

Министерство образования и науки Российской Федерации  
Федеральное государственное бюджетное образовательное  
учреждение высшего профессионального образования  
«Оренбургский государственный университет»

Кафедра программного обеспечения вычислительной  
техники и автоматизированных систем

С.А. Шезоков, Е.Н. Чернышова

## ПРОЕКТИРОВАНИЕ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ



Рекомендовано Ученым советом федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по направлению подготовки 230100.68 Информатика и вычислительная техника.

Оренбург 2012

Елена Чернопрудова

**Проектирование  
распределенных  
информационных систем**

«БИБКОМ»

2012

УДК 004.75(075.8)  
ББК 32.973.202я73

**Чернопрудова Е. Н.**

Проектирование распределенных информационных систем  
/ Е. Н. Чернопрудова — «БИБКОМ», 2012

В курсе лекций изложены общие теоретические основы баз данных и информационных систем, концептуальные основы распределенной обработки информации, методические основы и примеры практического проектирования распределенных информационных систем с использованием современных программных инструментальных средств.

УДК 004.75(075.8)  
ББК 32.973.202я73

© Чернопрудова Е. Н., 2012  
© БИБКОМ, 2012

# Содержание

Введение	5
Обозначение предметной области «распределенные информационные системы» и проблемных вопросов изучения дисциплины	6
1 Лекция 1. Назначение и основные компоненты информационной системы	17
Конец ознакомительного фрагмента.	33

# Щелоков С.А., Чернопрудова Е.Н. Проектирование распределенных информационных систем. Курс лекций

## Введение

Целью изучения дисциплины является овладение современными методами и средствами технологии исследования и проектирования, разработки и использования проблемно – ориентированных распределенных информационных систем (РИС). Для достижения поставленной цели предусматривается решение следующих основных задач: изучение распределенной обработки информации в автоматизированных информационных системах, архитектуры РИС, технологической базы РИС, распределенных информационных ресурсов и сетей, распределенных баз данных, принципов и технологий управления обменом информацией в РИС, методов и средств доступа к удаленным информационным ресурсам.

Дисциплина относится к дисциплинам вариативной части учебного профессионального цикла.

Выпускник должен обладать следующими общекультурными компетенциями (ОК):

– способен к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности (ОК-2);

– осознает сущность и значение информации в развитии современного общества; владеет основными методами, способами и средствами получения, хранения, переработки информации (ОК-11);

Выпускник должен обладать следующими профессиональными компетенциями (ПК):

– применять перспективные методы исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий (ПК-1);

– организовывать работу и руководить коллективами разработчиков программных средств информационных и автоматизированных систем (ПК-7).

## **Обозначение предметной области «распределенные информа-ционные системы» и проблемных вопросов изучения дисциплины**

Понятие «персональный компьютер», возникшее в уже далеком 1945 году и обозначающее индивидуальную работу пользователя в отдельно взятой комнате, изолированно от других пользователей, претерпело большие изменения и в реальной обыденной жизни и в виртуальном характере общения пользователя с информационными объектами. Начиная с середины восьмидесятых годов большие и дорогие майнфреймы уступают место компактным компьютерам с более мощными микропроцессорами. Следующий виток технологического развития обозначается появлением локальных сетей (Local-Area Networks, LAN), позволяющих объединить сотни компьютеров, находящихся в здании, таким образом, что машины в состоянии обмениваться небольшими порциями информации за несколько микросекунд. Большие массивы данных передаются с машины на машину со скоростью от 10 Мбит/с до 10 Гбит/с. Затем появляются глобальные сети (Wide-Area Networks, WAN), позволяющие миллионам машин во всем мире обмениваться информацией со скоростями, варьирующимися от 64 кбит/с (килобит в секунду) до гигабит в секунду.

В результате развития этих технологий сегодня не просто возможно, но и достаточно легко можно собрать компьютерную систему, состоящую из множества компьютеров, соединенных высокоскоростной сетью. Ее можно назвать простейшей распределенной информационной системой (РИС), в отличие от предшествовавших ей централизованных, или однопроцессорных систем, состоявших из одного компьютера, его периферии и, возможно, нескольких удаленных терминалов. Технологический скачок революционного развития вычислительной техники потребовал концептуальных изменений в использовании средств обработки информации.

Появился новый термин – «распределенная информационная система». Возникает научная задача – термину РИС нужно дать лаконичное и научно обоснованное определение. На сегодняшний день по утверждению известного специалиста в области информатики Э.Таненбаума, не существует общепринятого и в то же время строгого определения распределенной системы. В современной литературе можно выделить следующие научные толкования нашего термина как распределенная автоматизированная система (РАС):

– РАС – это автоматизированная система управления, которая приобрела специфику территориально рассредоточенной автоматизированной системы;

– РАС – это совокупность независимых объектов, которые взаимодействуют с целью решения проблемы, которая не может быть решена одним объектом индивидуально. При таком подходе распределенной является любая вычислительная система, где обработка данных разделена между двумя и более компьютерами;

– РАС – это совокупность независимых компьютеров, представляющаяся пользователям единой объединенной системой.

Такой подход к определению распределенной системы имеет свои недостатки. Например, все используемое в такой распределенной системе программное обеспечение могло бы работать и на одном единственном компьютере, однако с точки зрения приведенного выше определения такая система уже перестанет быть распределенной;

– распределенной является такая вычислительная система, в которой неисправность компьютера, о существовании которого пользователи ранее даже не подозревали, приводит к остановке всей их работы. Значительная часть распределенных вычислительных систем,

к сожалению, удовлетворяют такому определению, однако формально оно относится только к системам с уникальной точкой уязвимости.

Первый прототип РИС, имеющий структуру «клиент – сервер», следует рассматривать (рисунок 1) как некое типичное приложение, которое в соответствии с современными представлениями может быть разделено на следующие логические уровни:

- пользовательский интерфейс (ИП);
- логика приложения (ЛП);
- доступ к данным (ДД), работающий с базой данных (БД).

Пользователь системы взаимодействует с ней через интерфейс пользователя, база данных хранит данные, описывающие предметную область приложения, а уровень логики приложения реализует все алгоритмы, относящиеся к предметной области.

