциональной подсистеме. Состав реквизитов определяется постановкой задачи.
- ✧ Из общего состава необходимой информации по каждой подсистеме отбирается круг реквизитов для конкретной формы документа.
- ✧ Выбранные реквизиты подразделяются на две группы: переносимые на машинный носитель и непереносимые на машинный носитель.
- ✧ Осуществляется проектирование эскизов формы первичного документа и машинного носителя информации.
- ✧ Формы первичных документов уточняются, вычерчиваются на бланках стандартного размера.

Проектирование эскизов первичных документов ведется с учетом расположения реквизитов в соответствующих зонах. Любой бланк документа можно раз-

бить на зоны, предназначенные для записи определенных групп реквизитов, имеющих логическую или математическую зависимость, связанных одновременностью обработки или заполнения. Отдельные зоны документа могут быть построены по линейной, анкетной или табличной форме.

Наименование документа	Дата	Цех	Участок	Вид оплаты
	26.07	05	3	07

*Линейная форма* предусматривает для каждого реквизита две клетки: одну – для наименования, а другую – для записи данных. Во всех документах зона реквизитов обычно построена по линейной форме.

*Анкетная форма* предусматривает размещение реквизитов в вертикальной последовательности, друг за другом. При этом для наименования реквизита отводится левая часть строки, а для записи данных – правая.

*Табличная форма* предусматривает размещение реквизитов в виде таблицы с графами по вертикали и строками по горизонтали. Конкретные наименования и значения реквизитов проставляются в соответствующих строках и графах. По табличной форме построения многие зоны многострочных первичных документов.

Линейная анкетная и табличная формы в частном виде встречаются редко. В основном на практике получили распространение документы, сочетающие различные формы размещения реквизитов. При проектировании первичных документов учитываются преимущества и недостатки различных форм размещения реквизитов, и выбирается наилучшая для построения отдельных зон и документа в целом.

Первичные документы, проектируемые для использования в АС, должны быть унифицированы. *Унифицированные формы документов (УФД)* и бланки, применяемые для их оформления, следует строить на бумаге формата ряда А по ГОСТ 9327-60. При проектировании первичных документов необходимо учитывать следующие требования, предъявляемые к ним в связи с машинной обработкой и использованием в АС.

Реквизиты должны размещаться на бланке в определенной последовательности. Целесообразна такая последовательность: справочно-группировочные реквизиты-признаки, постоянные для группы документов, справочно-группировочные переменные реквизиты-признаки и реквизиты основания.

Из первичных документов необходимо исключить постоянную информацию (справочную, расценочную, нормативную, плановую), а также производственные показатели, которые могут быть получены в результате обработки. Постоянная информация в АС размещается во внешней памяти ЭВМ (в виде соответствующих разделов БД) и используется по мере надобности для решения комплекса задач. Выполнение этих требований существенно упрощает формы документов, их заполнение и обработку, а также повышает надежность информации.

Документы, отражающие однородные операции должны иметь определенный состав реквизитов и одинаковую, строго определенную последовательность их размещения на бланке. Унифицированные формы документов должны созда-

ваться на всех уровнях управления. Это обеспечит информационное единство АС различных уровней. В связи созданием общегосударственной автоматизированной системы управления (ОГАС) и отраслевых АС большое значение приобретает типизация первичных документов на общегосударственном и отраслевом уровнях.

## 2.2.2 ПОНЯТИЕ УНИФИЦИРОВАННОЙ СИСТЕМЫ ДОКУМЕНТАЦИИ

Существующие формы документов не отвечают требованиям, предъявляемым современной организационной и вычислительной техникой, применяемой для их обработки, не обеспечивает сопоставимость показателей, обрабатываемых в различных системах народного хозяйства. Поэтому создание унифицированной документации является задачей, которая решается на государственном уровне. ГОСТ 6.10.1-80 устанавливает требования к УСД общегосударственного значения, унифицированным формам документов различных уровней управления народным хозяйством, к составу и структуре реквизитов и показателей.

УСД общегосударственного назначения и УФД различных уровней являются составными частями ИО АС, объединенных в единую общегосударственную систему сбора и обработки информации для управления. Они должны быть пригодны для использования в условиях традиционных (ручных) методов сбора и обработки информации. УФД общегосударственного значения, отраслевые, субъектов федерации, предприятий должны быть включены соответственно в общегосударственные классификаторы управленческой документации, отраслевые классификаторы, классификаторы субъектов федерации и предприятий.

Регистрация УФД производится следующим образом:

- ✧ для форм общегосударственного назначения порядок регистрации определяется Госстандартом;
- ✧ для отраслевых и региональных – министерствами и департаментами;
- ✧ для УСД предприятий – руководством предприятий.

Регистрационные номера УФД общегосударственного назначения совпадает с кодом формы по общегосударственному классификатору управленческой документации. При установлении регистрационных номеров региональных и отраслевых УФД руководствуются соответствующими нормативными и методическими материалами Госстандарта. Регистрационные номера УСД предприятий устанавливают лица, ответственные за регистрацию этих форм на предприятиях.

К построению унифицированных документов (УД) в соответствии с ГОСТ 6.10.1-80 предъявляются следующие требования

УФД должны быть удобны для восприятия человеком и приспособлены для автоматизированной обработки:

- ✧ документы должны строиться на бумаге форматов А3, А4, А5 по ГОСТ 9327-60;
- ✧ основой построения УСД являются формуляры-образцы, содержащие следующие основные части: заголовочную, содержательную, оформляющую;

- ✧ в пределах указанных частей в формуляре-образце устанавливаются зоны для записи реквизитов, размеры и взаимное расположение зон. Установленная последовательность расположения реквизитов в пределах зон является обязательной;
- ✧ границы нанесения зон и реквизитов в формах документов определяются с применением конструкционной сетки, которая строится в соответствии с ГОСТ 8274-78.

Для информационной увязки АС различных уровней необходимо обеспечить сопоставимость реквизитов-признаков и показателей по наименованию и содержанию. Содержательная часть документа оформляется в табличной, линейной или анкетной форме в виде сплошного текста, заголовочная и оформляющая – в линейной форме.

Реквизиты, подлежащие переносу на машинные носители, должны быть сгруппированы и выделены в специальные зоны. Для кодирования реквизитов, содержащих информацию, которая подлежит машинной обработке, должны применяться коды соответствующих классификаторов технико-экономической информации (общегосударственные, отраслевые, предприятий). Места для записи кодов отводятся в зонах, предназначенных для обозначаемых реквизитов (например, ОКП – общегосударственный классификатор промышленной продукции). Реквизиты заголовочной и оформляющих частей документов предназначены для идентификации и придания юридической силы документу.

Обязательные реквизиты, идентифицирующие документ, включают:

- ✧ код формы;
- ✧ наименование формы;
- ✧ наименование должности лица, подписавшего документ;
- ✧ личная подпись и ее расшифровка;
- ✧ дата подписания или утверждения документа.

При построении содержательной части документа в виде таблицы следует соблюдать следующие требования:

- ✧ в таблицу включаются заголовки граф и строк (реквизиты-признаки), графы и строки;
- ✧ графы таблицы нумеруются;
- ✧ реквизиты-признаки могут быть постоянными и переменными.

Оформление текстов документов должны соответствовать требованиям ГОСТ 6.38-72 и ГОСТ 6.39-72. Сокращения слов в текстах производится в соответствии с правилами орфографии и пунктуации и сокращениями, принятыми в общегосударственных классификаторах технико-экономической информации. ГОСТ 6.10.2-83 устанавливает термины и определения основных понятий УСД.



## 2.3 НОРМАТИВНО-СПРАВОЧНАЯ ИНФОРМАЦИЯ

### 2.3.1 ОСНОВНЫЕ ПОНЯТИЯ

Нормативно-справочная информация (НСИ) занимает важное место в ИО АС и представляет систему научно и технически обоснованных нормативов, характеризующих количественную меру различных элементов процесса производства. НСИ предприятия является ядром всего ИО АС, и именно здесь в первую очередь закладывается научная обоснованность и комплексность управления предприятием. Объем нормативной информации составляет 60-70% общего объема информации, используемой в процессе управления предприятием. Нормативные массивы, зафиксированные на машинных носителях, составляют нормативную базу АС, т.е. нормативное хозяйство АС представляет собой размещенный на магнитных носителях систематизированный комплекс сведений и данных, обеспечивающих решение задач по управлению производством.

*Норматив* – это количественная и качественная характеристика объема управления, показывающая, каковы нормальные его параметры при данном уровне развития производства. Применительно к предприятию нормативы должны выражать зависимость отдельных показателей его деятельности от технических, экономико-организационных и общественно-политических условий.

*Норма* – это первичный количественный норматив, полученный внутри предприятия путем технического расчета, или установленный извне в качестве исходного. Нормы представляют собой входную информацию низкого уровня и выражаются в единицах физических величин.

В базовое ИО входят справочники. *Справочник* – это перечень данных, однозначно характеризующих состояние объекта на определенный период времени и позволяющих выделить этот объект из множества других.

### 2.3.2 ОРГАНИЗАЦИЯ И ВЕДЕНИЕ ФОНДА НСИ

Фонд НСИ является составной частью ИО АСП. Он представляет собой совокупность норм, нормативов, а так же условно-постоянных справочных и условных показателей, отражающих относительно устойчивые свойства объектов и явлений производственно-хозяйственной деятельности предприятия. Эти сведения представлены на различных машинных носителях информации и являются исходными данными при решении комплекса задач АСП.

В процессе создания и ведения фонда НСИ можно выделить следующие основные операции:

- ✧ создание системы обозначения технико-экономической информации (классификаторы ТЭИ, заводские классификаторы);
- ✧ заполнение первичной документации, содержащей НСИ;
- ✧ подготовка и сдача первичных документов в службу информативного хозяйства ИВЦ предприятия;

- ✧ прием и контроль первичных документов;
- ✧ создание массивов НСИ на машинных носителях;
- ✧ поддержание массивов НСИ в рабочем состоянии путем систематического внесения изменений (актуализация данных).

Нормативное хозяйство предприятия должно быть таким, чтобы обеспечить решение всего комплекса задач АСП, нормы должны быть едиными для всех нормативно-плановых расчетов, проводимых на предприятии. Использование единых норм для всех расчетов, выполняемых с помощью ЭВМ, обеспечивается в результате централизованного формирования и обновления в памяти ЭВМ массивов НСИ.

### ГЛАВА 3 ВНУТРИМАШИННАЯ ФАКТОГРАФИЧЕСКАЯ ИНФОРМАЦИОННАЯ БАЗА

В соответствии с ГОСТ 24.205-80 внутримашинной информационной базой (ИБ) называют совокупность всех данных на машинных носителях, сгруппированных по определенному признаку. В состав внутримашинной ИБ могут выделяться: база данных (фактографическая), документальная база данных, база знаний. Несмотря на то, что принципы хранения данных в системах обработки фактографической и документальной (текстовой) информации схожи, алгоритмы обработки в них заметно различаются. Поэтому в зависимости от характера информационных ресурсов, которыми оперируют такие системы, принято различать на два крупных класса – документальные и фактографические.

*Фактографическая БД* – это часть внутримашинной ИБ, представляющая совокупность массивов и выделенная для реализации определенных функций АС. Она оперирует фактическими сведениями, представленными в виде организованных совокупностей формализованных записей данных. Центральное функциональное звено фактографических ИнфС – *системы управления БД (СУБД)*. Фактографические системы используются не только для реализации справочных функций, но и для решения задач обработки данных. Под обработкой данных понимается специальный класс решаемых машиной задач, связанных с вводом, хранением, сортировкой, отбором и группировкой записей данных однородной структуры. В большинстве случаев эти задачи предусматривают предоставление пользователям итоговых результатов обработки в виде отчетов табличной формы.

Среди фактографических систем важное место занимают два класса: системы операционной обработки данных и системы, ориентированные на анализ данных и поддержку принятия решений. Первые рассчитаны на быстрое обслуживание относительно простых запросов большого числа пользователей. Сфера применения таких систем – это системы платежей, резервирования мест в гостини-

цах, поездах, самолетах, банковские и биржевые системы. Логическая единица функционирования систем операционной обработки данных – транзакция. Транзакция – это некоторое законченное, с точки зрения пользователя, действие над БД. В современной литературе для обозначения систем операционной обработки данных часто используют термин *OLTP* (*On-Line Transaction Processing – операционная обработка транзакций*).

Другой класс ИнфС – системы поддержки принятия решений (*СППР*) или аналитические системы. Эти системы ориентированы на выполнение более сложных запросов, **требующих синтаксической**

## 3.1 СИСТЕМЫ OLTP

### 3.1.1 ОБРАБОТКА ТРАНЗАКЦИЙ В OLTP-СИСТЕМАХ

Транзакцией называют неделимую с позиций воздействия на БД последовательность операций манипулирования данными. Транзакция может состоять из операций чтения, удаления, вставки, модификации данных. В OLTP – системах транзакция реализует некоторое осмысленное, с точки зрения пользователя, действие, например, перевод денег со счета на счет в платежной системе банка, резервирование места в поезде системой оформления железнодорожных билетов.

Чтобы использование механизмов обработки транзакций позволило обеспечить целостность данных и изолированность пользователей, транзакция должна обладать четырьмя основными свойствами: атомарности, согласованности, изолированности, долговечности.

Свойство атомарности означает, что транзакция должна выполняться как единая операция доступа к БД. Она должна быть выполнена полностью, либо не выполнена совсем. Другими словами, должны быть выполнены все операции манипулирования данными, которые входят в транзакцию, либо, если по каким-то причинам выполнение части операций не возможно, ни одна из операций не должна выполняться.

Свойство согласованности гарантирует взаимную целостность данных, т.е. выполнение ограничений целостности БД после окончания обработки транзакции. Следует отметить, что БД может обладать такими ограничениями целостности, которые сложно не нарушать, выполняя только один оператор ее изменения. Например, если в отношении А храниться число кортежей отношения В, то добавив новый кортеж в отношение В, не нарушив ограничений ценности невозможно. Поэтому такое нарушение целостности внутри транзакции допускается, но к моменту ее завершения БД должна быть в целостном состоянии. Несоблюдение этого условия приводит к отмене всех операций транзакции.

В многопользовательских системах с одной БД одновременно могут работать несколько пользователей или ПП. Поскольку каждая транзакция может изменять данные разделяемые данные, данные могут временно находиться в несогласованном состоянии. Доступ к этим данным другим транзакциям должен быть

запрещен, пока изменения не будут завершены. Свойство изолированности гарантирует, что они будут выполняться отдельно друг от друга.

Свойство долговечности означает, что если транзакция выполнена успешно, то произведенные ею изменения в данных не будут потеряны ни при каких обстоятельствах.

Результатом выполнения транзакции может быть ее фиксация или откат. Фиксация транзакции – это действие, обеспечивающее запись в БД всех изменений, которые были произведены в процессе ее выполнения. До того, как транзакция зафиксирована, возможна отмена всех сделанных изменений и возврат БД в то состояние, в котором она была до начала выполнения транзакции. Фиксация транзакции означает, что все результаты ее выполнения становятся видимыми другими транзакциям.

Если нормальное завершение транзакции невозможно, например, нарушены ограничения целостности БД, пользователь выдал специальную команду, происходит откат транзакции. БД возвращается в исходное состояние, все изменения аннулируются.

Механизм корректного отката и фиксации транзакций основан на использовании журнала транзакций. Для того, чтобы иметь возможность сделать откат, СУБД должна сохранять все изменения, которые транзакция внесла в БД. Однако необходимости каждый раз сохранять всю информацию БД, нет. Реляционные операции изменяют строки отношений БД, поэтому, чтобы обеспечить возможность отката, СУБД должна хранить те строки, которые были модифицированы. При выполнении любой операции, изменяющей БД, СУБД автоматически сохраняет в журнале транзакций состояние модифицируемых строк до операции и после нее. Только после этого изменения вносятся в БД. Если по окончании обработки транзакция фиксируется, то в журнале делается соответствующая отметка. Если же производится откат транзакции, то СУБД по журналу восстанавливает те строки отношений, которые были модифицированы, отменяя, таким образом, все изменения.