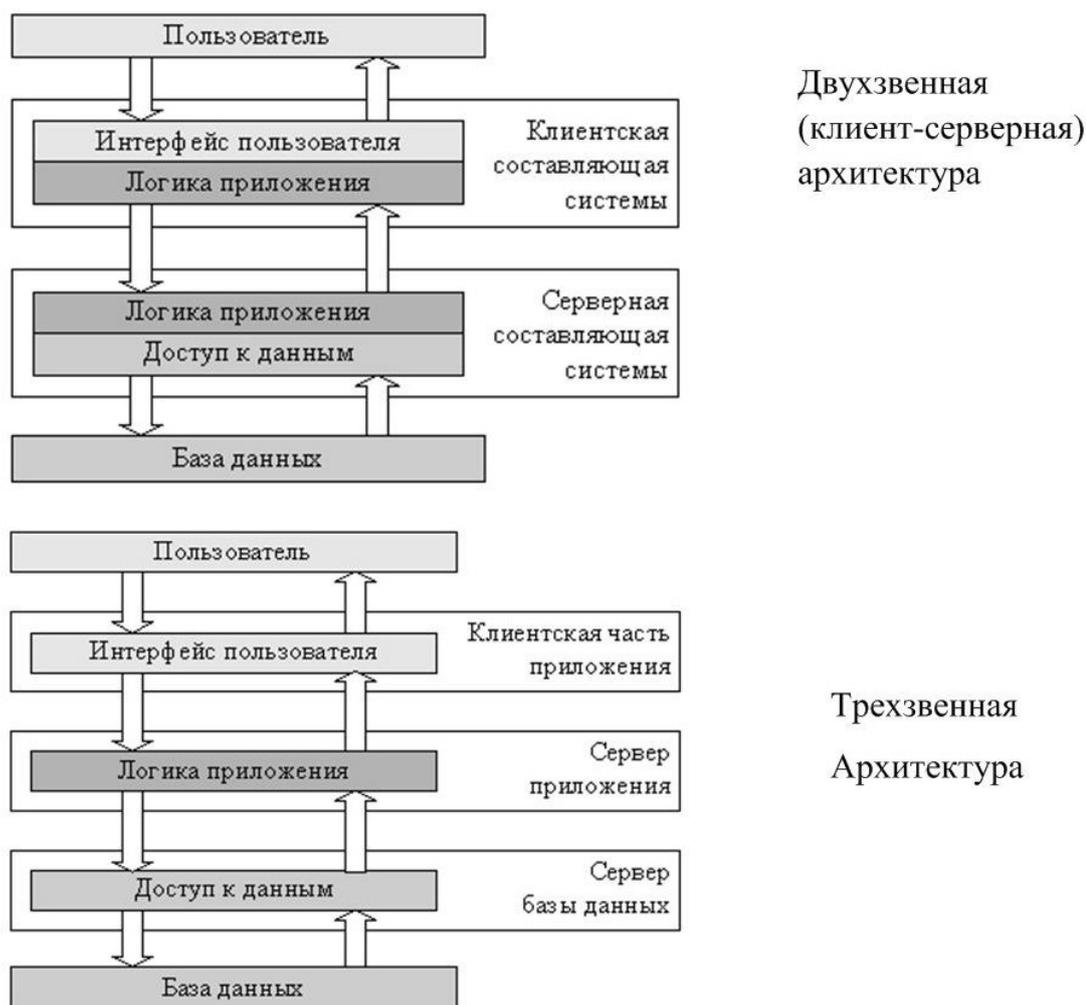


Рисунок 1 – Архитектура «клиент – сервер»

Поскольку на практике разных пользователей системы обычно интересует доступ к одним и тем же данным, наиболее простым разнесением функций такой системы между несколькими компьютерами будет разделение логических уровней приложения между одной серверной частью приложения, отвечающим за доступ к данным, и находящимися на нескольких компьютерах клиентскими частями, реализующими интерфейс пользователя. Логика приложения может быть отнесена к серверу, клиентам, или разделена между ними. Архитектуру построенных по такому принципу приложений называют клиент-серверной

или двухзвенной. На практике подобные системы часто не относят к классу распределенных, но формально они могут считаться простейшими представителями распределенных систем.

Развитием архитектуры клиент-сервер является трехзвенная архитектура, в которой интерфейс пользователя, логика приложения и доступ к данным выделены в самостоятельные составляющие системы, которые могут работать на независимых компьютерах. Запрос пользователя в подобных системах последовательно обрабатывается клиентской частью системы, сервером логики приложения и сервером баз данных. Однако обычно под распределенной системой понимают системы с более сложной архитектурой, чем трехзвенная.

Применительно к приложениям автоматизации деятельности предприятия, распределенными обычно называют системы с логикой приложения, распределенной между несколькими компонентами системы, каждая из которых может выполняться на отдельном компьютере. Например, реализация логики приложения системы розничных продаж (рисунок 2) должна использовать запросы к логике приложения третьих фирм, таких как поставщики товаров, системы электронных платежей или банки, предоставляющие потребительские кредиты. Таким образом, в обиходе под распределенной системой часто подразумевают рост многозвенной архитектуры "в ширину", когда запросы пользователя не проходят последовательно от интерфейса пользователя до единственного сервера баз данных.

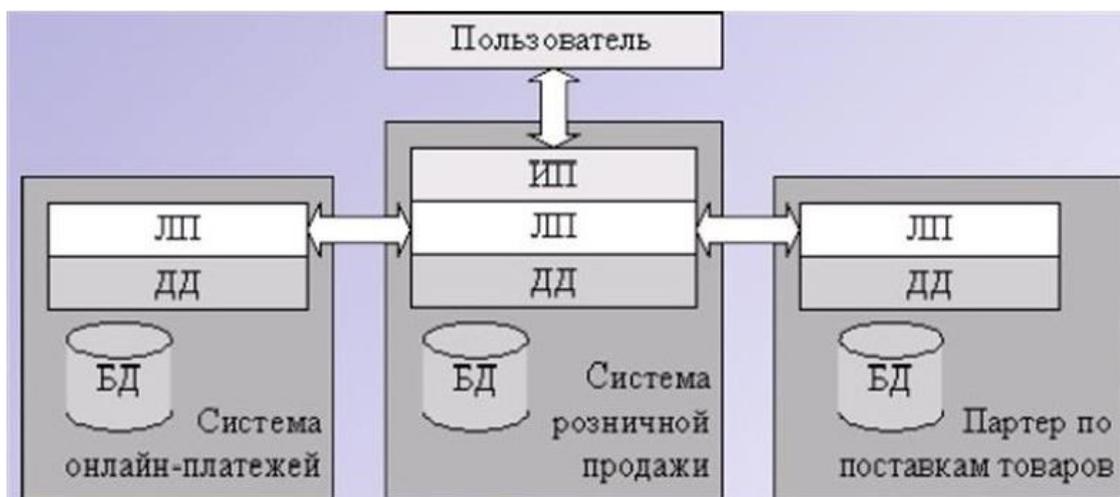


Рисунок 2

В качестве другого примера (рисунок 3) распределенной системы можно привести **сети прямого обмена данными между клиентами** (peer-to-peer networks). Если предыдущий пример имел "древовидную" архитектуру, то сети прямого обмена организованы более сложным образом. Подобные системы являются в настоящий момент, вероятно, одними из крупнейших существующих распределенных систем, объединяющие миллионы компьютеров.

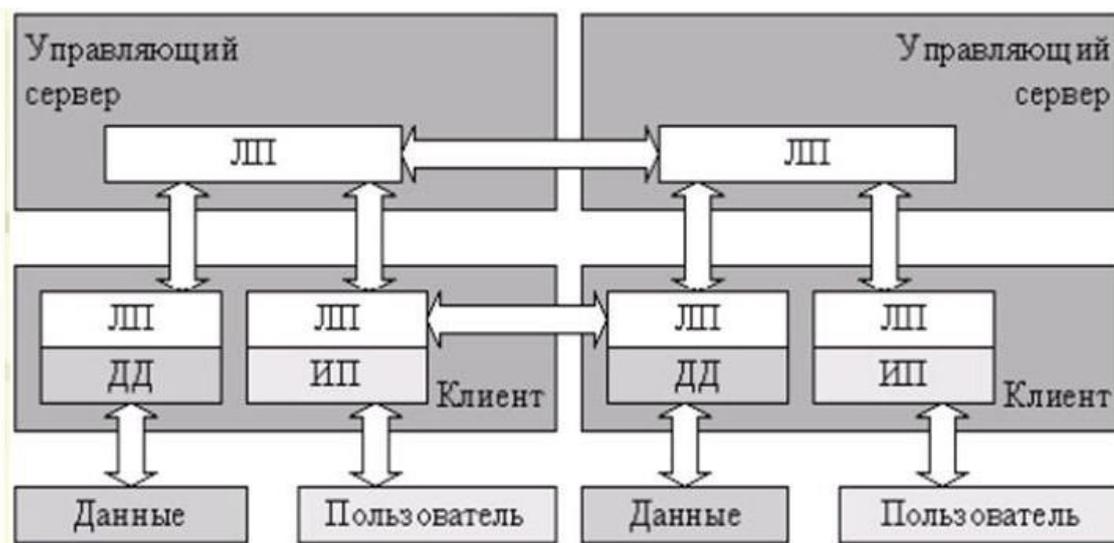


Рисунок 3

Другая популярная архитектура – **Grid**. Грид вычисления – это форма распределённых вычислений, в которой «супер и виртуальный компьютер» представлен в виде кластера соединённых с помощью сети, слабосвязанных компьютеров, работающих вместе для выполнения огромного количества заданий (операций, работ). Эта технология была применена для решения научных, математических задач, требующих для решения значительных вычислительных ресурсов. Грид вычисления используются также и в коммерческой инфраструктуре для решения таких трудоёмких задач как экономическое прогнозирование, сейсмоанализ, разработка и изучение свойств новых лекарств. Если рассмотреть какие же **характеристики** стали присущи к распределённым информационным системам, то можно отметить следующие:

- прозрачность реализации;
- открытость;
- легкая масштабируемость и расширяемость;
- устойчивость к авариям;
- наличие промежуточного уровня.

Первая из характеристик состоит в том, что от пользователей скрыты различия между компьютерами и способы связи между ними. То же самое относится и к внешней организации распределённых систем.

Другой важной характеристикой распределённых систем является способ, при помощи которого пользователи и приложения единообразно работают в распределённых системах, независимо от того, где и когда происходит их взаимодействие.

Распределённые системы должны также относительно легко поддаваться расширению, или масштабированию. Эта характеристика является прямым следствием наличия независимых компьютеров, но в то же время не указывает, каким образом эти компьютеры на самом деле объединяются в единую систему.

Распределённые системы обычно существуют постоянно, однако некоторые их части могут временно выходить из строя. Пользователи и приложения не должны уведомляться о том, что эти части заменены или починены или что добавлены новые части для поддержки дополнительных пользователей или приложений.

Для того чтобы поддержать представление различных компьютеров и сетей в виде единой системы, организация распределённых систем часто включает в себя дополнительный уровень программного обеспечения, находящийся между верхним уровнем, на котором

находятся пользователи и приложения, и нижним уровнем, состоящим из операционных систем.

Соответственно, такая распределенная система обычно называется **системой промежуточного уровня (middleware)**.

Использование протокола TCP/IP посредством сокетов предоставляет стандартный, межплатформенный, но низкоуровневый сервис для обмена данными между компонентами. Для выполнения сформулированных выше требований к распределенным системам функции сеансового и представительского уровня должна взять на себя некоторая **промежуточная среда (middleware)**, называемая так же промежуточным программным обеспечением. Такая среда (рисунок 4) должна помогать разработчикам создавать открытые, масштабируемые и устойчивые распределенные системы.

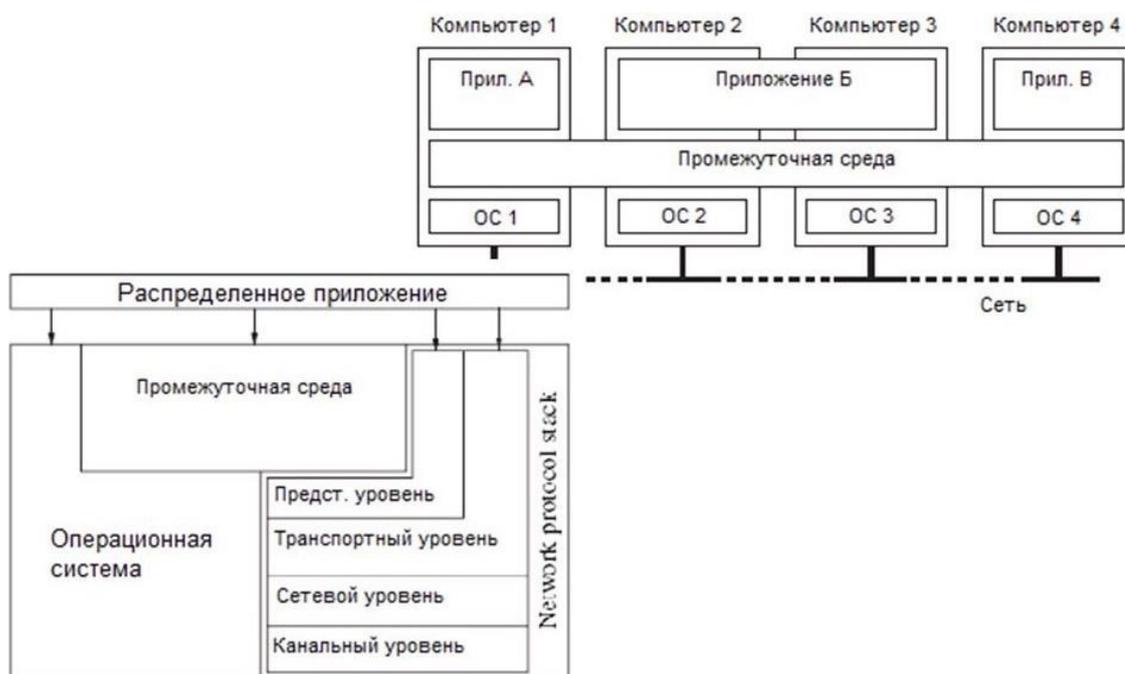


Рисунок 4

Для достижения этой цели промежуточная среда должна обеспечить сервисы для взаимодействия компонент распределенной системы. К таким сервисам относятся:

- обеспечение единого и независимого от операционной системы механизма использования одними программными компонентами сервисов других компонент;
- обеспечение безопасности распределенной системы: аутентификация и авторизация всех пользователей сервисов компоненты и защита передаваемой между компонентами информации от искажения и чтения третьими сторонами;
- обеспечение целостности данных: управление транзакциями, распределенными между удаленными компонентами системами;
- балансировка нагрузки на серверы с программными компонентами;
- обнаружение удаленных компонент.

Чтобы сделать разработку и интеграцию распределенных приложений как можно более простой, основная часть программного обеспечения промежуточного уровня базируется на некоторой модели, или парадигме, определяющей распределение и связь. Относительно простой моделью является **представление всех наблюдаемых объектов в виде файлов**, построенной по принципу распределенной файловой системы (distributed file system). Во многих случаях это программное обеспечение всего на один шаг ушло от сете-

вых операционных систем в том смысле, что прозрачность распределения поддерживается только для стандартных файлов (то есть файлов, предназначенных только для хранения данных). Процессы, например, часто должны запускаться исключительно на определенных машинах. Программное обеспечение промежуточного уровня, основанное на модели распределенной файловой системы, оказалось достаточно легко масштабируемым, что способствовало его популярности.

Другая важная ранняя модель программного обеспечения промежуточного уровня основана на **удаленных вызовах процедур** (Remote Procedure Calls, RPC). В этой модели акцент делается на сокрытии сетевого обмена за счет того, что процессу разрешается вызывать процедуры, реализация которых находится на удаленной машине. При вызове такой процедуры параметры прозрачно передаются на удаленную машину, где, собственно, и выполняется процедура, после чего результат выполнения возвращается в точку вызова процедуры. За исключением, вероятно, некоторой потери производительности, все это выглядит как локальное исполнение вызванной процедуры: вызывающий процесс не уведомляется об имевшем место факте сетевого обмена.