Для того, чтобы транзакцией, как единой логической единицей, СУБД должна уметь определять ее границы, т.е. первую и последнюю входящие в нее операции. Стандарт языка SQL предусматривает следующий принцип выделения транзакций как некоторой закономерной последовательности действий. Предполагается, что транзакция начинается с первого SQL – оператора, вводимого пользователем или содержащегося в ПП. Все следующие далее операторы составляют тело транзакции. Тело транзакции завершается SQL – операторами COMMIT WORK или ROLLBACK WORK. Выполнение транзакции заканчивается также при завершении программы, которая сгенерировала транзакцию. Транзакция фиксируется, если ее тело оканчивается оператором COMMIT WORK, либо в случае конца приложения, сгенерировавшее транзакцию, завершилось с ошибкой.

Применение транзакций – эффективный механизм организации многопользовательского доступа к БД. Однако при реализации этого механизма СУБД приходится сталкиваться с целым рядом проблем.

Во-первых, необходимо избежать потери изменений БД в ситуации, когда программы читают одни и те же данные, изменяют их, и пытаются записать результат на прежнее место. В БД могут быть сохранены изменения, выполненные только одной программой, результаты работы всех остальных программ будут потеряны.

Во-вторых, требуется исключить возможность чтения незафиксированных изменений. Это может произойти в случае, когда одна транзакция вносит изменения, они тут же считываются в другой транзакции, но затем другая транзакция прерывается оператором ROLLBACK WORK.

Чтобы избежать этих проблем, должна быть использована специальная дисциплина совместной обработки (сериализации) транзакций. В ее основе лежат следующие принципы:

Транзакция не может получить доступ к незафиксированным данным, в которых произведены изменения, но эти изменения еще не зафиксированы.

Результат совместного выполнения транзакций должен быть эквивалентен результату их последовательного выполнения. Т.е. если две транзакции выполняются параллельно, то предполагается, что результат такой же, как если бы сначала выполнялась первая, а затем вторая транзакция, или наоборот.

В современных СУБД сериализация транзакций реализуется через механизм блокировок. На время выполнения транзакции СУБД блокирует часть БД, к которой транзакция обращается. Блокировка сохраняется до момента фиксации транзакции. Если в процессе параллельной обработки другой транзакции делается попытка обратиться к заблокированным данным, обработка транзакции приостанавливается и возобновляется только после завершения транзакции, заблокировавшей данные и снятия блокировки.

### 3.1.2 ВЫПОЛНЕНИЕ ТРАНЗАКЦИЙ В РАСПРЕДЕЛЕННЫХ БД

Современные ИнфС работают с распределенными БД, поэтому в одной транзакции могут модифицироваться отношения, физически хранящиеся на удаленных БД. Транзакция, обновляющая данные на нескольких узлах сети, называется распределенной. Если транзакция работает с БД, расположенной на одном узле, то ее называют локальной. Т.о., логически распределенная транзакция состоит из нескольких локальных.

С точки зрения пользователя, локальные и распределенные транзакции должны обрабатываться одинаково, т.е. СУБД должна организовать процесс выполнения распределенной транзакции так, чтобы все локальные транзакции, входящие в нее, синхронно фиксировались на затрагиваемых ими узлах распределенной системы. Однако распределенная транзакция должна фиксироваться только в том случае, когда зафиксированы все локальные транзакции, ее составляющие. Если прерывается хоть одна из локальных транзакций, должна быть прервана и распределенная транзакция.

Для практической реализации этих требований в СУБД используют меха-

низм двухстадийной фиксации транзакций. При его использовании фиксация распределенных транзакций осуществляется в два этапа.

На первой стадии сервер БД, фиксирующий распределенную транзакцию, посылает команду «приготовиться к фиксации» всем узлам сети (серверам БД), задействованным для выполнения локальных транзакций, инициированных локальной транзакцией. Все серверы локальных БД должны в ответ сообщить, что они готовы к фиксации. Если хотя бы от одного из серверов ответ не получен, то сервер распределенной БД производит откат локальных транзакций на всех узлах.

Вторая стадия начинается, когда все локальные СУБД готовы к фиксации. Сервер, обращающий распределенную транзакцию, заканчивает ее фиксацию, посылая команду «зафиксировать транзакцию» всем локальным серверам.

Описанный подход выполнения транзакций в распределенных системах не единственно возможный. Альтернатива ему – тиражирование данных. Эта технология предполагает отказ от распределенных данных – во всех узлах вычислительной системы должна находиться своя копия БД. Средства тиражирования автоматически поддерживают согласованное состояние информации в нескольких БД посредством копирования изменений, вносимых в любую из них. Любая транзакция в такой системе выполняется локально, поэтому нет необходимости в сложной процедуре фиксации.

Узкое место такого подхода – обеспечение тождественности данных в узлах сети. Процесс переноса изменений исходной БД в базы, принадлежащие различным узлам распределенной системы, принято называть тиражированием данных. Функции тиражирования данных выполняет специальный модуль СУБД – сервер тиражирования данных (репликатор). При любых изменениях в тиражируемых данных репликатор копирует их на все остальные узлы системы. Схема тиражирования может быть построена на полном обновлении содержимого таблицы на удаленных серверах или же обновление только изменившихся записей. Процесс тиражирования данных скрыт от ПП пользователей, репликатор автоматически поддерживает БД в согласованном состоянии.

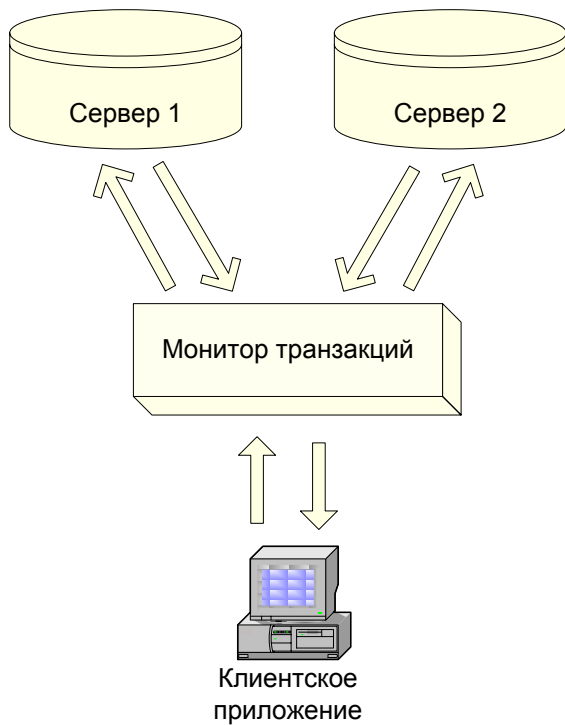
### 3.1.3 МОНИТОРЫ ТРАНЗАКЦИЙ

С ростом сложности распределенных вычислительных систем возникают проблемы эффективного использования их ресурсов. Для решения этих проблем в состав распределенных OLPT-систем вводят дополнительный компонент – монитор транзакций (*TRM – transaction processing monitor*).

- ✧ динамическое распределение запросов в системе (выравнивание нагрузки);
- ✧ оптимизация числа выполняющихся серверных приложений.

Если в системе функционируют несколько серверов, представляет одинаковый сервис, например, доступ к БД, то для оптимизации пропускной способности системы 9числа обрабатываемых запросов в единицу времени0 необходимо добиться их сбалансированной нагрузки, т.е. необходимо обеспечить, чтобы на каждый из них поступало примерно равное число пользовательских запросов. При

распределении запросов может учитываться также удаленность серверов, их готовность, содержимое запроса. Реализуется функция выравнивания нагрузки следующим образом (см.рис.)



*Рис.10. Упрощенная схема работы монитора транзакций.*

Запрос клиентского приложения поступает на монитор транзакций, который, действуя от имени клиентского приложения, определяет получателя этого запроса. Для этого он обращается к динамической маршрутной таблице, по которой определяет систему, предоставляющую соответствующий сервис. Если нужный сервис предлагают несколько систем, то в зависимости от используемого алгоритма маршрутизации выбирается одна из них, после чего ей перенаправляется запрос клиентского приложения. Результат выполнения запроса через монитор транзакций перенаправляется приложению, пославшему запрос. Клиентские приложения не знают о том, какой системе будут направлены их запросы, предлагается ли нужный им сервис одним или несколькими серверами, расположен ли нужный сервер локально или удаленно, - в любом случае их запрос будет обработан оптимально.

Скорость обработки транзакций напрямую зависит от числа запущенных серверных приложений. Чем больше приложений одновременно обслуживает запросы, тем выше пропускная способность системы. Это увеличение наиболее заметно на многопроцессорных системах, где каждое приложение может работать на отдельном процессоре.

В идеале для эффективного использования системных ресурсов нужно по мере необходимости увеличивать или уменьшать число серверных приложений в зависимости от числа обрабатываемых запросов. Для решения этой задачи мониторы транзакций периодически измеряют отношение числа запросов в очереди к числу работающих серверных приложений. Если это отношение превышает некоторое максимальное пороговое значение, то запускается дополнительная копия серверного приложения. Если это отношение падает ниже минимального порогового значения, то одна из копий завершается.



## 3.2 СИСТЕМЫ, ОРИЕНТИРОВАННЫЕ НА АНАЛИЗ ДАННЫХ

### 3.2.1 ХРАНИЛИЩА ДАННЫХ

К середине 80-х в развитых странах завершился первый этап оснащения бизнеса и органов государственного управления средствами ВТ. Военные ведомства и крупные корпорации установили распределенные ВС, состоящие из мощных мейнфреймов. С появлением ПК ЭВМ стали доступны множеству средних фирм и организаций. Исторически эти системы в первую очередь реализовывали потребности в операционной обработке данных – они обслуживали информационные архивы, телефонные сети, системы резервирования билетов и др. Использование мощных средств ВТ позволило накапливать большие объемы информации: документы, сведения о банковских операциях, клиентах, предоставленных услугах. Однако период хранения этой информации был относительно невелик – сохранялись только данные за текущий календарный период.

Вскоре возникло понимание, что сбор данных – не самоцель, и накопленные информационные массивы могут быть полезны. Системы операционной обработки способны выполнять тривиальный анализ данных – вычислять максимальные, минимальные и средние значения атрибутов. Но из накопленных данных можно почерпнуть намного более глубокие сведения как о функционировании организации, которая обслуживается ИнфС, так и о сфере ее деятельности. В информационных массивах можно попытаться выявить скрытые, на первый взгляд, закономерности и вывести из них правила, которым подчиняется ПрдО ИнфС. Впоследствии эти правила можно использовать для стратегического планирования, принятия решений и прогнозирования их последствий

Осознание пользы накапливаемой информации и возможности использования ее для решения аналитических задач привело к появлению нового класса вычислительных систем – систем поддержки принятия решений (СППР), ориентированных на аналитическую обработку данных. Под СППР понимают человеко-машинный вычислительный комплекс, ориентированный на анализ данных и обеспечивающий получение информации, необходимой для разработки решений в сфере управления. Следует заметить, что аналитические системы существовали и ранее, но именно возможность обработки больших объемов накапливаемых данных дала новый толчок их развитию.

Для получения интересующей информации лица, принимающие решения, или аналитики обращаются к СППР с запросами. Эти запросы в большинстве случаев более сложные, чем те, которые применяются в системах оперативной обработки данных. Поэтому аналитический запрос сложно сформулировать в терминах языка SQL, и для получения информации применяют специализированные языки, ориентированные на аналитическую обработку данных. Также для выполнения аналитических запросов могут быть использованы приложения, написанные специально для решения тех или иных аналитических задач.

Для того, чтобы можно было извлекать полезную информацию из данных,

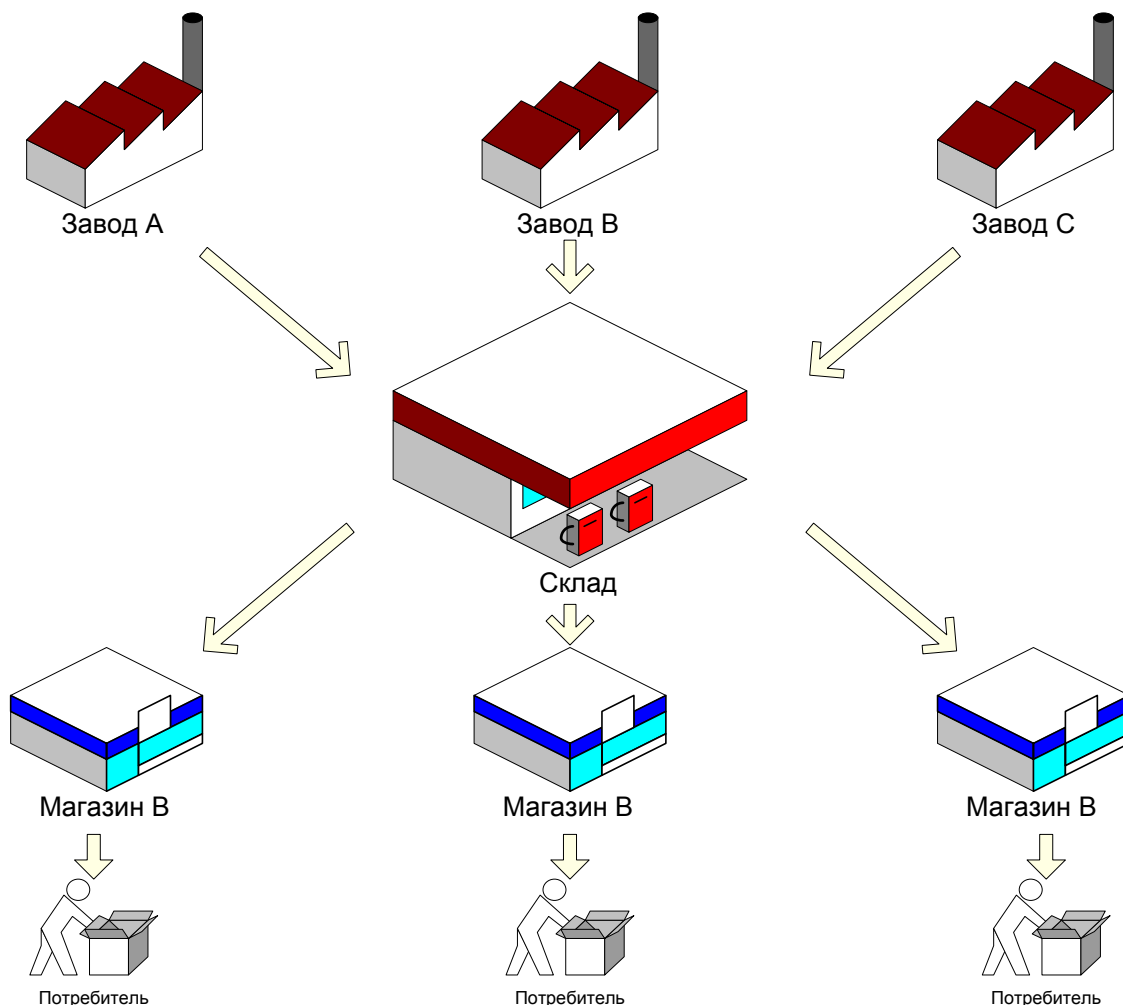
они должны быть организованы особым, отличным от принятого в OLTP-системах образом. Связано это со следующими факторами.

Во-первых, для выполнения аналитических запросов необходима обработка больших информационных массивов. Чем выше степень нормализации данных, и чем больше в ней таблиц, тем медленнее выполняется анализ. Происходит это потому, что увеличивается число операций соединения отношений. В системах обработки транзакций нормализация таблиц баз данных позволяет устранить избыточность данных, уменьшив тем самым объем действий, необходимых при обновлении информации. В аналитических системах данные практически не обновляются – в системе происходит лишь их накопление и чтение. Поэтому проблема нормализации в них не актуальна, как в системах обработки транзакций.