По мере того как все более входит в моду ориентированность на объекты, становится ясно, что если вызов процедуры проходит через границы отдельных машин, он может быть представлен в виде прозрачного обращения к объекту, находящемуся на удаленной машине. Это привело к появлению разнообразных систем промежуточного уровня, реализующих представление о распределенных объектах (distributed objects). Идея распределенных объектов состоит в том, что каждый объект реализует интерфейс, который скрывает все внутренние детали объекта от его пользователя. Интерфейс содержит методы, реализуемые объектом, не больше и не меньше. Все, что видит процесс, – это интерфейс. Когда процесс вызывает метод, реализация интерфейса на машине с процессом просто преобразует вызов метода в сообщение, пересылаемое объекту. Объект выполняет запрашиваемый метод и отправляет назад результаты. Затем реализация интерфейса преобразует ответное сообщение в возвращаемое значение, которое передается вызвавшему процессу. Microsoft DCOM (Distributed COM – распределённая COM) основана на технологии DCE/RPC (разновидности RPC). DCOM позволяет COM-компонентам взаимодействовать друг с другом по сети. Технология DCOM обеспечивает базовые установки безопасности позволяя задавать, кто и из каких машин может создавать экземпляры объекта и вызывать его методы; OMG CORBA (Common Object Request Broker Architecture – общая архитектура брокера объектных запросов) – это технологический стандарт, продвигаемый консорциумом OMG, задачей которого является осуществить интеграцию изолированных систем, дать возможность программам, написанным на разных языках, работающим на разных узлах сети, взаимодействовать друг с другом так же просто, как если бы они находились в адресном пространстве одного процесса; Java RMI (Remote Method Invocation) – программный интерфейс вызова удаленных методов в языке Java.

Как модели могут упростить использование сетевых систем, вероятно, наилучшим образом видно на примере World Wide Web. Успех среды Web в основном определяется тем, что она построена на базе потрясающе простой, но высокоэффективной модели распределенных документов (distributed documents). В модели, принятой в Web, информация организована **в виде документов**, каждый из которых размещен на машине, расположение которой абсолютно прозрачно. Документы содержат ссылки, связывающие текущий документ с другими. Если следовать по ссылке, то документ, с которым связана эта ссылка, будет извлечен из места его хранения и выведен на экран пользователя. Концепция документа не ограничивается исключительно текстовой информацией. Например, в Web поддерживаются аудио- и видеодокументы, а также различные виды документов на основе интерактивной графики.

Итак, еще раз перечислим и кратко охарактеризуем модели ПУ:

**1 Распределенная файловая система** обозначение – «Distributed File System»; достоинство – Относительно простая модель; цель: обеспечить прозрачный доступ удаленных пользователей к файловой системе; пример: NFS.

**2 Удаленный вызов процедур** обозначение – «Remote Procedure Call (RPC)»; цель: обеспечение прозрачности удаленного исполнения кода; особенности функционирования:

- реализация процедуры находится на сервере;
- клиент передает параметры процедуры;
- сервер исполняет процедуру и возвращает результат – некоторая потеря производительности;

– весь сетевой обмен скрыт от процесса. **3 Распределенные объекты** обозначение – «Distributed Objects»;

особенности функционирования:

- каждый объект реализует интерфейс;
- интерфейс содержит методы, реализуемые объектом;
- процесс видит только интерфейс;
- наиболее популярные технологии распределенных объектов в настоящее время:
- Microsoft DCOM;
- OMG CORBA;
- Java RMI.

**4 Распределенные документы** обозначение «Distributed Documents»;

– реализация: World Wide Web – цель: Прозрачность размещения документов; особенности функционирования:

- ссылки связывают документы;
- содержимое не ограничено текстовой информацией.

**Кратко формулируем задачи промежуточного уровня:**

- обеспечение интероперабельности;
- обеспечение безопасности;
- обеспечение целостности данных;
- балансировка нагрузки;
- обнаружение удаленных компонент.

Чтобы достигнуть цели своего существования – улучшения выполнения запросов пользователя – распределенная информационная система должна удовлетворять некоторым необходимым требованиям.

Можно сформулировать следующий **набор требований**, которым в наилучшем случае должна удовлетворять РИС.

**Открытость.** Все протоколы взаимодействия компонент внутри распределенной системы в идеальном случае должны быть основаны на общедоступных стандартах. Это позволяет использовать для создания компонент различные средства разработки и операционные системы. Каждая компонента должна иметь точную и полную спецификацию своих сервисов. В этом случае компоненты распределенной системы могут быть созданы независимыми разработчиками. При нарушении этого требования может исчезнуть возможность создания распределенной системы, охватывающей несколько независимых организаций.

**Масштабируемость.** Масштабируемость вычислительных систем имеет несколько аспектов. Наиболее важный из них – возможность добавления в распределенную систему новых компьютеров для увеличения производительности системы, что связано с понятием балансировки нагрузки (load balancing) на серверы системы. К масштабированию относятся так же вопросы эффективного распределения ресурсов сервера, обслуживающего запросы клиентов.

**Поддержание логической целостности данных.** Запрос пользователя в распределенной системе должен либо корректно выполняться целиком, либо не выполняться вообще. Ситуация, когда часть компонент системы корректно обработали поступивший запрос, а часть – нет, является наихудшей.

**Устойчивость.** Под устойчивостью понимается возможность дублирования несколькими компьютерами одних и тех же функций или же возможность автоматического распределения функций внутри системы в случае выхода из строя одного из компьютеров. В идеальном случае это означает полное отсутствие уникальной точки сбоя, то есть выход из строя одного любого компьютера не приводит к невозможности обслужить запрос пользователя.

**Безопасность.** Каждый компонент, образующий распределенную систему, должен быть уверен, что его функции используются авторизованными на это компонентами или пользователями. Данные, передаваемые между компонентами, должны быть защищены как от искажения, так и от просмотра третьими сторонами.

**Эффективность.** В узком смысле применительно к распределенным системам под эффективностью будет пониматься минимизация накладных расходов, связанных с распределенным характером системы. Поскольку эффективность в данном узком смысле может противоречить безопасности, открытости и надежности системы, следует отметить, что требование эффективности в данном контексте является наименее приоритетным. Например, на поддержку логической целостности данных в распределенной системе могут тратиться значительные ресурсы времени и памяти, однако система с недостоверными данными вряд ли нужна пользователям.

Классическим примером системы, в значительной мере отвечающей всем представленным выше требованиям, является система преобразования символьных имен в сетевые IP-адреса (DNS). Система имен – организованная иерархически распределенная система, с дублированием всех функций между двумя и более серверами.

**Повышение отношения производительности к затратам.** Любая задача может быть разделена между различными компьютерами в распределенной системе. Такая конфигурация обеспечивает лучшее соотношение производительности к стоимости системы. Это особенно актуально для конфигурации «сеть рабочих станций» (NOW).

**Масштабируемость.** Компьютеры, как правило, подключены к глобальной компьютерной сети, поэтому установка новых компьютеров непосредственно не создает узких мест в компьютерной сети.

**Модульность и дополнительная расширяемость.** Гетерогенные единицы могут быть добавлены в систему без снижения производительности, так как используется промежуточный уровень взаимодействия. Аналогично, существующие единицы могут быть легко заменены новыми.

Итак, можно считать, что предметная область обозначена и выделены проблемные вопросы, которые решаются специалистами информационных технологий. Исследования в области распределенных систем достаточно сложные, поэтому за выдающуюся статью по распределенным вычислениям ежегодно вручается «Премия Дейкстры» (<http://www.podc.org/dijkstra/>).