Во-вторых, выполнение некоторых аналитических запросов, например, анализ тенденций и прогнозирование, требует хронологической упорядоченности данных. Реляционная модель не предполагает существования порядка записей в таблице.

В-третьих, данные, используемые для целей анализа, как правило, отличаются от данных систем обработки транзакций. При обслуживании аналитических запросов чаще используются не детальные, а обобщенные (агрегированные) данные. Так, например, для прогнозирования объема продаж сети универмагов будет излишним иметь информацию о каждой сделанной покупке. Для этого достаточно знать значения прогнозируемой величины за несколько предыдущих лет.

Принципы, лежащие в основе СППР, позволяют эффективно обрабатывать транзакции, поэтому данные, применяемые для анализа, стали выделять в отдельные базы данных. Впоследствии эти базы данных стали называть хранилищами данных (ХД).



*Рис.11. Схема процесса промышленного производства и реализации продукции.*

Концепция ХД – это концепция подготовки данных для последующего анализа. Она предполагает выполнение следующих положений:

- ✧ интеграции и согласования данных из различных источников: традиционных OLTP, информации из внутренних и внешних по отношению к организации электронных архивов;
- ✧ разделения наборов данных, использованных системами обработки транзакций и СППР.

Рассмотрим схему формирования СППР, основанной на концепции ХД, проведя аналогию с процессом производства и реализации промышленной продукции.

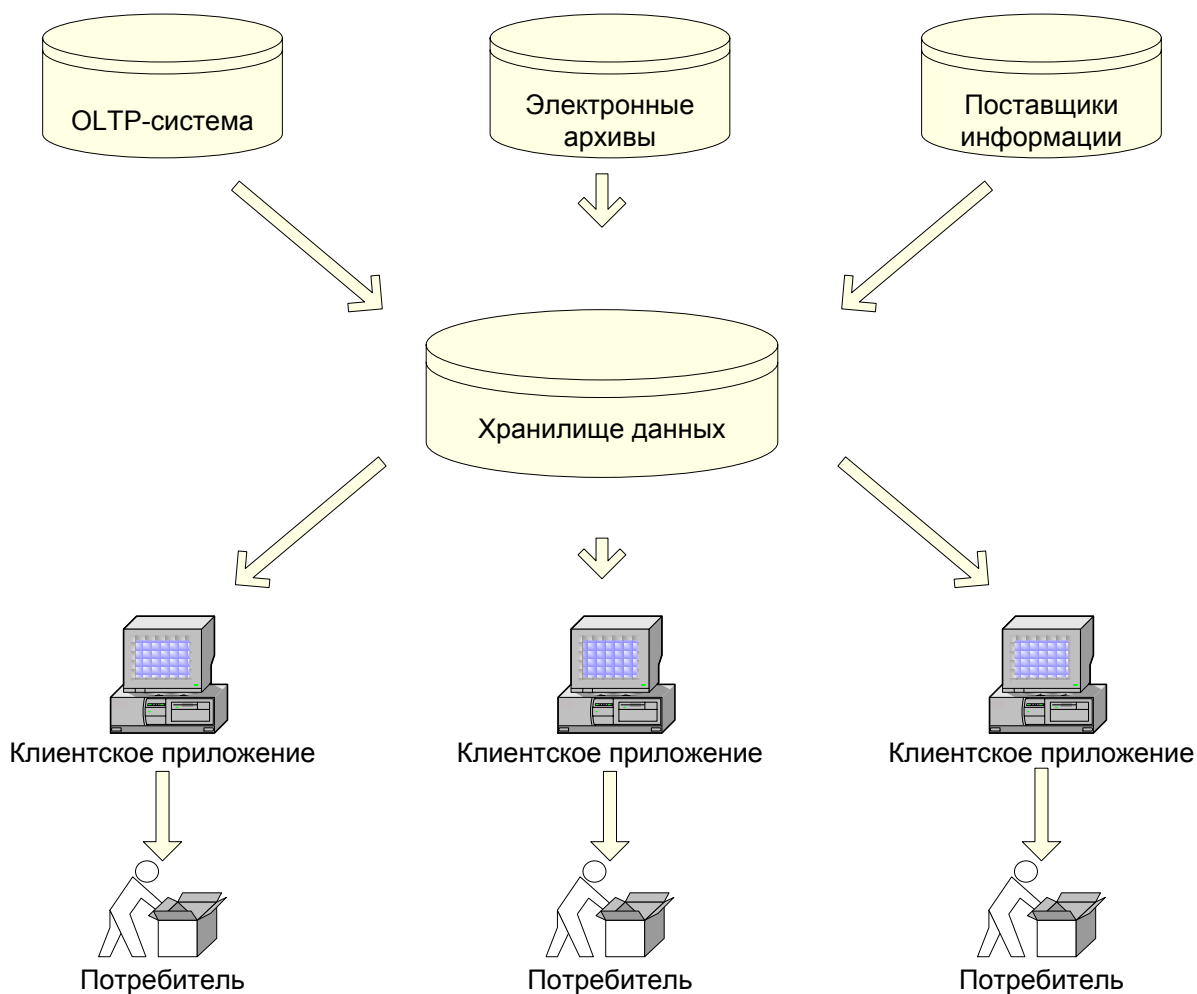


Рис.12. Упрощенная логическая схема аналитической системы.

Производство и реализация товаров имеет много общего с анализом данных: на предприятии из сырья получается готовая продукция, которая затем доставляется потребителю; в процессе анализа из накопленных данных добывается и предоставляемая полезная специалистам информация для разработки решений.

Упрощенный процесс производства и реализации промышленной продукции может быть описан следующим образом (см. рис.).

Любая продукция, прежде чем быть доставленной потребителю, должна быть изготовлена. Этим занимаются заводы. Промышленная продукция отправляется на склад, откуда поступает в магазины. Именно там она находит своего потребителя. Подобная схема обработки и снабжения справедлива и для аналитической системы (см. рис.).

Исходные данные для анализа производятся OLTP-системами, поступают из электронных архивов и от поставщиков информации. Эти источники слабо связаны между собой, поэтому и данные, которые они предоставляют, имеют различную структуру и форматы представления. Необходимо произвести согласование данных различных источников, чтобы ими удобно было оперировать при ана-

лизе. Это подразумевает приведение их к единому формату, устранение дублирующихся и некорректных значений.

Подготовленные данные загружаются в хранилище. Пользователи-

<i>Свойство</i>	<i>OLTP</i>	<i>СППР</i>
Назначение данных	Оперативный поиск, несложные виды обработки	Аналитическая обработка, прогнозирование, моделирование
Уровень агрегации данных	Детализированные	Агрегированные
Период хранения данных	От нескольких месяцев до одного года	От нескольких лет до десятков лет
Частота обновления	Высокая частота, обновление маленькими порциями	Низкая частота, обновление большими порциями

аналитики осуществляют доступ к нему через клиентское приложение

Эти приложения могут осуществлять трансляцию запросов потребителей информации, либо производить аналитическую обработку данных ХД. В отличие от OLTP в СППР, использующих концепцию ХД, критерии поиска и состав выдаваемой в виде отчета информации не фиксируются при ее разработке, пользователи оперируют основным заранее не регламентированными запросами. Использование концепции ХД в СППР преследует следующие цели:

- ✧ своевременное обеспечение аналитиков всей информацией, необходимой для выработки решений;
- ✧ создание единой модели данных организации;
- ✧ создание интегрированного источника данных, предоставляющего удобный доступ к разнородной информации.

Отличия СППР и OLTP приведены в таблице.

В последнее время сформировался новый класс СППР – системы оперативной аналитической обработки (OnLine Analysis Processing – OLAP ). Под OLAP системой принято понимать СППР, основанную на концепции ХД и обеспечивающую малое время выполнения аналитических запросов.

К числу основных задач, которые требуют решения при создании ХД, относятся:

- ✧ выбор оптимальной структуры хранения данных с точки зрения обеспечения приемлемого времени отклика на аналитические запросы и требуемого объема памяти;
- ✧ первоначальное заполнение и последующее пополнение ХД данными;
- ✧ обеспечение удобства доступа пользователей к данным.

### 3.2.2 МОДЕЛИ ДАННЫХ, ИСПОЛЬЗУЕМЫЕ ДЛЯ ПОСТРОЕНИЯ ХРАНИЛИЩ

Задачи, решаемые OLTP и СППР, существенно различаются, поэтому их БД

тоже построены на разных принципах. Критерием эффективности для OLTP систем служит число транзакций, которые они способны выписать в единицу времени. Для СППР важнее скорость выполнения сложных запросов и прозрачность структуры хранения информации для потребителей. Важная особенность СППР на основе ХД состоит в том, что загрузка данных выполняется сравнительно редко, но большими порциями, поэтому в таких системах не предусматриваются развитые средства обеспечения целостности, восстановления, устранения взаимных блокировок. Это не только существенно облегчает и упрощает сами средства реализации, но и значительно снижает внутренние накладные расходы при доступе к информации и, следовательно, повышает производительность анализа.

В настоящее время существуют два подхода к построению ХД:

- ✧ подход, основанный на испытании многомерной модели БД (Multidimensional OLAP- MOLAP);
- ✧ подход, использующий реляционную модель БД (Relational OLAP – ROLAP).

Прежде, чем рассказать о каждом из них, попытаемся разобраться, какие данные могут храниться в ХД, и как они могут быть представлены. Чаще всего там содержатся сведения о значении некоторых параметров, характеризующих ПрдО в определенные моменты или за определенные промежутки времени. Пусть, например, требуется создать ХД, накапливающее информацию об изменении социально-экономической обстановки в России. Эта обстановка характеризуется многими параметрами, в числе которых:

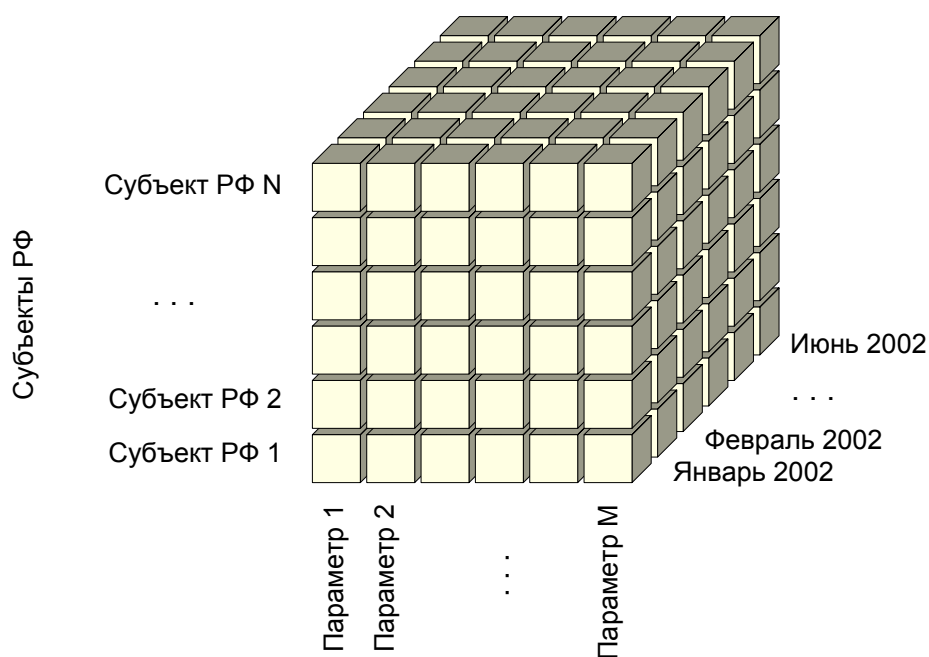
- ✧ объем промышленного производства;
- ✧ индекс потребительских цен;
- ✧ уровень занятости населения и др.

Госкомстат России собирает их значения для различных субъектов федерации ежемесячно, поквартально или за год. В ХД должны попадать факты вида: НАЗВАНИЕ ПАРАМЕТРА в СУБЪЕКТЕ РФ в МОМЕНТ ВРЕМЕНИ был равен {ЗНАЧЕНИЕ}.

Например, индекс потребительских цен в г. Москве в декабре 1998 г. был равен 101%. В рассмотренном примере значение связано с точкой в трехмерном пространстве (N, S, T) с измерениями:

- ✧ N – название параметра;
- ✧ S – субъект федерации;
- ✧ T – момент времени.

Число возможных параметров, субъектов РФ, а также рассматриваемых моментов времени, конечно, поэтому всевозможные значения можно представить в виде гиперкуба. В этом гиперкубе каждое значение находится в строго определенной ячейке, что значительно упрощает обращение к ней. Представление данных в виде гиперкуба более наглядно, чем совокупность нормализованных таблиц, оно понятно не только администратору БД, но и рядовым сотрудникам. Гиперкуб может быть реализован в рамках реляционной модели или существовать как отдельная БД специальной многомерной структуры.



*Рис.13. Представление данных в виде гиперкуба: A – значение «Параметра N» для «Субъекта РФ1» в январе 2002 года.*

**Многомерная модель ХД.** Многомерная модель БД появилась довольно давно, однако в силу присущих ей ограничений применение получила лишь в последнее время. При использовании этой модели данные хранятся не в виде плоских таблиц, как в РБД, а в виде гиперкубов – упорядоченных многомерных массивов. Конечно, такой подход требует большого объема памяти для хранения данных, при его испытании сложно модифицировать структуру данных. Например, добавление еще одного измерения требует полной перестройки куба. Однако многомерные СУБД обеспечивают более быстрый по сравнению с реляционными системами поиск и чтение данных, избавляют от необходимости многократно соединять таблицы. Среднее время ответа на сложный аналитический запрос при использовании многомерных СУБД обычно в 10-100 раз меньше, чем в случае РСУБД с нормализованной структурой.

Основные понятия многомерной модели – измерение и значение (ячейка). Измерение – это множество, образующее одно из граней гиперкуба (аналог домена в реляционной модели). Измерения играют роль индексов, используемых для идентификации конкретных значений в ячейках гиперкуба. Значения – это подвергаемые анализу количественные или качественные данные, которые находятся в ячейках гиперкуба.

В многомерной модели вводятся следующие основные операции манипулирования измерениями: сечение, вращение, детализация, свертка.

При выполнении операции сечения формируется подмножество гиперкуба, в котором значение одного или более измерений фиксировано. Например, если на [рис. 3.3](#) зафиксировать значение измерения «Время» равным «Январь 1991 года», то мы получим двухмерную таблицу с информацией о значениях всех пара-

метров для всех субъектов федерации в январе 1991 года.

Операция вращения изменяет порядок представления измерений. Она обычно применяется к двумерным таблицам, обеспечивая представление их в более удобной для восприятия форме. Если в исходной таблице по горизонтали были расположены субъекты РФ, а по вертикали параметры социально-экономической сферы, то после операции вращения параметры будут размещены по горизонтали, а названия субъектов РФ – по вертикали.

Для выполнения операций свертка и детализации должна существовать иерархия значений измерения, т.е. некоторая подчиненность одних значений другим. Например, 12 месяцев образуют год, субъекты РФ образуют регионы. При выполнении операции свертки одно из значений измерения заменяется значением более высокого уровня иерархии. Например, аналитик, узнав значения параметров для января 1991 года, желает получить их значения за весь 1991 год. Чтобы это сделать, необходимо выполнить операцию свертки. Операция детализации – это операция, обратная свертке. Она обеспечивает переход от обобщенных к детализированным данным.

Основное назначение СУБД, поддерживающих многомерную модель, - реализация систем ориентированных на аналитическую обработку. Многомерные СУБД лучше других справляются с задачами выполнения сложных нерегламентированных запросов.