На данном этапе изучения дисциплины можно запомнить определение:

**РИС** – это совокупность автономных компьютеров, взаимодействующих через компьютерную сеть и промежуточную среду, которая позволяет компьютерам координировать свою деятельность и предоставлять доступ к ресурсам системы так, что пользователям система представляется единой и целостной.

**Распределенная система обладает следующими свойствами:**

- отсутствие общей физической шины;
- отсутствие общей памяти;

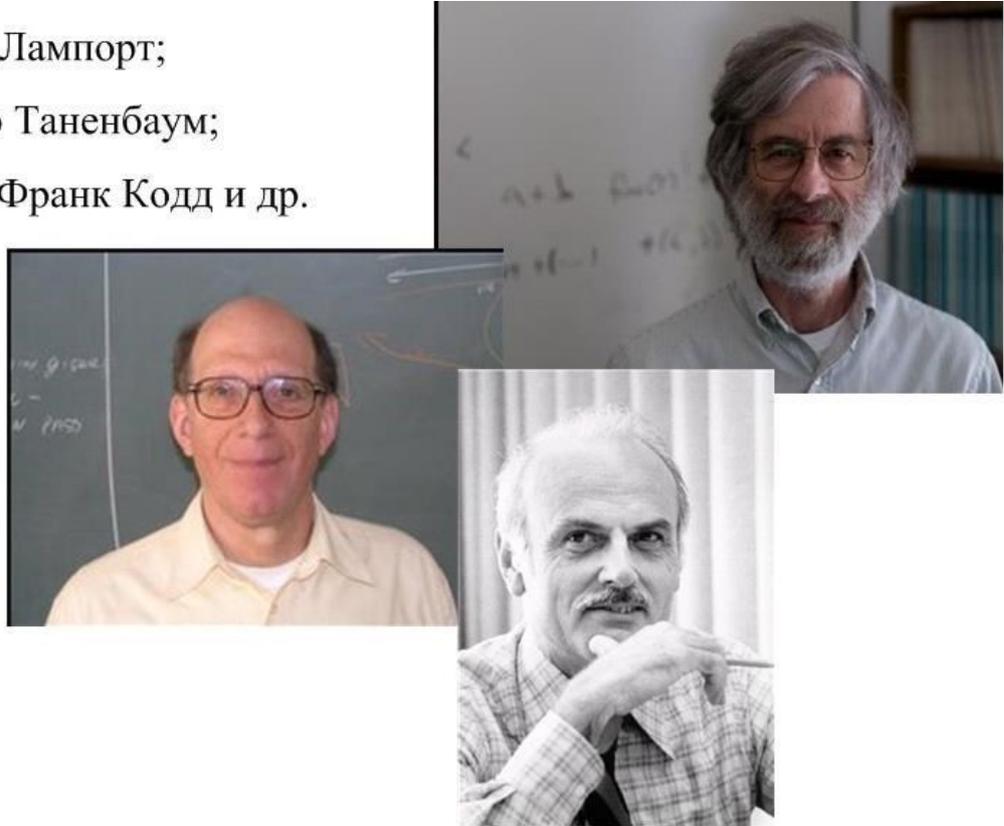
- географическое распределение;
- автономность и гетерогенность.

Известных исследователей в области распределенных информационных систем и баз данных нужно знать:

Лесли Лампорт;

Эндрю Таненбаум;

Эдгар Франк Кодд и др.



Чтобы подробно разобраться со всеми проблемами, обозначенными в теории и практике РИС, ну проштудировать многотомное собрание сочинений. В данном учебном пособии обозначим рамки и направления изучения распределенных информационных систем. Объектом изучения обозначим корпоративную распределенную информационную систему организации, предприятия (рисунок 5).

Типовой пример РИС такой сети можно представить схемой компьютерного интегрированного полиграфического производства (рисунок 6).

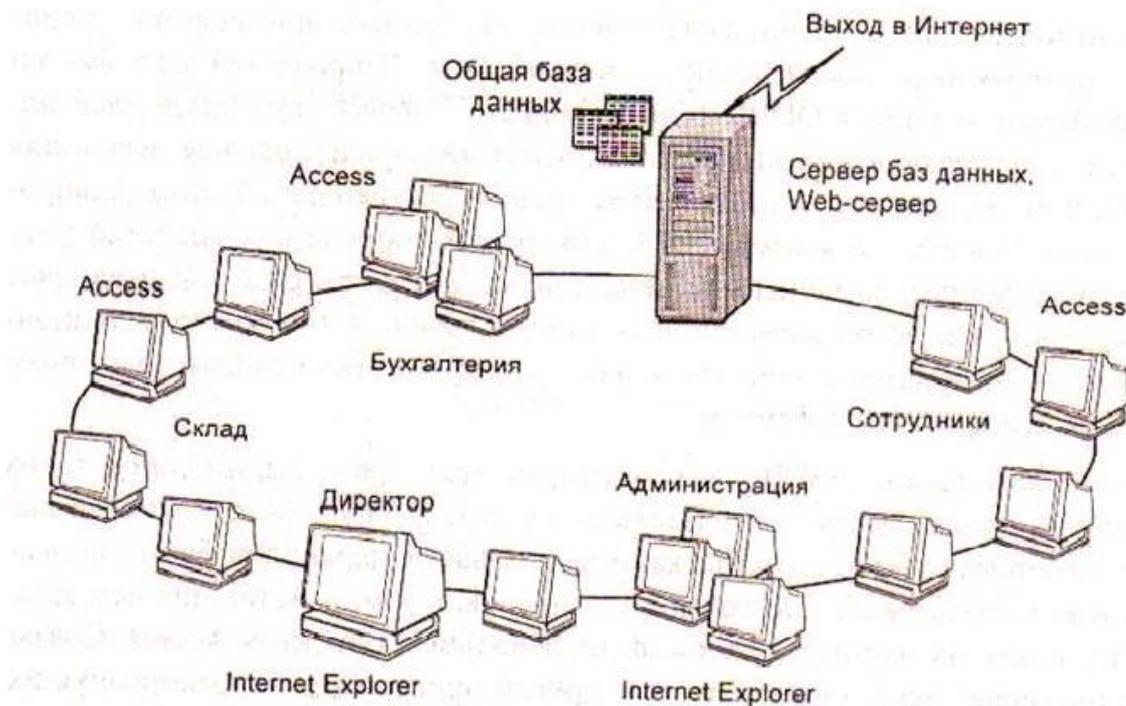


Рисунок 5 – Корпоративная сеть с SQL – и WEB – серверами

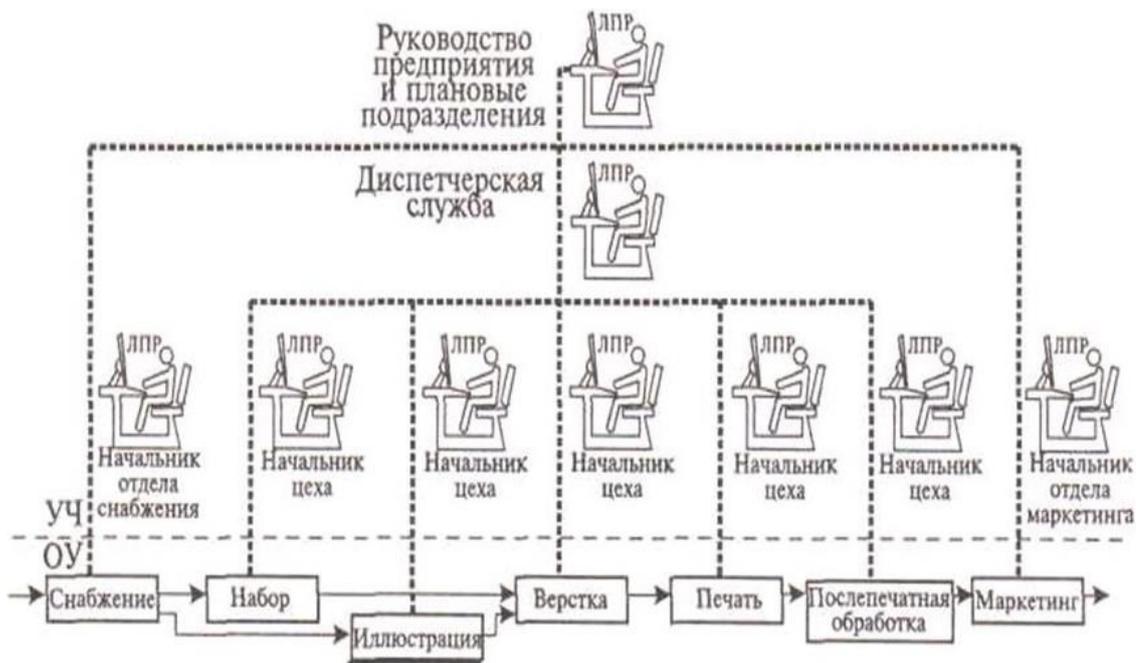


Рисунок 6

УЧ – управляющая часть; ОУ – объект управления

Лекционный материал учебного пособия можно сгруппировать по трем разделам. Первый раздел предназначен для изучения основных составляющих автоматизированной информационной системы (АИС). Проведем маленький эксперимент. Ответьте на два простых вопроса: что такое данные? и что такое информация? Большинство отвечающих относятся к одному варианту ответа: данные – это информация; информация – это данные. Круг замыкается! Простые и обыденные понятия, но мы с ними свыклились и не вдаемся в суть

определений. Поэтому в первом разделе учебного пособия уделено основное внимание разъяснению терминов, определений объектов, составляющих информационную систему (ИС), концепции построения баз данных как основы ИС.

Второй раздел раскрывает содержание распределенной обработки информации в РИС, определяет концепцию построения распределенных баз данных (хранилищ и витрин данных) в структурах OLTP и OLAP. Важно уяснить как распределяется презентационная логика, логика аналитической обработки, логика баз данных в РИС.

Третий раздел посвящен изучению технологий и методик моделирования ИС. Должен быть понятен алгоритм проектирования РИС с учетом требований CASE – и CALS – технологий, используемых за рубежом и российских аналогов. Специалист по РИС должен знать весь перечень стадий и этапов проектирования РИС. Для проектирования РИС необходимо уметь пользоваться современными инструментальными средствами разработки. Заканчивается курс лекций практическими примерами проектирования и разработки РИС в современной инструментальной среде Visual Studio 2005, 2010, при этом проектировщику достаточно знать основы событийно управляемого программирования и разработки Windows-приложения на языке C#. Разработка базы данных демонстрируется на примере СУБД SQL Server 2005, 2008. Важно и необходимо уяснить, как проектируется и создается удаленный запрос с клиентского приложения на сервер базы данных с помощью функций пользователя или хранимых процедур. Специалист по РИС должен в обязательном порядке выполнять фрагментацию и локализацию данных в распределенной базе данных.

Итак, предметная область изучения обозначена. Содержание предмета изучения изложено в семи лекциях.

# 1 Лекция 1. Назначение и основные компоненты информационной системы

Первая лекция предназначена для общего введения в теорию и практику построения информационных систем на основе баз данных, рассмотрения и обсуждения основных терминов и определений компонентов информационных систем. Лекция посвящена рассмотрению общетеоретических вопросов, касающихся системы баз данных как информационной системы, использующей информационные технологии по обработке информации. Некоторые термины в системе баз данных не имеют четко определенных государственным стандартом определений, что составляет основную проблему однозначного понимания некоторых определений и понятий баз данных как информационной системы. Для решения данной основной проблемы обсудим различные подходы и общепринятые соглашения (нотации), существующие в современной научной литературе.

Ключевые слова: система, управление, информация, данные, базы данных, система управления базами данных (СУБД), уровни представления баз данных, модели данных.

## **Введение.**

В прошлом информация считалась сферой бюрократической работы и ограниченным инструментом для принятия решений. Сегодня информацию рассматривают как один из основных ресурсов развития общества, а информационные системы и технологии как средство повышения производительности и эффективности деятельности людей.

Наиболее широко информационные системы и технологии используются в производственной, управленческой и финансовой деятельности, хотя начались подвижки в сознании людей, занятых и в других сферах, относительно необходимости их внедрения и активного применения.

Основные идеи современной информационной технологии базируются на концепции *баз данных* (БД), ранее упоминаемых достаточно часто без какого-либо их детального пояснения.

Согласно данной концепции основой информационной технологии являются *данные*, организованные в БД, адекватно отражающие реалии действительности в той или иной предметной области и обеспечивающие пользователя актуальной информацией в соответствующей предметной области.

Развитие современного промышленного производства и бизнеса невозможно без создания автоматизированных информационных систем (АИС), одно из назначений которых – предоставление пользователю достоверной информации, необходимой для принятия оптимального решения. В настоящее время ни одна из задач управления производством и бизнесом не должна выполняться без применения автоматизированных информационных систем. Сегодня мы должны рассматривать любую деятельность любого специалиста как некоторую систему принятия решений, поэтому специалисту и нужна достоверная информация. Таким образом, одной из важнейших функций информационной системы является информационное обеспечение процесса управления.

Итак, что же такое *База данных* и *Система управления базами данных*!

К сожалению, в большинстве книг по этому направлению информационных технологий нет достаточно четких определений. Рассмотрим и обсудим термины и определения, касающиеся баз данных.

## **Понятие об информационных системах.**

Под *системой* понимают любой объект, который одновременно рассматривается и как единое целое, и как объединенная в интересах достижения поставленных целей совокупность разнородных элементов. Системы значительно отличаются между собой как по составу, так и по главным целям.

Приведем примеры нескольких систем, состоящих из разных элементов и направленных на реализацию различных целей (таблица 1.1).

Таблица 1.1

Система	Элементы системы	Главная цель системы
Организация	Люди, оборудование, материалы, здания и др.	Производство
Компьютер	Электронные и электромеханические элементы, линии связи и др.	Обработка данных
Телекоммуникационная система	Компьютеры, модемы, кабели, сетевое программное обеспечение и др.	Передача информации
Информационная система	Компьютеры, компьютерные сети, люди, информационное и программное обеспечение	Производство профессиональной информации

В информатике понятие “*система*” широко распространено и имеет множество смысловых значений. Чаще всего оно используется применительно к набору технических средств и программ. Системой может называться аппаратная часть компьютера. Системой может также считаться множество программ для решения конкретных прикладных задач, дополненных процедурами ведения документации и управления расчетами.

Добавление к понятию “*система*” слова “*информационная*” отражает цель ее создания и функционирования. Информационные системы обеспечивают сбор, хранение, обработку, поиск и выдачу информации, необходимой в процессе принятия решений задач в любой предметной области. Появление электронных вычислительных машин и персональных компьютеров предопределило создание и внедрение автоматизированных информационных систем (АИС), которые значительно повысили производительность и результативность информационных технологий по обработке и выдаче информации.