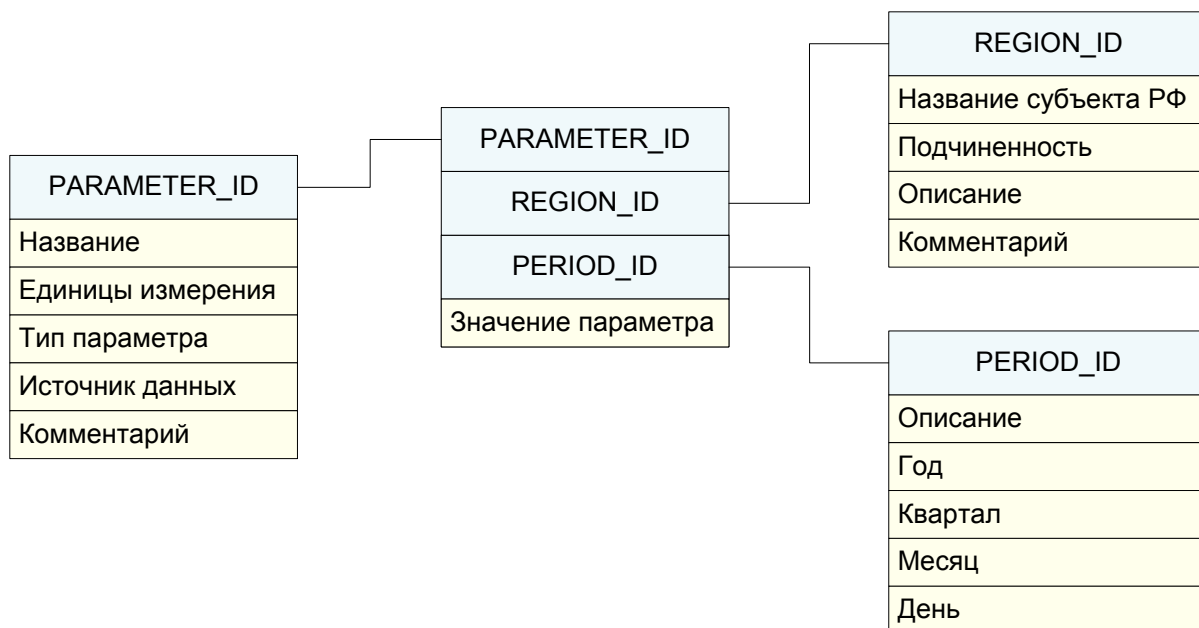
Однако у многомерных БД имеются серьезные недостатки, сдерживающие их применение. Многомерные СУБД, по сравнению с реляционными, неэффективно используют память. В многомерной СУБД заранее резервируется место для всех значений, даже если часть из них заведомо будет отсутствовать. Другой недостаток состоит в том, что выбор высокого уровня детализации при реализации гиперкуба может очень сильно увеличить размер многомерной БД.

**Реляционная модель ХД.** Основой при построении ХД может служить и традиционная реляционная модель. В этом случае гиперкуб эмулирует СУБД на логическом уровне. В отличие от многомерных СУБД, реляционные СУБД способны хранить огромные объемы данных, однако они проигрывают в скорости выполнения аналитических запросов. При использовании РСУБД для организации ХД данные организуются специальным образом. Чаще всего используется так называемая радиальная схема. Другое ее название – «звезда». В этой схеме используются два типа таблицы: таблица фактов (фактологическая таблица) и несколько справочных таблиц (таблицы измерений).

В таблице фактов обычно содержатся данные, наиболее интенсивно используемые для анализа. Если проводить аналогию с многомерной моделью, то запись фактологической таблицы соответствует ячейке гиперкуба. В справочной таблице перечислены возможные значения одного из измерений гиперкуба. Каждое измерение описывается своей собственной справочной таблицей. Фактологическая таблица индексируется по сложному ключу, скомпонованному из индивидуальных ключей справочных таблиц. Это обеспечивает связь справочных таблиц с фактологической по ключевым атрибутам. В качестве примера на рис.3.4 приве-



дена упрощенная схема структуры ХД, используемого для накопления информации из рассмотренного ранее примера (см. рис.3.3).



*Рис.14. Пример БД с радиально связанными таблицами (схема «звезда»): ключевые атрибуты выделены.*

В реальных системах количество строк в фактологической таблице может составлять 10 и 100 миллионов. Число справочных таблиц обычно не превышает двух десятков. Для увеличения производительности анализа в фактологической таблице могут храниться не только детализированные, но и предварительно вычисленные данные.

Если БД включает большое число измерений, можно использовать схему «снежинка». В этой схеме атрибуты справочных таблиц могут быть детализированы в дополнительных справочных таблицах (см. рис. 3.5).

Для сокращения времени, требуемого для получения отклика от аналитической системы, можно использовать некоторые специальные средства. В состав мощных реляционных СУБД обычно входят оптимизаторы запросов. При создании ХД на основе РСУБД их наличие приобретает особую важность. Оптимизаторы анализируют запрос и определяют лучшую, с позиций некоторого критерия, последовательность операций обращения к БД для его выполнения.

**Комбинация многомерного и реляционного подхода: киоски данных.** Каждая из описанных моделей имеет как преимущества, так и недостатки. Многомерная модель позволяет проводить быстрый анализ данных, но не позволяет хранить большие объемы информации. Реляционная модель, напротив, практически не имеет ограничений по объему накапливаемых данных, однако СУБД на ее основе не обеспечивает такой скорости выполнения аналитических запросов, как МСУБД. Нельзя ли совместить эти два подхода так, чтобы скрыть их недостатки

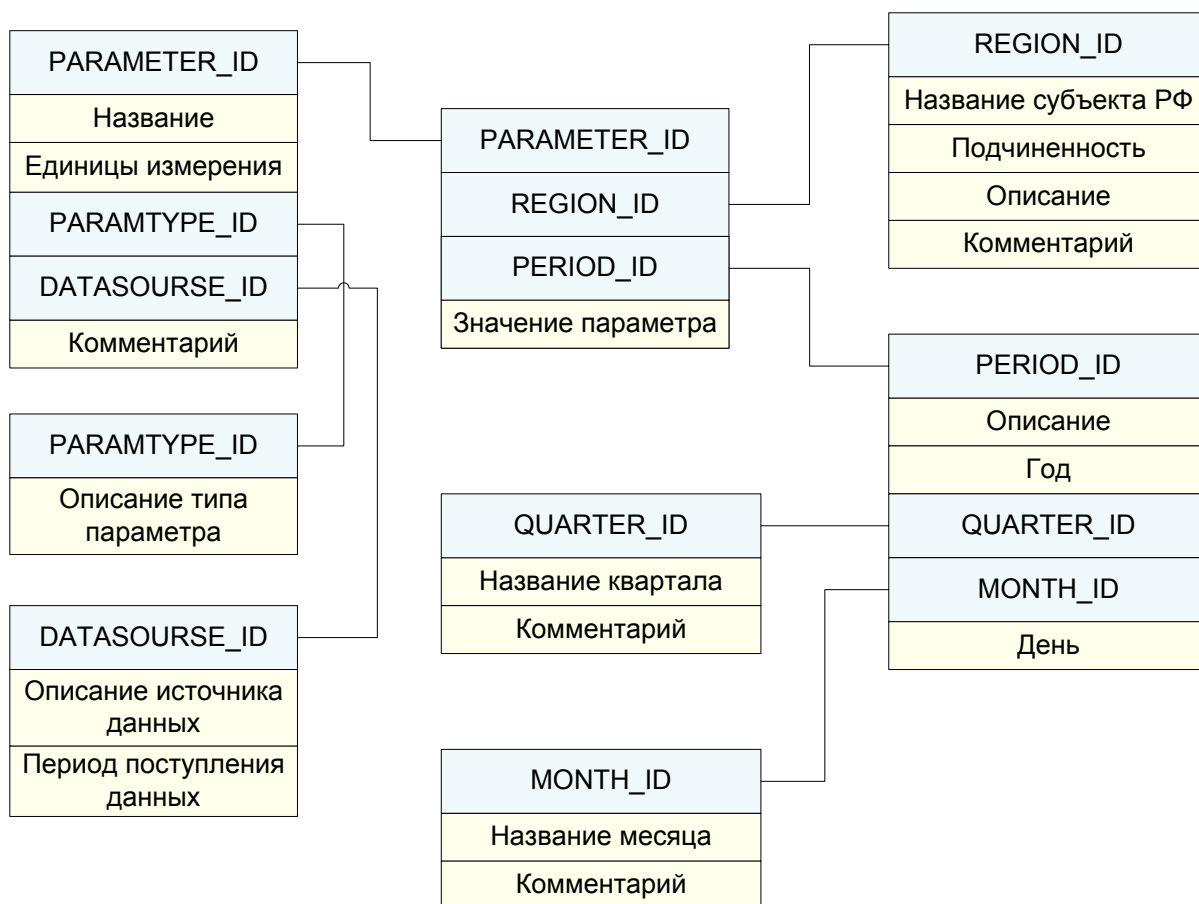


Рис.15. Пример БД со схемой «снежинка»

и сделать более заметными их достоинства?

Ситуация, когда для анализа необходима вся информация, находящаяся в ХД, возникает довольно редко. Обычно каждый аналитик обслуживает одно из направлений деятельности организации, поэтому в первую очередь ему необходимы данные, характеризующие именно это направление. Реальный объем этих данных не превосходит ограничений, присущих многомерным СУБД. Возникает идея выделить данные, которые реально нужны конкретным аналитическим приложениям, в отдельный набор. Такой набор мог бы быть реализован в многомерной БД. Источником данных для него должно быть центральное хранилище организации.

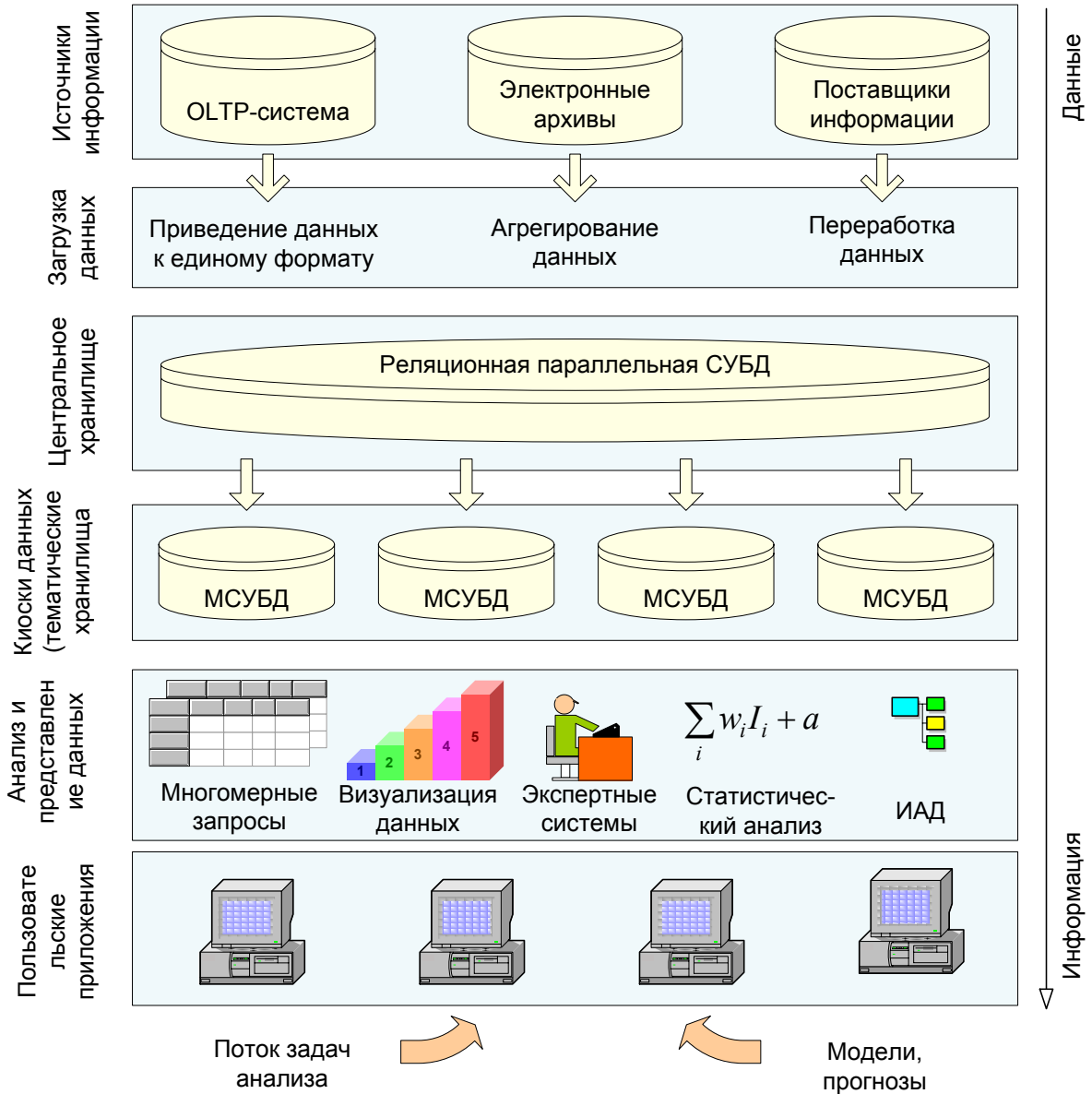


Рис.16. Логическая схема СППР, использующей ХД и киоски данных

Если проводить аналогии с производством и реализацией продукции, то многомерные БД играют роль мелких складов. В концепции ХД их принято именовать киосками данных. Киоск данных – это специализированное тематическое ХД, обслуживающее одно из направлений деятельности организации. Логическая схема СППР, использующей центральное ХД организации и киоски данных аналитических отделов, представлена на рис. 3.6.

Такая схема позволяет эффективно использовать возможности РСУБД по хранению огромных объемов информации и способность МСУБД обеспечивать высокую скорость выполнения аналитических запросов.

## ГЛАВА 4 ВНУТРИМАШИННАЯ ДОКУМЕНТАЛЬНАЯ ИНФОРМАЦИОННАЯ БАЗА

Классические модели в теории БД изначально ориентировались на организацию хранения и обработки детально структурированных данных. Чаще всего эти данные представляли собой числовые значения, описывающие те или иные характеристики информационных объектов. Однако на практике оказалось, что чаще информация представлена не в виде структурированных массивов данных, а в виде простых текстовых документов. Вследствие этого документальные БД сразу выделялись в особый тип баз данных. Исторически сложилось так, что за системами, ориентированными на работу с текстовыми документами укоренилось название информационно-поисковые системы (ИПС). Хотя, если быть точнее, их следует называть документальными ИПС (ДИПС), поскольку традиционные СУБД также являются ИПС, только фактографическими (ФИПС). В настоящее время ведутся интенсивные исследования систем выборки и моделирования документов, основанных на методах теории вероятностей, лингвистики и понимания ЕЯ.

В ИПС документы могут быть представлены (смоделированы) прямо либо косвенно. При прямом представлении документ хранится в памяти в обычной форме, а при косвенном представлении используются различные способы *индексирования*. По индексу можно получить адрес или идентификатор документа, который возможно, хранится в памяти. Как при прямом, так и при косвенном представлении документы можно хранить в виде полного текста или в усеченном виде. Например, из документов могут быть удалены незначимые слова, а оставшиеся могут быть приведены к некоторой основной форме.

- ✧ В одном варианте поиск документов основывается на точном совпадении ключевых слов с использованием поиска в строке для прямого представления и различных методов обработки индексов – для косвенного.

- ✧ В другом варианте поиск основывается на вычислении сходства между документом и запросом или документом или группой документов.

Для поиска в строке используются различные схемы сравнения, например, компараторы, управляемые конечными автоматами и обрабатывающие поток данных. Простейшие запросы могут сводиться к совпадению с ключевыми словами, более сложные запросы состоят из логической комбинации простых запросов и содержат указание контекста, возможно, с заданием относительного расстояния между ключевыми словами.