В качестве основного классификационного признака АИС целесообразно рассматривать особенности автоматизируемой профессиональной деятельности процесса переработки входной информации для получения требуемой выходной информации, в котором АИС выступает в качестве инструмента должностного лица или группы должностных лиц, участвующих в управлении организационной системой.

В соответствии с предложенным классификационным признаком можно выделить следующие классы АИС (рисунок 1.1):

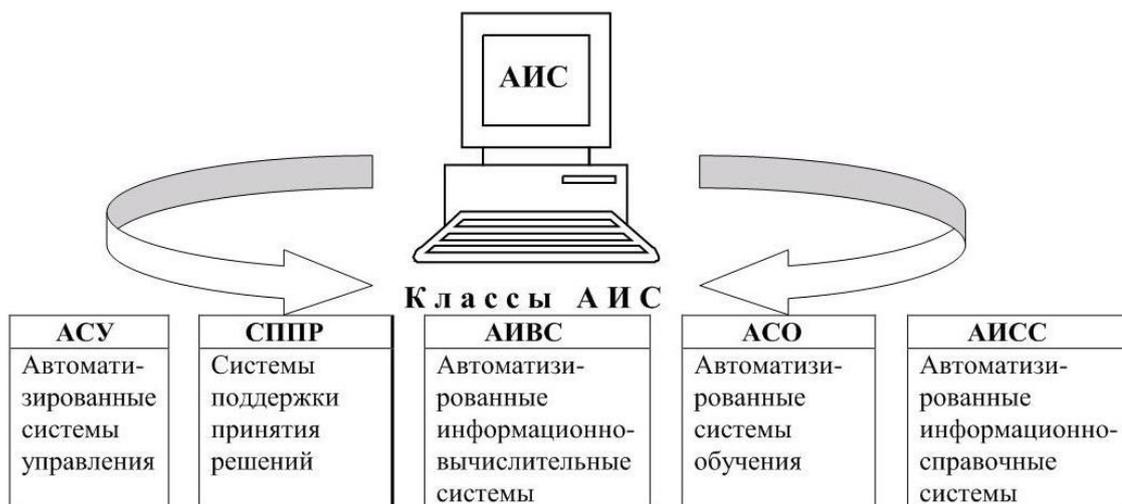


Рисунок 1.1 – Классификация АИС

АИС – взаимосвязанная совокупность средств, методов и персонала, используемых для хранения, обработки и выдачи информации в интересах достижения поставленной цели.

Современное понимание АИС как системы предполагает использование в качестве основного технического средства переработки информации персонального компьютера. Кроме того, техническое воплощение информационной системы само по себе ничего не будет значить, если не будет учтена роль человека, для которого предназначена производимая информация и без которого невозможно ее получение и представление.

Структуру АИС составляет совокупность отдельных ее частей, называемых *подсистемами*.

Подсистема – это часть системы, выделенная по какому-либо признаку.

Общую структуру информационной системы можно рассматривать как совокупность подсистем независимо от сферы применения. В этом случае говорят о *структурном признаке* классификации, а подсистемы называют *обеспечивающими*. Таким образом, структура любой информационной системы может быть представлена совокупностью обеспечивающих подсистем (рисунок 1.2).

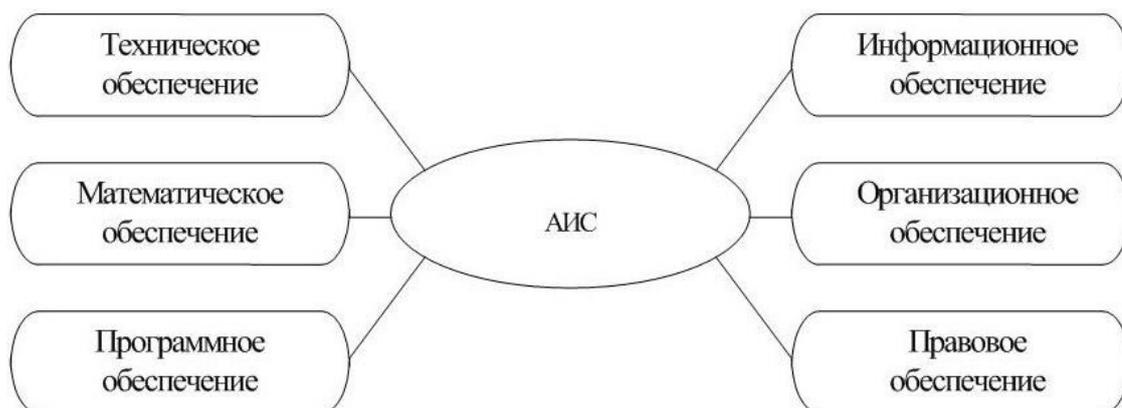


Рисунок 1.2 – Структура АИС как совокупность обеспечивающих подсистем

Среди обеспечивающих подсистем обычно выделяют информационное, техническое, математическое, программное, организационное и правовое обеспечение. В понятии АИС присутствуют три очень важных, емких, ключевых, философско-методологических и специально-научных понятия:

- система;
- управление;
- информация.

И поэтому, чтобы разобраться в АИС, необходимо прежде всего выяснить:

- что такое система;
- что такое управление;
- что такое информация.

**Система** (от греч. SYSTEMA –целое, составленное из частей соединением) – это совокупность элементов, взаимосвязанных друг с другом, образующая определенную целостность, единство.

Введем набор понятий, связанных с использованием слова “система”.

**Элемент** – некоторый объект (материальный, энергетический, информационный), обладающий определенным функциональным назначением, отличающимся от назначения системы.

Введем обозначения:

$M$  – элемент;

$\{M\}$  – совокупность элементов;

$M \in \{M\}$  – принадлежность элементов совокупности.

Вопрос. Сколько или какое количество элементов необходимо, чтобы их совокупность стала системой? Древние философы спорили – сколько нужно камней, сложенных вместе, чтобы они образовали кучу. Вся доступная разуму человека природа состоит из систем, которые могут быть различны по масштабам: от бесконечно больших (галактика) до бесконечно малых (атом); различны по природе: материальные, энергетические, информационные. Практически любой объект с определенной точки зрения может рассматриваться как система, если совокупность элементов обладать двумя признаками: а) связями, которые позволяют посредством переходов по ним от элемента к элементу соединить два любых элемента совокупности (рисунок 1.3).

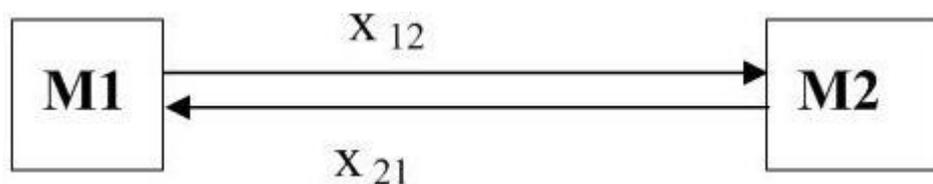


Рисунок 1.3 – Связь двух элементов

$X_{12}$  – воздействие элемента  $M1$  на элемент  $M2$ ;

$X_{21}$  – воздействие  $M2$  на  $M1$ .

б) Свойством (назначением, функцией), отличным от свойств отдельных элементов совокупности.

По степени автоматизации решения своих функциональных задач системы могут быть:

- ручные, т.е. без средств автоматизации;
- автоматизированные, т.е. со средствами автоматизации при участии человека;
- автоматические, т.е. типа «автомат» без участия человека.

Дадим определение АСУ согласно ГОСТ 19675-74:

АСУ – это человеко-машинная система, обеспечивающая автоматизированный сбор и обработку информации, необходимой для оптимизации управления в различных сферах человеческой деятельности.