В косвенном представлении и при кластеризации (классификации) документов мы используем ключевые слова или термины, выбираемые в соответствии с некоторой автоматической или полуавтоматической схемой индексирования. Ключевые слова и документы можно объединять в группы (кластеры). Группы ключевых слов используются для составления тезаурусов (словарей синонимов или ключевых слов), а группы документов – для разбиения очень больших документальных БД с целью их более эффективной обработки.

*Документальная БД* представляется совокупностью документов на машинном носителе, объединенных по единству *ПрдО*, функций управления. Документальные системы служат для работы с документами на ЕЯ – монографиями, публикациями в периодике, текстами законодательных актов. Наиболее распространенный тип документальных систем – информационно-поисковые системы (*ИПС*), предназначенные для накопления и поиска по различным критериям документов на ЕЯ.

#### 4.1 НАЗНАЧЕНИЕ И ОСНОВНЫЕ ПОНЯТИЯ

В отличие от традиционных БД, ориентированных на полное и точное представление данных достаточно простой смысловой структуры, документальные БД ориентированы на частичное, приближенное представление данных, имеющих значительно более сложную смысловую структуру, представленных на входе в форме текста.

Информационный поиск в системе проводится на основе поступившего от потребителя запроса на отыскание необходимой ему информации. потребность человека в определенной информации в процессе его практической деятельности носит название информационной потребности. Под действием получаемой информации информационная потребность людей постоянно меняется и трансформируется. Однако информационная потребность может быть представлена в виде некоторой последовательности частных ее значений в фиксированные моменты времени. Такое частное значение информационной потребности потребителя в определенные моменты времени, выраженное на естественном языке (ЕЯ), представляет собой информационный запрос, с которым пользователь обращается к системе.

В СУБД мы формулируем запрос и отыскиваем удовлетворяющие этому запросу кортежи в одном или нескольких отношениях. В ИПС мы отыскиваем на-

бор документов, связанных с интересующей нас темой или имеющих к ней отношение. Критерий поиска основывается на отношении, называемом сходством, и постепенно уточняется до достижения требуемой цели. В то время как в СУБД мы имеем тождество между атрибутами запроса и атрибутами экземпляров записей, в ИПС обнаруженный документ может соответствовать атрибутам поиска лишь частично. Рассмотрим примеры запросов в СУБД и ИПС:

### *Запрос в СУБД*

```
SELECT NAME, AGE  
FROM EMPLOYEE  
WHERE SALARY > 30000  
AND DEPT = ENGINEERING
```

### *Запрос в ИПС*

*Найти все документы, которые содержат в одном предложении слова ДОКУМЕНТ и КЛАСТЕРИЗАЦИЯ или КЛАССИФИКАЦИЯ*

В запросе к СУБД отношение `EMPLOYEE` и атрибуты запроса находятся во взаимно однозначном соответствии и кортежи из отношения `EMPLOYEE` в ответе удовлетворяют запросу в точности – это кортежи тех сотрудников, которые работают в конструкторском отделе и получают более 30000 долларов в год. В запросе к ИПС ответом может быть последовательность документов, содержащих искомые слова. Однако запрос может быть неправильно сформирован потребителем, и не отражать его истинной информационной потребности в момент обращения к системе. Таким образом, при проведении информационного поиска в ИПС фактически рассматривается не информационная потребность пользователя, а только информационный запрос, в ответ на который выдаются те или иные документы системы. Следовательно, реакцию системы необходимо рассматривать не только по отношению к информационной потребности, но и по отношению к информационному запросу.

Для выражения данных отношений в теории ИПС введены два фундаментальных понятия: пертинентность и релевантность.

Под *пертинентностью* понимается соответствие смыслового содержания документа информационной потребности пользователя. Документы, содержание которых удовлетворяет информационной потребности пользователя, называют пертинентными.

*Релевантность* представляет собой соответствие содержания документа к информационному запросу в том виде, в каком он сформулирован, а документы, содержание которых отвечает запросу пользователя, носят название релевантных.

Автоматизация процесса информационного поиска потребовала формализации представления основного смыслового содержания информационного запроса и документов в виде соответственно поискового предписания (ПП) и поисковых образов документов (ПОД). Для записи ПП и ПОД применяются специальные языки, называемые информационно-поисковыми. В процессе проведения информационного поиска в ИПС определяется степень соответствия содержания документов и запроса пользователя путем сопоставления ПОД с ПП. А на основе такого сопоставления принимается решение о выдаче документа (он признается реле-

вантным) или о его невыдаче (он признается нерелевантным).

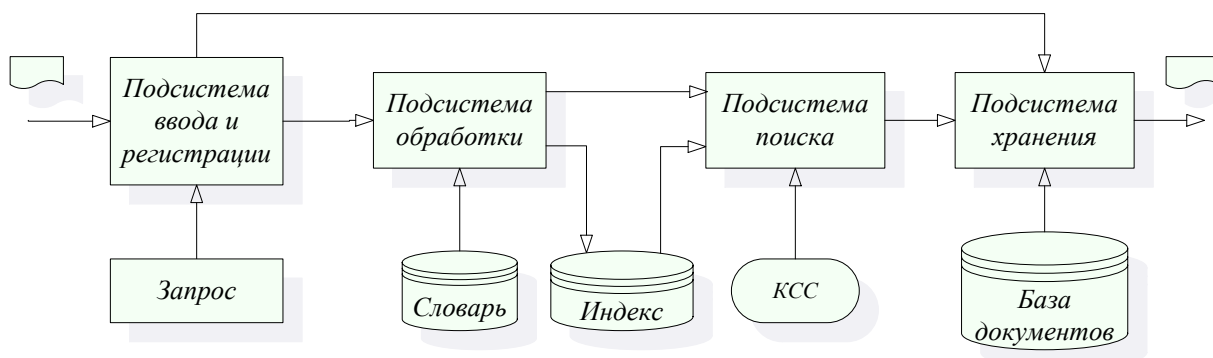
Решение о выдаче или невыдаче документа в ответ на запрос принимается на основе некоторого набора правил, по которому данной ИПС определяется степень смысловой близости между ПОД и ПП. Такой набор правил получил название *критерия смыслового соответствия* (КСС). Критерий может задаваться явно или неявно. На самом деле КСС базируется не на ранее введенном понятии релевантности, а на понятии формальной релевантности – соответствия содержания ПОД и ПП. Фактическая релевантность, понимаемая как смысловое соответствие содержания документа информационному запросу, может быть установлена только человеком в процессе осмысления содержания документа и запроса.

Рис.17. Общая функциональная структура ИПС

## 4.2 ОБЩАЯ ФУНКЦИОНАЛЬНАЯ СТРУКТУРА ДОКУМЕНТАЛЬНЫХ ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМ

В состав типичной ИПС входят четыре основные подсистемы:

- ✧ подсистема ввода и регистрации;
- ✧ подсистема обработки;
- ✧ подсистема поиска;
- ✧ подсистема хранения.



Текстовые документы, поступающие на вход системы, могут быть представлены как в бумажном, так и в электронном виде. Поэтому *подсистема ввода и регистрации* решает следующие основные задачи:

- ✧ создание электронных копий документов (например, сканирование с последующим распознаванием текста или ввод с клавиатуры);
- ✧ обеспечение подключения к каналам доставки электронных документов;
- ✧ распознавание, а при необходимости и преобразование формата электронных документов;
- ✧ присвоение электронным документам уникальных идентификаторов (регистрация).

Все поступившие документы без внесения в них каких-либо изменений направляются в *подсистему хранения* для сохранения в базе документов. База документов может представлять собой простую совокупность файлов, распределенную по каталогам жесткого диска. Однако такой тип представления базы документов характеризуется следующими недостатками:

- ✧ неэффективным использованием дискового пространства;
- ✧ низкой скоростью доступа при большом количестве файлов.

Поэтому для хранения документов применяют средства сжатия и быстрого поиска информации. В этом случае подсистема хранения представляет собой совокупность стандартных или специализированных средств архивации, СУБД и т.п., обеспечивающих возможность доступа к данным по предъявляемому идентификатору.

Далее документы поступают на вход *подсистемы обработки*, задачей которой является формирование для каждого документа ПОД, в который заносится информация, необходимая для последующего поиска документа.

ПОД сохраняются в *индексе*, представляющем собой таблицу, строки которой соответствуют документам, а столбцы информационным признакам, на основе которых строится ПОД. В ячейках таблицы могут храниться либо 1, либо 0 – в зависимости от наличия или отсутствия данного признака в данном документе.

При поступлении на вход системы запроса пользователя он преобразуется в ПП и передается в подсистему поиска, задачей которой является отыскание в индексе ПОД, удовлетворяющих ПП с точки зрения КСС. Идентификаторы релевантных документов подаются с выхода подсистемы поиска на вход подсистемы хранения, которая осуществляет выдачу пользователю самих релевантных документов.

### 4.3 АНАЛИЗ ТЕКСТА И ИНДЕКСИРОВАНИЕ.

В автоматизированных документальных системах поиск информации обычно проводится по формализованным описаниям документов, которые составляются на основе их заголовков и текстов рефератов. В процессе формализации смысловое содержание документов переводится на формализованный язык. Процесс такого перевода принято называть индексированием. Индексирование связано потерей части информации, содержащейся в исходных текстах, и эта потеря бывает тем большей, чем беднее изобразительные средства информационного языка, на который осуществляет перевод. Получаемые в процессе индексирования формализованные описания документов (их поисковые образы) обычно бывают непригодны для восстановления исходных текстов. Поэтому для выдачи результатов поиска из ЭВМ приходится хранить в памяти машины наряду с формализованными описаниями документов также их заголовки или тексты рефератов.

Поиск документов может выполняться и по их неформализованным описаниям. При этом программы поиска строятся, исходя из определенной формализованной модели “понимания” текстов, и интерпретация содержания текстов осу-



ществляется на уровне этой модели. Такой подход к построению документальных систем позволяет отказаться от хранения в них формализованных описаний документов и дает возможность совершенствовать модели «понимания» текстов без изменения массивов ранее накопленной информации.

### 4.3.1 СТРУКТУРА ИНДЕКСА

Выше мы говорили, что документы могут быть представлены своим полным текстом или с помощью какой-нибудь схемы индексирования. Выборка основывается либо на поиске образца в полном тексте, либо на совпадении или сходстве запроса и индекса документа. Индексы документов реализуются как простой системой инвертированных файлов по отдельным ключам, так и полным инвертированием текста.

Простой индекс есть бинарное отношение  $Q(v, a)$ , в котором  $v$  – значение атрибута, а  $a$  – список адресов элементов хранения (записей или документов), соответствующих данному значению атрибута. В СУБД хранимым объектом может быть тип записи в физической памяти, соответствующий файлу. В ИПС хранимый объект – это документ такого же размера, как файл. Индекс  $Q(v, a)$  часто называют инвертированным индексом или инвертированным файлом в том смысле, что значения атрибутов извлекаются из элементов хранения на поверхность, т.е. инвертируются. Каждый элемент инвертированного индекса называется инвертированным списком. Соответственно можно полностью инвертировать хранимый объект (файл), построив индекс для каждого из его атрибутов. Слово атрибут больше подходит к структурам форматированных СУБД.

Если мы хотим инвертировать полный текст в ИПС, мы должны инвертировать каждое слово этого документа, добавив к нему информацию о контексте и окружении. В частично инвертированных системах только некоторые ключевые слова (термины), лучше всего представляющие документы, выбирают в качестве атрибутов инверсии.

Так как индекс частично повторяет содержимое памяти, большие БД имеют большие индексы или индексные файлы, которые, в свою очередь, также требуют эффективного доступа. Обычно строят иерархии индексов (индекс индекса и т.д.) до тех пор, пока корневой индекс не станет достаточно маленьким. Например, можно организовать инвертированные индексы в виде многоуровневых индексно-последовательных файлов, или, что эквивалентно, в виде инвертированного файла со словарями. Один из способов реализации таких индексных структур, обеспечивающий эффективную динамическую поддержку, – использование В-деревьев, листьями которых являются ссылки на документы. Приведен пример одноуровневого частично инвертированного файла документов. Рассмотрим следующий пример.

Номер документа
Зона1 (Автор)
Зона2 (Заголовок)
Зона3 (Аннотация)
Зона4 (Текс)

Формат документа

MACHINE(S), DBMS.

ФАЙЛ ДОКУМЕНТОВ содержит:

*DOC-1*

-

*ЗОНА 3: generalized database management systems with ANSI/SPARC architecture using a conceptual and three operational data models*

-

-

*DOC-2*

-

-

*ЗОНА 3: A network data model must be available in distributed databases*

-

-

*DOC-3*

-

-

*ЗОНА 3: A database machine is hardware solution to von Neuman bottlenecks in nonnumeric processing*

-

-

*DOC-4*

-

-

*ЗОНА 3: Powerful data languages are needed in DBMS*

-

-

*DOC-5*

-

-

*ЗОНА 3: ... a data definition languages as an integral part of a DBMS ... query processing in distributed DBMS using database machines*

ДОКУМЕНТАЛЬНАЯ БД: содержит 5 документов, формат которых приведен на рисунке.

ИНДЕКСАЦИЯ: только по зоне 3.

СТРУКТУРА ХРАНЕНИЯ: состоит из файла, содержащего 5 полных документов, и частично инвертированного индекса по ключевым словам, извлеченным из зоны 3.

АВТОМАТИЧЕСКОЕ ИНДЕКСИРОВАНИЕ по следующим ключевым словам: DATA, MODEL(S), LANGUAGE(S), DISTRIBUTED, DATABASE(S),

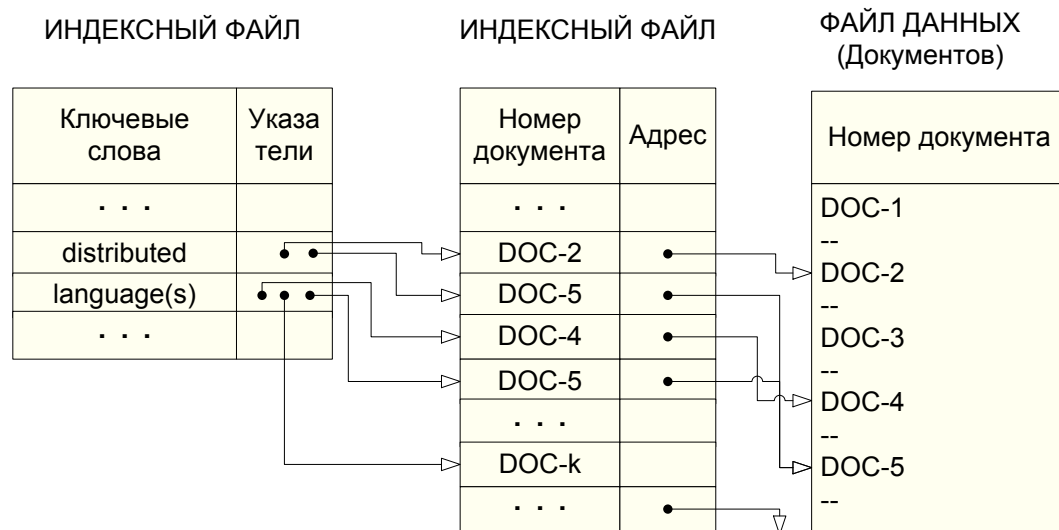
### ИНВЕРТИРОВАННЫЙ ИНДЕКС:

Ключевые слова ( v )	Адрес документа ( a )
----------------------	-----------------------

Рис.18. Трехуровневая структура инвертированного индекса

DATA	DOC-1, DOC-2, DOC-4, DOC-5
DATABASE (S)	DOC-1, DOC-2, DOC-3, DOC-5
DBMS	DOC-4, DOC-5
DISTRIBUTED	DOC-2, DOC-5
LANGUAGE (S)	DOC-4, DOC-5
MACHINE (S)	DOC-3, DOC-5
MODEL (S)	DOC-1, DOC-2

Для эффективности поиска значений и слияния/пересечения списков *a* значения *v* и адреса *a* хранятся в упорядоченном виде. Списки *a* в документальных базах данных значительно больше, чем в СУБД. Поэтому в ИПС инвертированный файл для документов обычно хранится в трех разных файлах, связанных указателями (см. Рис.18).



### 4.3.2 ИНВЕРТИРОВАНИЕ ТЕКСТА

Если документы в базе данных представлены с помощью индексов, возможны следующие варианты индексирования:

1. *Частичное инвертирование.* В этом случае для каждого из терминов, представленных в индексе (см. пример), файл пересылок, содержит адреса документов, в которых встречается этот термин.
2. *Полное инвертирование.* Все слова документа, в том числе и незначащие, включаются в индекс. Кроме того, каждая запись файла пересылок содержит информацию о позиции каждого вхождения термина в каждый из содержащих его документов.
3. *Неполное инвертирование.* Особая форма полного инвертирования, в которой незначащие слова удаляются, а остальные объединяются по своим основам при хранении документов в базе данных.

Существуют реализации (в том числе и коммерческие) систем, использующих варианты индексирования 1 и 3. Реализации систем с полным инвертированием, основанных на косвенном представлении текста, отсутствуют из-за огромного расхода памяти, однако вполне возможно создание базы данных с полными текстами в прямом представлении.

## 4.4 ПОИСК

При поиске в ИПС необходимо уметь найти релевантные документы в соответствии с терминами, заданными в запросе пользователя. Существуют два основных подхода:

- ✧ поиск на точное совпадение терминов;
- ✧ поиск с использованием меры сходства

Поиск на точное совпадение можно классифицировать на:

- ✧ поиск с инвертированием (индексированием) текста;
- ✧ поиск по образцу в полном тексте;
- ✧ поиск в полном тексте с использованием КА.

### 4.4.1 ПОИСК НА ТОЧНОЕ СОВПАДЕНИЕ ТЕРМИНОВ.

Поиск инвертированием текста. В этом случае задается полное совпадение терминов (ключевых слов) для поиска идентификаторов документов, содержащих эти ключевые слова. Например, если нас интересуют документы, содержащие термин MODEL, поиск по полному совпадению обнаружит все документы, в которых есть одно или несколько вхождений этого термина. Однако документы, которые содержат вхождения слова MODELS, не будут найдены. Для того чтобы отыскать и эти документы, пользователь может задать условие частичного совпадения, включив в поисковый образец несущественный символ, например: MODEL?.

Существуют различные способы указания несущественных символов фик-

сированной длины и несущественных символов переменной длины. Вот некоторые примеры:

```
L??SE      - LOOSE, LEASE;
L*SE       - LSE, LESE, LEASE, LOOSE, LICENSE;
DISTRIBUT* - DISTRIBUTE, DISTRIBUTD,
              DISTRIBUTION, DISTRIBUTIONS,
              DISTRIBUTIVE;
*IVE       - GIVE, FORGIVE, MSGIVE, DISTRIBUTIVE.
```

Можно использовать логические выражения для составления сложных запросов из элементарных. Логическое выражение состоит из поисковых терминов и логических связок AND, OR, XOR, NOT между отдельными поисковыми подвыражениями. Например,

1. DATA AND MODEL
2. DATA OR DATABASE
3. DATA XOR DATABASE (содержат только один из терминов DATA или DATABASE) .
4. DATADASE NOT MACHINES.

Для обработки этих запросов мы можем использовать инвертированный файл из вышеприведенного примера. Для запроса 1 мы имеем

$$(DOC-1, DOC-2, DOC-3, DOC-4, DOC-5) \wedge (DOC-1, DOC-2) = DOC-1, DOC-2$$

т.е. документы 1 и 2 содержат термины DATA и MODEL.

Для запроса 2

$$(DOC-1, DOC-2, DOC-4, DOC-5) \vee (DOC-1, DOC-2, DOC-3, DOC-5) = DOC-1, DOC-2, DOC-3, DOC-4, DOC-5,$$

т.е. все документы БД содержат либо оба термина DATA и DATABASE, либо один из них.

Для примера 3

$$\begin{aligned} \text{Результат запроса 2} - \text{результат запроса (DATA AND DATABASE)} &= (DOC- \\ 1, DOC-2, DOC-3, DOC-4, DOC-5) - ((DOC-1, DOC-2, DOC-4, DOC-5) \wedge (DOC-1, DOC- \\ 2, DOC-3, DOC-5)) &= (DOC-1, DOC-2, DOC-3, DOC-4, DOC-4, DOC-5) - (DOC-1, DOC-2, \\ DOC-5) &= DOC-3, DOC-4 \end{aligned}$$

т.е. DOC-3 и DOC-4 содержат только термины DATA или DATABASE, но не содержат их одновременно.

Для примера 4

$$(DOC-1, DOC-2, DOC-3, DOC-5) - (DOC-3, DOC-5) = DOC-1, DOC-2$$

т.е. документы 1 и 2 содержат термин DATABASE, но не содержат термины MACHINE.

#### 4.4.2 ПОИСК ПО ОБРАЗЦУ В ПОЛНОМ ТЕКСТЕ.

В процессе поиска по образцу мы имеем образец, который нужно найти, строку текста, в которой надо выполнить поиск, и указатель на текущую позицию в строке. Перед поиском образец размещается в начале текста, но сравнение на-

чинается с конца образца. Последний символ образца сравнивается с соответствующим символом строки. Если они совпадают, сравнение продолжается справа налево. Если же они различны, образец сдвигается вправо и поиск продолжается с пропуском некоторых символов текста. Чтобы достигнуть конца текста как можно скорее, было предложено использовать два вида сдвигов  $\delta_1$  и  $\delta_2$ , которые выбираются по следующим критериям:

- ✧ когда обнаруживается несовпадение, делается попытка найти сравниваемый символ текстовой строки в остальной (расположенной левее) части образца. Если эта попытка успешна, образец сдвигается вправо на  $\delta_1$  с тем, чтобы выровнять найденные совпадающие символы в образце и тексте. Если же такого символа в образце нет, образец сдвигается вправо на  $\delta_1$  так, чтобы первый его символ находился сразу за рассмотренным символом текста. Наибольшее значение  $\delta_1$  равно длине образца.
- ✧ если часть образца совпадает с текстом, делается попытка найти вхождение этого фрагмента образца в левой его части. Как только такое вхождение найдено, образец сдвигается вправо на  $\delta_2$ , чтобы выровнять совпадающие части.

Проиллюстрируем эти операции на примере.

```
Образец:   AT_THAT
Текст:     WHICH_FINALY_HALTS._AT_THAT_POINT...
                ↑
```

F не встречается в образце, поэтому сдвигаем его вправо на  $\delta_1=7$  (длина абзаца).

```
Образец:           AT_THAT
Текст:     WHICH_FINALY_HALTS._AT_THAT_POINT...
                ↑
```

Пробел в образце есть, поэтому выполняем сдвиг вправо на  $\delta_1=4$ .

```
Образец:           AT_THAT
Текст:     WHICH_FINALY_HALTS._AT_THAT_POINT...
                ↑
```

T совпадает, но L в образце отсутствует. Из двух возможностей сдвинуть на  $\delta_2=3$  (чтобы выровнять T), либо на  $\delta_2=5$  (поскольку L отсутствует) выбираем больший сдвиг.

```
Образец:           AT_THAT
Текст:     WHICH_FINALY_HALTS._AT_THAT_POINT...
                ↑
```

Совпадает AT, но не совпадает «пробел». Опять две возможности:  $\delta_1=2$  выровнять «пробелы», либо  $\delta_2=5$  выравнивает подстроки AT. Выбираем больший сдвиг.

```
Образец:           AT_THAT
Текст:     WHICH_FINALY_HALTS._AT_THAT_POINT...
                ↑
```

Теперь образец полностью совпадает с текстом. Поиск закончен.

### 4.4.3 ПОИСК С ИСПОЛЬЗОВАНИЕМ КА

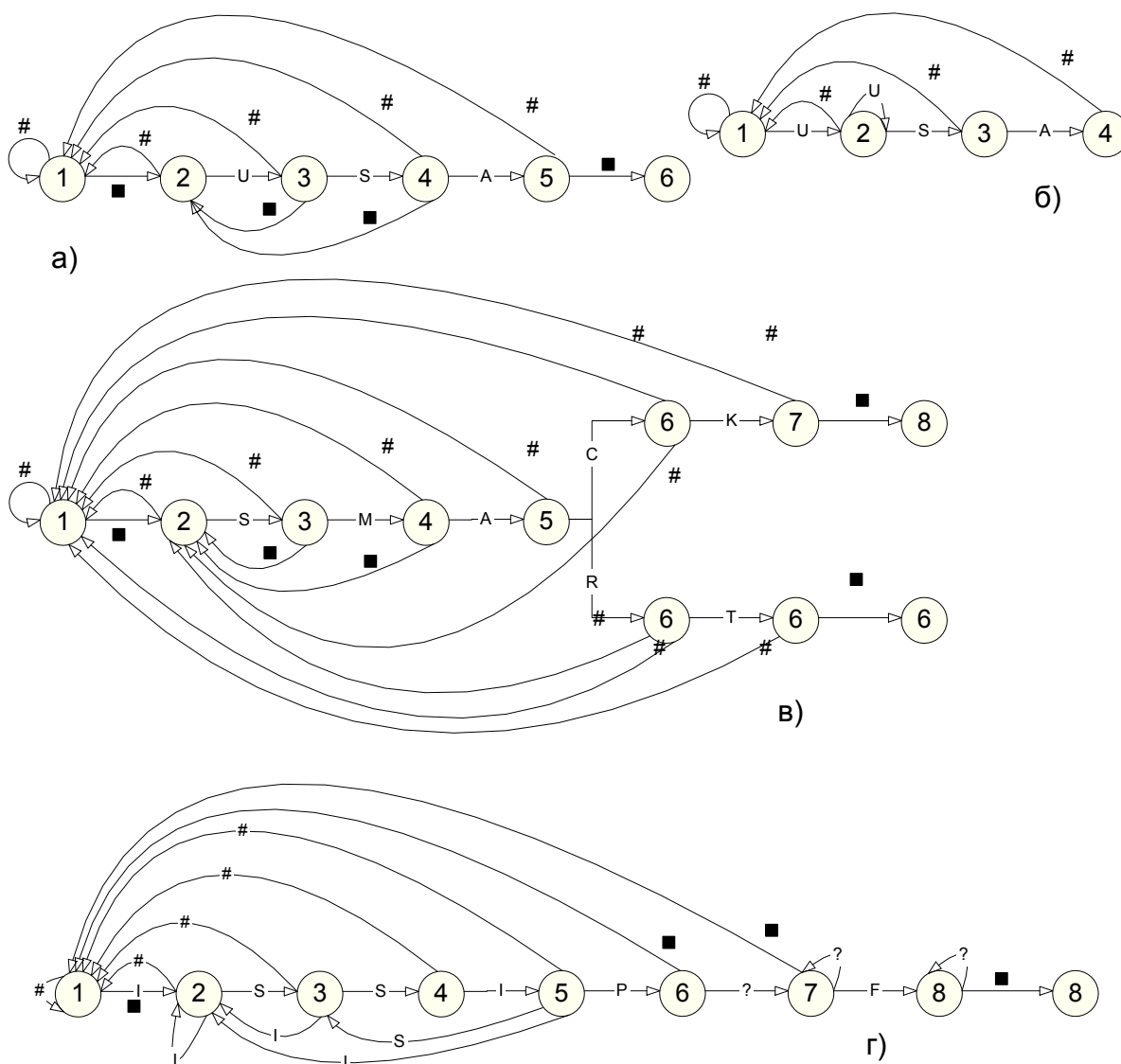
Конечные автоматы (КА) используются в качестве концептуального средства при моделировании вычислительных процессов в машине. Эта модель основана на понятии дискретных последовательных преобразований, используемом для описания вычислительных процессов. Один из таких процессов – поиск по образцу в тексте. Для программы, выполняющей поиск по образцу, мы можем построить КА со следующими данными:

- ✧ алфавит  $A$ , состоящий из входных символов;
- ✧ множество состояний  $S$ ;
- ✧ отображение  $M: A \times S \rightarrow S$ ;
- ✧ начальное и конечное состояния.

Алфавит для поиска по образцу состоит из всех символов входного текста и образца. Мы строим КА для заданного образца, указывая все состояния и возможные переходы. Во время поиска на вход КА подаются последовательно все символы текста и КА проходит через различные состояния. Если при этом достигается коечное состояние, говорят, что автомат распознает данную текстовую строку.

На Рис.19. приведено несколько КА. КА на Рис.19(а) распознает слово USA, на Рис.19(б) – слово \*USA\*, на Рис.19(в) – слова SMACK и SMART и на Рис.19(г) – \*ISSIP\*F. Знак # обозначает любой символ, отличный от тех, которые изображены на дугах переходов, а черный квадрат представляет символ конца слова.

Рис.19. Конечный автомат для поиска по образцу



В этих КА вершины с входящими и выходящими дугами представляют отображение  $A \times S \rightarrow S$ . Сама вершина соответствует текущему состоянию автомата.



Если текущий входной символ равен одному из символов, нарисованных на дугах, выходящих из текущей вершины, автомат переходит в новое состояние (справа от соответствующей стрелки). Входящие в вершину дуги представляют переходы, которые приводят КА в это состояние.

Реализация КА очевидна. Для каждого КА надо построить матрицу (таблицу) состояний, имеющую блок для каждого состояния КА. Каждый блок состояния содержащий в свою очередь массив, индексируемый или адресуемый входными символами (например, из 256 элементов для 8-битных символов). Каждый элемент матрицы содержит блока для следующего состояния. Однако такая непосредственная реализация неэкономно расходует память. Например, для автомата, имеющего 1000 состояний, потребовалось бы 256К элементов длиной в слово и более.

## ГЛАВА 5 ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ИНФС

Повышение эффективности использования ВТ при управлении предприятиями и организациями связано, прежде всего, перепоручением машине все большего числа не только вычислительных, но и логических операций по переработке информации. Развитие способности машины к логической переработке информации уже не может происходить только за счет исследования средств и методов выражения информации, а требует обращения к ее содержанию. Для этого важно выработать такие способы записи информации, которые обеспечивали бы в наибольшей мере использование ее смысла в процессе переработки. Поэтому создания ИнфЯ тесно связан с процессом создания ИнфО ИнфС.

Под использованием смысла информации следует понимать, прежде всего, возможность идентификации того или иного сообщения в целом, многоаспектного ее рассмотрения, установления значения входящих в ее состав элементов, определения их взаимных связей. Чем детальнее отображен смысл информации, чем рациональнее организованы средства выражения этого смысла, тем проще при прочих равных условиях программы обработки информации.

*Совокупность правил записи сообщений, отображающих определенный набор объектов и ситуаций, позволяющая на основе формальных процедур использовать в процессе обработки их смысл, и составляет информационный язык (ИнфЯ).*

Основные цели разработки ИнфЯ следующие:

- ✧ перепоручение машине большего числа логических операций, связанных переработкой смысла сообщений;
- ✧ упрощение общения с машиной за счет приближения языка общения с машиной к естественному;
- ✧ унификация языка общения с ИнфС, что является необходимой предпосылкой реализации принципа интеграции в обработке информации;

- ✧ объединение подсистем единой (глобальной) АС и достижение их согласованного функционирования посредством утверждения единой смысловой базы в языке их взаимного общения;
- ✧ упрощение управленческой документации за счет внутримашинного хранения информации, в частности всех промежуточных показателей.

Как всякий человеко-машинный язык, ИнфЯ должен обладать рядом свойств, налагаемых прагматикой человеко-машинного общения. Так его единицы и структура должны быть близкими к ЕЯ для удобства работы с ним людей. С другой стороны, тексты на таком языке должны легко вводиться в машину, его структура и логика должны быть максимально приближены к машинной и в этом смысле напоминать структуру и логику алгоритмических языков. В ИнфЯ должны оптимальным способом сочетаться достоинства ЕЯ – гибкость и эффективность в передаче мыслей, и искусственных языков – точность, однозначность, возможность формализации анализа и синтеза высказываний, их семантического анализа и синтеза.

Задачи организации ИнфО системы во многом совпадают с задачами анализа и синтеза языка ИнфС. Определение состава информации, выявление ее структуры и характера преобразований представляют собой неотъемлемый элемент исследования языка системы. Все компоненты структуры ИнфО взаимосвязаны и взаимозависимы. Основу структурной организации ИнфО системы составляет выделенная и упорядоченная система показателей управления объектом. Показатель является основным объектом классификации информации, центральным высказыванием и определяющим элементом в информационном языке (ИнфЯ), служит основным наполнением форм документов и массивов информации, т. е. БД. Документация составляется в основном из наборов показателей системы, служит основным средством передачи информации для решения задач системы, находится в сложной зависимости от других компонентов ИнфЯ системы.

Классификация информации определяет способы записи информации в массивах, т. е. информационный язык, структуру массивов, способы поиска и во многом процедуры простейшей переработки информации. ИнфЯ оказывает влияние на структурную организацию массивов хранимой информации (в то же время находясь в зависимости от нее) и в значительной степени определяет возможности и характер автоматизированной обработки: поиска, сортировки, группировки простейших логических преобразований данных.

## **5.1 ПОНЯТИЕ ЛИНГВИСТИЧЕСКОГО ОБЕСПЕЧЕНИЯ ИнфС**

Информационный процесс на предприятии представляет собой обмен информацией по установленным каналам между работниками, занятыми управлением. В течение длительного процесса совместной деятельности отдельные группы работников, участвующие в процессе управления, используя естественный язык, адекватно воспринимают смысл передаваемых и получаемых сообщений. Однако ввиду относительной разобщенности между персоналом различных

функций управления отдельные группы работников даже на одном предприятии зачастую используют различные термины для идентичных понятий и, наоборот, для выражения однозначных по смыслу высказываний применяют различные совокупности слов ЕЯ. Поэтому структурные единицы информационной базы (ИБ) предприятия необходимо описывать по единым правилам, обеспечивающим их представление в сопоставимых выражениях. В связи с этим требуется решить ряд «языковых проблем», которые решаются путем организации лингвистического обеспечения, тесно связанного с ИО и составляющего в совокупности с ним информационно-лингвистическое обеспечение (ИЛО) АС.

Информационный процесс на предприятии обслуживается ЕЯ человека. Применение при автоматизации процессов управления компьютерной техники вынуждает вводить в системе управления машинные языки, «понятные» машине, а также промежуточные алгоритмические языки и трансляторы для перевода с ЕЯ на язык машины и, наоборот, с машинного на ЕЯ. Но в любом случае первичная и конечная информация, которой пользуются работники управления, должна быть выражена на ЕЯ.

ЕЯ не приспособлен к использованию в качестве инструмента формализации при проведении операций по поиску семантических элементов, их сравнению, определению смысловой тождественности вследствие того, что с записью высказывания, насчитывающего 10-20 слов, затруднительно проводить эти операции. Наличие в языке таких свойств, как синонимия, омонимия, возможность фиксации одного явления, факта, события с помощью разного набора языковых конструкций также не способствуют решению этой задачи.

## **5.2 ПОНЯТИЕ ЯЗЫКА «ДЕЛОВОЙ ПРОЗЫ»**

Вместе с тем на предприятиях, в документах применяется весьма ограниченная часть словарного запаса ЕЯ, так как перечень наименований показателей, признаков и их совокупностей конечен. Это создает возможность разработки информационного языка, позволяющего решить задачу описания ИБ предприятия. Ее решению способствуют также следующие особенности языка «деловой прозы»:

- ✧ насыщенность специальной терминологией;
- ✧ относительно узкий набор грамматических конструкций;
- ✧ преобладание назывных именных конструкций;
- ✧ возможность отнесения большинства слов из наименований показателей одной системы к общим номенклатурам и т.д.

Поэтому на сегодняшний день актуальна задача разработки информационного языка, предназначенного для формализованного описания наименований структурных единиц ИБ. Можно сформулировать следующие основные требования, которые должны быть предъявлены к языку формализованного описания информационных совокупностей:

- ✧ однозначное описание высказываний;

- ✧ возможность идентификации (упорядочения структурных единиц ИБ СУ с целью присвоения им соответствующего идентификатора) наименований структурных единиц ИБ СУ;
- ✧ лаконичность и компактность выражений;
- ✧ простота структурных и конструктивных отображений, обеспечивающая быстрое овладение языком;
- ✧ относительно быстрый поиск структурных единиц, их группировка, сопоставление и анализ;
- ✧ возможность моделирования информационных связей и зависимостей между элементами;
- ✧ способность к развитию и совершенствованию.

Создание такого языка и методологии ее применения должно пройти следующие стадии:

- ✧ изучение ИБ СУ предприятия;
- ✧ классификация СЕ ИБ;
- ✧ подготовка правил формирования и формализованного отображения языковых конструкций;
- ✧ формирование словарей наименований СЕ ИБ.

Такой язык представляет собой семантическую систему, предназначенную для описания содержания СЕ ИБ, их совокупностей и связей между ними. В отличие от алгоритмических языков, этот язык не предназначен для описания операций в последовательности, регламентируемой алгоритмами.

### 5.3 СРЕДСТВА ОПИСАНИЯ

Информационный процесс на производстве обслуживается ЕЯ человека. Применение при автоматизации процессов управления ВТ вынуждает вводить в систему управления машинные языки, «понятные» машине, а также промежуточные алгоритмические языки и трансляторы для перевода с ЕЯ на язык машины, и наоборот, с машинного на ЕЯ. Однако какими бы ни были промежуточные «человеко-машинные» языки, первичная и конечная информация, которой пользуются работники управления, должна быть выражена на ЕЯ.

## **ГЛАВА 6 ЭТАПЫ РАЗРАБОТКИ ИО АС**

### **6.1 РАЗРАБОТКА ИО АС НА ПРЕДПРОЕКТНОЙ СТАДИИ.**

#### **6.1.1 ОСНОВНЫЕ МЕТОДЫ СОЗДАНИЯ ИО АС.**

ИО АС включает системы документов, совокупность классификаторов (ко-дификаторов), справочников, данных на машинном носителе (БД), загрузочные файлы на МН, проектную и эксплуатационную документацию, программные средства ведения БД, в том числе и СУБД. Разработка рационального варианта ИО АС, имеющую такую сложную структуру, требует значительных средств и усилий проектировщиков высокой квалификации, создания методики, обеспечивающей успех разработки.

Одним из методических принципов создания сложных систем, какой является ИО АС, служит структуризация, разделение сложной проблемы и решение ее по частям с учетом обеспечения взаимосвязи между отдельными частями. Поскольку разработка ИО АС является составной, неразрывной частью более общей проблемы создания АС, то и необходимые для этого работы делят на стадии, этапы так, как это регламентировано для АС или ИАСУ, например, НПО.

В настоящее время основным регламентирующим документом, определяющим состав стадий и этапов на каждой стадии при создании АС, ИАСУ является ГОСТ 24.601-86 «Единая система создания стандартов АС. Автоматизированные системы. Стадии создания». В соответствии с ним выделяют стадии:

- 1) исследование и обоснование создания АС;
- 2) техническое задание;
- 3) эскизный проект;
- 4) технический проект;
- 5) рабочая документация;

- б) разработка несерийных компонентов комплекса средств автоматизации;
- 7) ввод в действие.

При этом обязательными являются стадии 1,4,5,7. Остальные выполняются по необходимости при согласовании с Заказчиком.

В свою очередь, работы, выполняемые на каждой стадии, делят на этапы. Применительно к ИО АС выделяют следующие этапы его разработки:

- сбор и анализ данных;
- разработка требований к ИО;
- разработка технических решений по ИО;
- рабочая документация по ИО;
- комплексная отладка, опытная эксплуатация, проведение испытаний.

По используемым средствам технологии проектирования АС и ИО АС удобно разделить на следующие классы:

- индивидуальное проектирование;
- проектирование с использованием ППП;
- проектирование с использованием САПР АС или CASE-средств;
- комбинированные подходы.

С точки зрения организации работ, при создании ИО АС используются «процедурный» и «непроцедурный» подходы.

«Процедурный» подход создания ИО АС предполагает изучение, анализ информационных потребностей задач, состав которых должен быть предварительно определен, т.е. выполнена функциональная структуризация АС. В результате анализа указанных информационных потребностей «практических приложений» выявляют реквизитный состав информации, ее характеристики, в том числе совокупности совместно используемых данных, частоту использования (активность) и т.д. На основании результатов такого анализа структурирование данных, которое отвечало бы информационным потребностям задач. Следовательно, в основе этого подхода лежит первичная разработка функциональной структуры АС.

При «непроцедурном» или «объектном» подходе создания ИО АС предварительно определяют перечень объектов, автоматизацию управления которыми предполагают осуществить. Предполагается наличие у разработчиков ИО типовых информационно-логических моделей соответствующих объектов управления, которые адаптируются, видоизменяются в соответствии с особенностями конкретных объектов управления.

### 6.1.2 СБОР И АНАЛИЗ ДАННЫХ.

Нельзя рассчитывать на построение рациональной АС, не изучив предварительно все стороны объекты управления и существующую систему управления этими объектами. Описание ОУ, а также процессов управления ими в основном содержится в имеющихся системах документов. Поэтому изучение содержания документов, путей их движения, процедур создания, оформления, ведения (внесения изменений), хранения, передачи, использования для выполнения функций

управления является основным содержанием работ предпроектной стадии. Цель выполнения этих работ – получение точной информации о всех характеристиках объектов и системы управления.

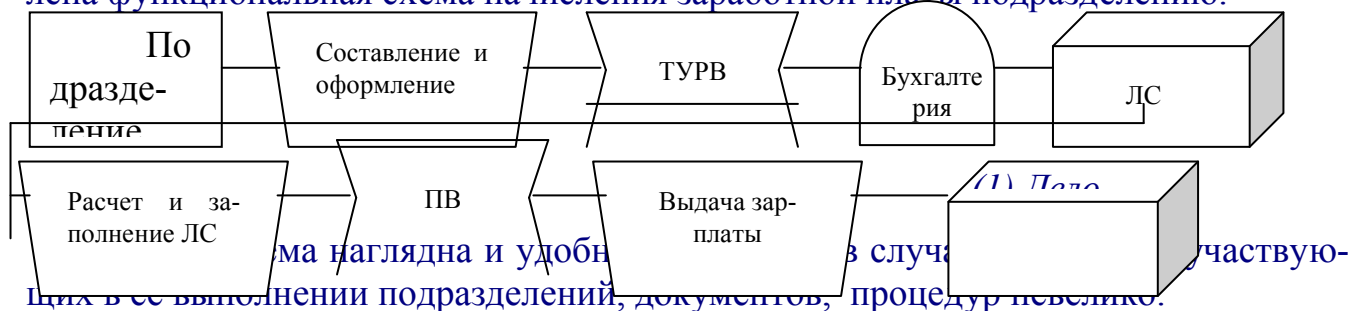
Полностью достичь этой цели только изучением документов не удастся, поскольку многие функции выполняются либо без документального оформления, либо только частично соответствуют тому, что отражено в документах. Поэтому, наряду с изучением документопотоков очень важно правильно организовать и провести собеседование с различными группами специалистов и руководителями.

Целью составления описания существующей системы управления является представление процесса управления в наглядном и анализируемом виде. Этого

*Рис.20. Функциональная схема начисления заработной платы: ТУРВ – табель учета рабочего времени. ЛС – лицевые счета, ПВ – платежные ведомости.*

обычно достигают разработкой компактных функциональных, функционально-технологических схем выполнения функций управления. Эти схемы должны содержать информацию, достаточную для анализа, должны быть компактными и обозримыми.

Функциональная и функционально-технологическая схемы представляют собой описание процесса реализации на изучаемом объекте отдельных функций управления, выполняемыми ими процедуры и последовательность их выполнения. Описание выполняют в соответствии с ГОСТ 19.002-80 «ЕСПД. Схемы алгоритмов и программ. Правила выполнения» и ГОСТ 19.003-80 «ЕСПД. Схемы алгоритмов и программ. Обозначения условные графические». На рис. Представлена функциональная схема начисления заработной платы подразделению.



В противоположном случае она становится труднопонимаемой из-за своей неупорядоченности. В таких случаях используют функционально-технологические схемы, имеющие упорядоченную структуру.

Функционально-технологические схемы выполняют в виде таблицы, разделенной на графы. Каждая графа отводится для размещения в ней определенной информации (описания состава процедур, документов), непосредственно относящейся к подразделению, наименование которого указывают в заголовке графы.

Пример функционально-технологической схемы для расчета заработной платы подразделению представлен на рис.



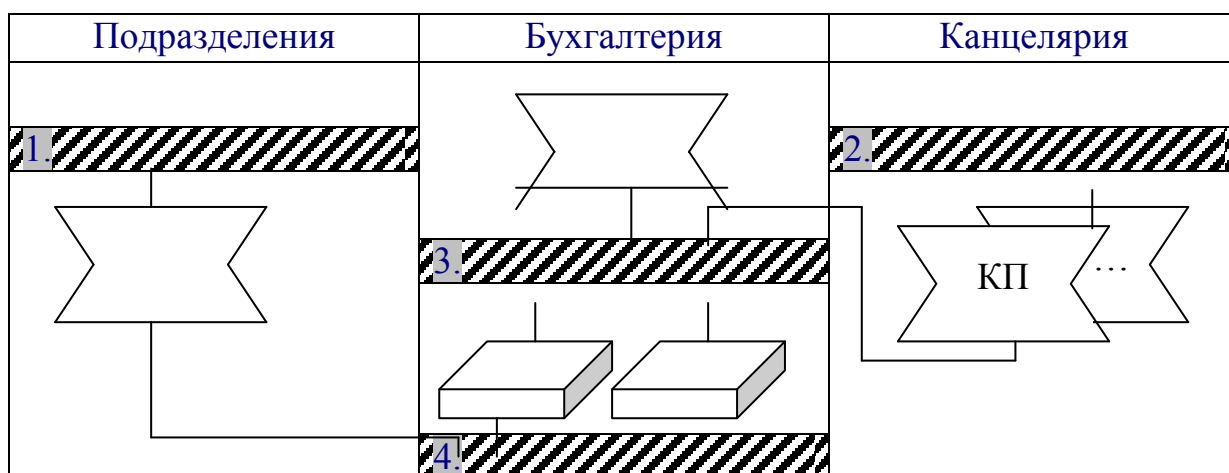
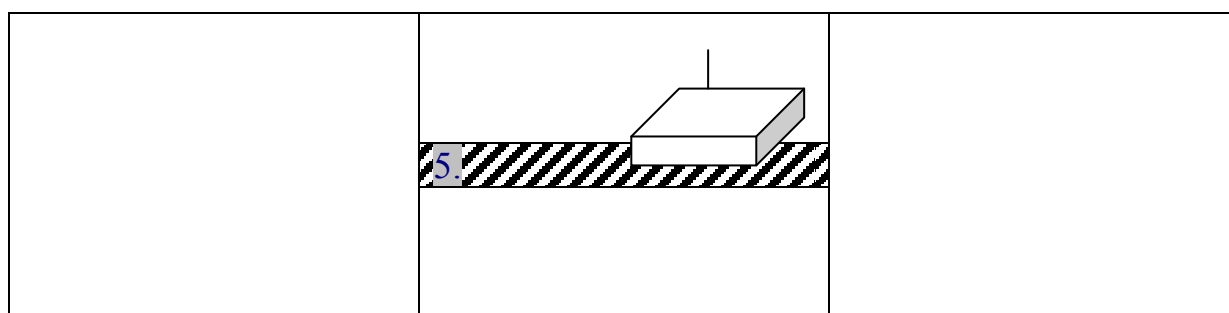


Рис.21. Рис. Функционально – технологическая схема расчета зарплаты: И – инструкция по начислению заработной платы, ИП – копии приказов по бухгалтерии, КЛС – карточка лицевых счетов, КП – дело копий приказов, №1 – дело ТУРВ. №2 – дело ПВ.



Процедуры на данной схеме обозначены в виде узких заштрихованных прямоугольников, которые нумеруются в порядке их выполнения и которые прочерчиваются через всю графу, отводимую для подразделения. Сверху процедуры указывают исходные документы, снизу – выходные для этой процедуры. На рис. Отображены следующие процедуры:

- составление, оформление ТУРВ, передача бухгалтерии;
- регистрация копий приказов, передача в бухгалтерию;
- расчет зарплаты, занесение результатов расчета в карточку лицевого счета бухгалтерии;
- составление, оформление ПВ;
- выдача зарплаты кассирам, помещение ПВ в дело бухгалтерии.

Процесс создания таких схем является первым этапом анализа, позволяющим осмыслить содержание процесса управления. Вторым этапом является разработка предложений по рационализации процесса выполнения рассмотренных функций управления. Эта работа должна выполняться заказчиком и разработчиком совместно.

### 6.1.3 РАЗРАБОТКА ТРЕБОВАНИЙ К ИО АС.

В соответствии с ГОСТ 24.104-85 «Единая система стандартов автоматизированных систем управления. Общие требования», основными требованиями к ИО АС является его достаточность для выполнения всех автоматизированных функций АС. Для выполнения этого требования необходимо разработать схему функций проектируемой АС, содержащую как их состав, так и связи между ними.

На схеме функций не указывают документы, процедуры, подразделения, а лишь совокупность взаимосвязанных функций, каждая из которых является обобщением по отношению к функционально-технологической схеме. Так, рассмотренная выше функция расчета и выдачи зарплаты на схеме функций будет

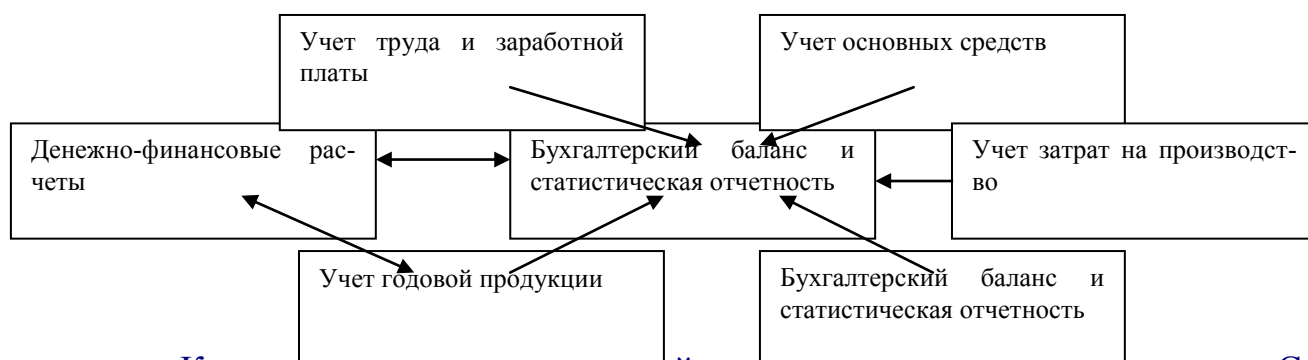
представлена прямо-угольником и взаимосвязана с другими функциями бухгалтерского учета.

Рис.22. Рис. Схема функций бухгалтерского учета

Цель построения и анализа схемы функций – обеспечить необходимую их полноту в АС, рационализировать состав с учетом использования ЭВМ в процессе их выполнения.

При разработке схемы функций важным является выбор степени детализации их описания. Так, например, в качестве функций можно указать «бухгалтерский учет», либо представить это в виде совокупности следующих функций:

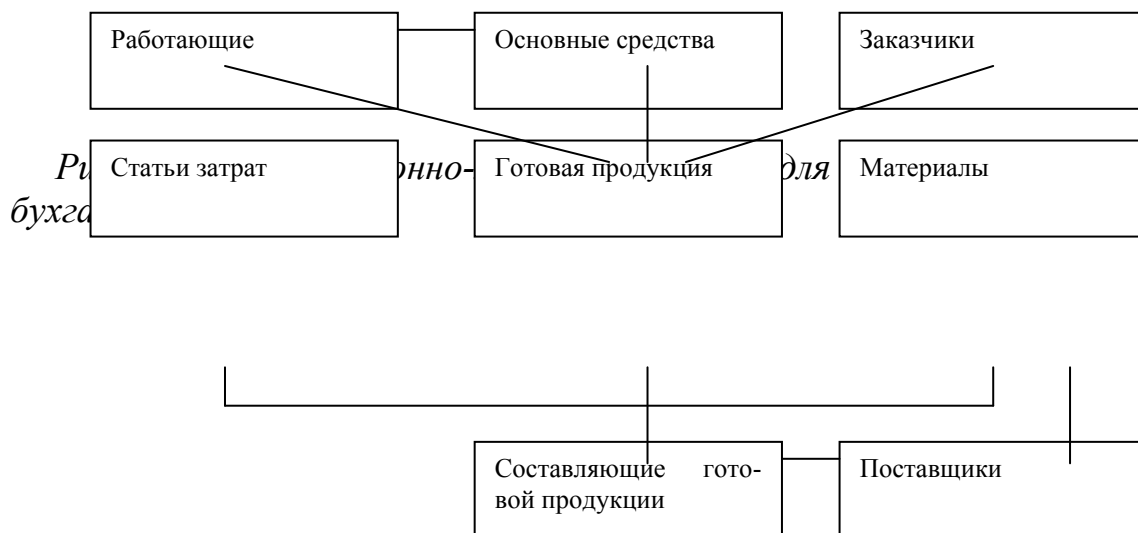
- учет труда и заработной платы;
- учет основных средств;
- учет материальных ценностей;
- учет затрат на производство;
- учет готовой продукции;
- денежно-финансовые расчеты;
- бухгалтерский баланс и сводная бухгалтерская отчетность.



Каждая из указанных функций может быть еще детализирована. Степень детализации описания функций или, что то же самое, степень структуризации схемы функций выбирают исходя из удобства построения на ее основе информационно-логических моделей (ИЛМ), выполняемых обычно в рамках отдельных функциональных подсистем. Каждой подсистеме на предприятии чаще всего соответствует функциональное подразделение.

Если изучение, описание и анализ объекта и системы управления проводились в соответствии с описанной выше методикой, т.е. по функциональным подразделениям, то и схему функций логично строить в рамках каждого подразделения, т.е. по подсистемам. При этом следует учитывать и отображать информационные взаимосвязи между подсистемами.

Основным элементом ИЛМ является информационный объект, выделяемый на основе общности материального объекта. На рис. представлена ИЛМ для бухгалтерского учета.



Из этого примера видно, что знание схемы функций недостаточно для построения ИЛМ; такие информационные объекты, как «Статьи затрат», «Поставщики», «Заказчики», не имеют прямого соответствия в схеме функций. Поэтому для построения ИЛМ необходимо привлекать функционально-технологические схемы, первичные материалы описания.

Построенные схемы функциональной структуры и ИЛМ дают возможность разработать основные решения по видам обеспечения, в том числе, по техническому, программному, информационному, и ответить на следующие вопросы:

будет ли система строиться как централизованная, распределенная, смешанная;

будут ли применяться СУБД, и какие;

каковы основная структура и организация данных;

какие общепромышленные, отраслевые унифицированные системы и формы документов, зарегистрированные унифицированные документы предприятия будут использованы;

какие общегосударственные, отраслевые классификаторы, а также зарегистрированные классификаторы предприятия будут применены;

каков будет в условиях АС технологический процесс сбора, передачи, обработки и представления данных.

Ответы на эти и другие вопросы оформляют в виде документов ТЗ на проектирование АС в соответствии с требованиями ГОСТ 24.201-85 «Единая система стандартов АС. Техническое задание на АС». Они определяют требования к ИО

АС, подлежащей техническому и рабочему проектированию.

Таким образом, на предпроектной стадии проводятся следующие этапы работ по созданию АС, в том числе разработка ИО АС:

- ✧ сбор и описание документов, проведение собеседований;
- ✧ составление функционально-технологических схем;
- ✧ составление структурно-функциональных схем;
- ✧ построение ИЛМ (исходный вариант);
- ✧ разработка основных требований по обеспечивающим частям АС, в том числе по немашиной и внутримашинной частям ИО АС;
- ✧ оформление ТЗ на АС.

## **6.2 РАЗРАБОТКА ИО АС НА СТАДИЯХ ПРОЕКТИРОВАНИЯ**

### **6.2.1 РАБОТЫ ПО СОЗДАНИЮ ИО АС НА СТАДИИ ТЕХНИЧЕСКОГО ПРОЕКТА**

Приступая к техническому проектированию, разработчик ИО АС имеет ТЗ, в котором определены требования к ИО АС, в том числе:

основная (исходная) ИЛМ по функциональным подсистемам, системам; использованные системы документов, классификаторы; использованные СУБД, ОС ЭВМ, ППП окружения, функциональные ППП; типы применяемых ЭВМ, периферийные средства ввода-вывода..

В результате технического проектирования ИО АС в соответствии с ГОСТ 24.201-88 «Система технической документации на АС. Виды и комплектность документов» должны быть составлены следующие документы:

описание ИО АС;  
описание организации внутримашинной информационной базы;  
описание организации немашиной информационной базы;  
описание системы классификации и кодирования;  
описание массивов информации;  
описание форм документов;  
перечень входных документов;  
перечень выходных документов.

ГОСТ 2.205-88 «Система технической документации на АС. Требования к содержанию документов по информационному обеспечению» определяет содержание этих документов.

После того как выявлены общие потребности в данных для всех задач АС, проектирование внутримашинной информационной базы выполняют в такой последовательности:

разрабатывают концептуальную модель по системам, подсистемам;  
разрабатывают логическую модель хранения данных для конкретной СУБД;  
определяют и проектируют внешние представления для выполнения отдельных функций.

## 6.2.2 РАБОТЫ ПО ИО АС НА СТАДИИ РАБОЧЕЙ ДОКУМЕНТАЦИИ

Основным содержанием этой стадии являются работы, связанные с завершением загрузки ИО АС и составление рабочей документации. Для выполнения загрузки БД и файлов, входящих в информационную базу, необходимо:

- завершить работы по формированию и контролю загрузочных файлов;
- завершить отладку программных средств и сервисных программ;
- выполнить непосредственную загрузку ИБ;
- сформировать на машинных носителях дубли ИБ и организовать их хранение;

- организовать постоянную актуализацию ИБ;

администратору БД осуществить приемку всей необходимой рабочей документации, сгенерированных СУБД, ППП окружения, оригинальных программ загрузки и ведения ИБ.

Собственно загрузке ИБ должна предшествовать организационная работа, которая включала бы ИБ в общий документооборот в части внесения необходимых изменений от различных служб системы управления. Если это не будет выполнено, ИБ быстро потеряет актуальность, данные в ней не будут отражать текущего состояния объектов управления, оно не может быть использована для целей управления.

При подготовке, загрузке, ведении данных важно соблюсти принцип: данные готовятся в месте возникновения информации. Это означает, что данные по конструкции изделий должны готовить сами конструкторы, по технологии – сами технологи. Как подготовку данных, загрузку и ведение БД, так и ее дальнейшую эксплуатацию невозможно осуществить без разработки и использования необходимых рабочих документов. К ним, в соответствии с ГОСТ 24.201-80, относятся следующие этапы:

- описание технологического процесса обработки данных;
- должностные инструкции;
- технологические инструкции;
- инструкции по эксплуатации.

## 6.2.3 РАБОТЫ ПО СОЗДАНИЮ ИО АС НА СТАДИИ ВВОДА В ДЕЙСТВИЕ АС

На завершающей стадии создания АС основные работы связаны с подготовкой организации к вводу в действие, комплектацией необходимыми компонентами технического и программного обеспечения, комплексной отладкой техники и программ. В завершение проводят государственные приемочные испытания всей АС, устраняют выявленные при этом недостатки и осуществляют приемку АС в промышленную эксплуатацию и сопровождение.

Применительно к ИО АС эти работы заключаются в следующем:

- пользователи АС обучаются действиям, необходимым для использования

ИБ АС и для поддержания ее в актуальном состоянии;

персонал, обслуживающий ИБ, фактически должен перейти в режим постоянной эксплуатации, после того как для них будет выполнена загрузка соответствующей части ИБ;

с этого момента ИБ должна иметь постоянную связь с соответствующими источниками корректирующей информации, т.е. теми службами, которые вносят изменений в документы и информацию.

Приемка ИБ и ИО в опытную эксплуатацию может осуществляться по частям, по мере готовности ИБ для отдельных приложений, необходимого для этого программного, технического, организационного (инструкций) обеспечения. Для этого проводится проверка завершения всех необходимых работ **по составляется акт завершения работ.**

По завершении необходимых работ и составлении об этом акта осуществляют опытную эксплуатацию завершенной части ИО АС. Целью проведения опытной эксплуатации является практическая проверка работоспособности и эффективности проверяемой части ИО АС, освоение работы с БД, обеспечение взаимодействия служб, персонала, участвующего в технологической цепи ведения и использования ИБ АС. Для проведения опытной эксплуатации назначается комиссия, которая составляет акт приемки в опытную эксплуатацию.

Для проведения опытной эксплуатации составляется и объявляется приказ о начале опытной эксплуатации, в котором указываются состав комиссии по проведению опытной эксплуатации, сроки начала и завершения работ, форма завершения работ.

Комиссия составляет программу опытной эксплуатации ИБ, БД для определения порядка и объема проведения опытной эксплуатации.

По результатам опытной эксплуатации составляют протокол испытаний. В нем указывают пункты программы опытной эксплуатации, по которым проводились испытания, режимы, условия испытаний, полученные результаты.

После устранения выявленных недостатков ИБ ее части передаются в промышленную эксплуатацию, которая проводится в аналогичном порядке. При этом составляются следующие документы:

- ✧ акт приемки в промышленную эксплуатацию ИБ, БД;
- ✧ приказ о проведении работ;
- ✧ приказ о составе приемочной комиссии;
- ✧ программа работ;
- ✧ программа испытаний;
- ✧ протокол испытаний;
- ✧ приказ о вводе в промышленную эксплуатацию ИБ в целом или ее частей.

Состав документов, разрабатываемых на стадии «Ввод в действие», и их содержание регламентируется ГОСТ 24.208-87 «Система технологической документации на АС. требования к содержанию документов стадии «Ввод в действие».

## СПИСОК ЛИТЕРАТУРЫ

1. Дракин В.Н., и др. Общение конечных пользователей с системами обработки данных. М.: Радио и связь, 1988 г.
2. Каутс Р., Впейминг Н. *Интерфейс 'человек-компьютер'*. М.: Мир, 1990г.
3. Филлипович Ю. Н. Организация взаимодействия человека с техническими средствами АС. В 7 книгах. Книга 2. Языковые средства общения человека с ЭВМ. М.: ВШ, 1990 г.
4. Попов Э.В. *Общение с ЭВМ на языке ЕЯ*. М.:Наука,1982 г.
5. Лингвистические вопросы алгоритмической обработки сообщений. М.:Наука, 1983 г.
6. Белоногов Г.Г., Языковые средства автоматизированных систем. М.:Наука, 1983 г.