Большой системой называют систему, включающую большое количество однотипных элементов и однотипных связей. Пример: гирлянда, трубопровод.

Сложной системой называют систему, состоящую из элементов разных типов и обладающую разнородными связями между ними. Пример: космический корабль, автомобиль, морское судно и т.п.

Простая система – это система, состоящая из небольшого числа элементов и не имеющая разветвленной структуры (нельзя выявить иерархические уровни).

Структура системы – совокупность внутренних, устойчивых связей между элементами системы, определяющая ее основные свойства.

Простейшие типы структур систем показаны на рисунке 1.4.

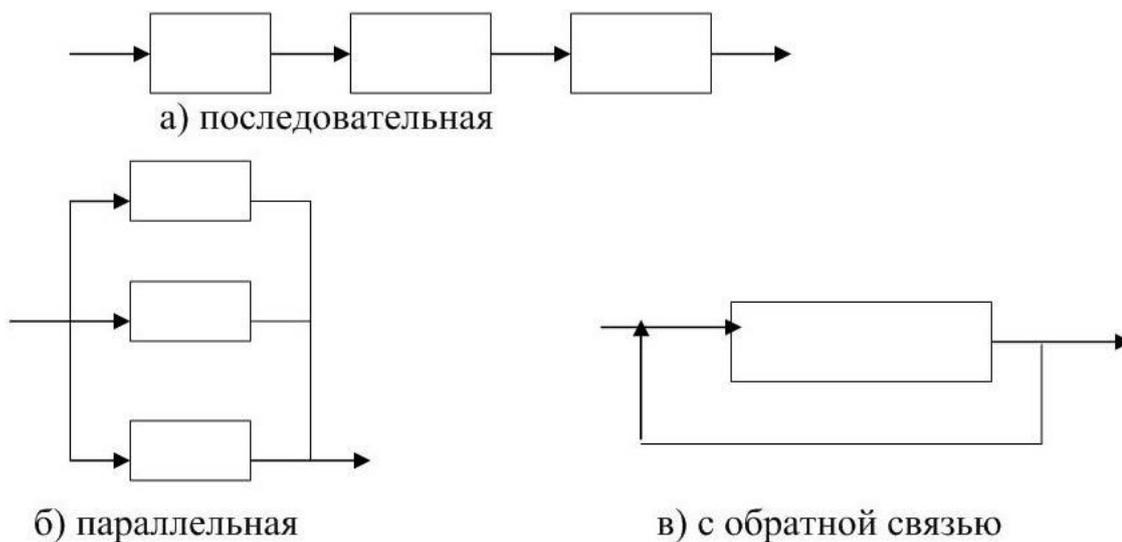


Рисунок 1.4 – Типы структур систем

Иерархические структуры – это структуры с наличием подчиненности (рисунок 1.5).

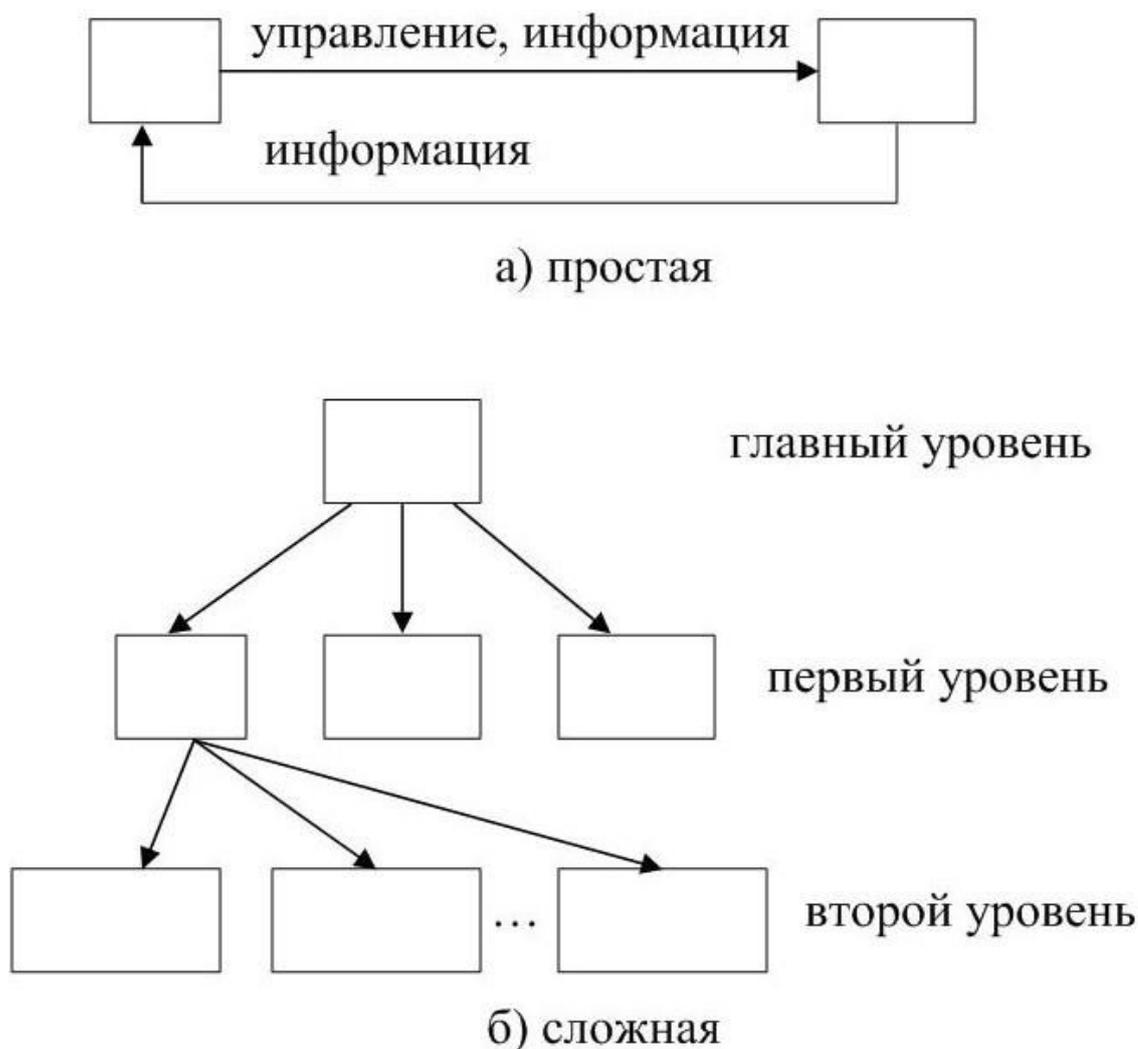


Рисунок 1.5 – Иерархические структуры

С понятием структура тесно связан термин «декомпозиция» – это деление системы на части, удобное для каких-то операций.

Примеры: автомобиль состоит из систем смазки, охлаждения, питания и др. Любая книга имеет содержание (оглавление).

Различают также статистические и динамические системы. Состояние статической системы с течением времени остается постоянным, динамические системы, наоборот, изменяют свое состояние во времени.

Динамические системы разделяют на детерминированные, т.е. полностью определенных в любой момент времени, и вероятностные (стохастические).

По характеру взаимодействия системы и внешней среды различают закрытые и открытые системы.

Системы обладают свойством целостности – это принципиальная несводимость свойств системы к сумме свойств составляющих ее элементов и, в то же время, зависимость свойств каждого элемента от его места и функции внутри системы.

Группа элементов системы, описываемая только своими входами и выходами и обладающая определенной целостностью, называется модулем (рисунок 1.6).

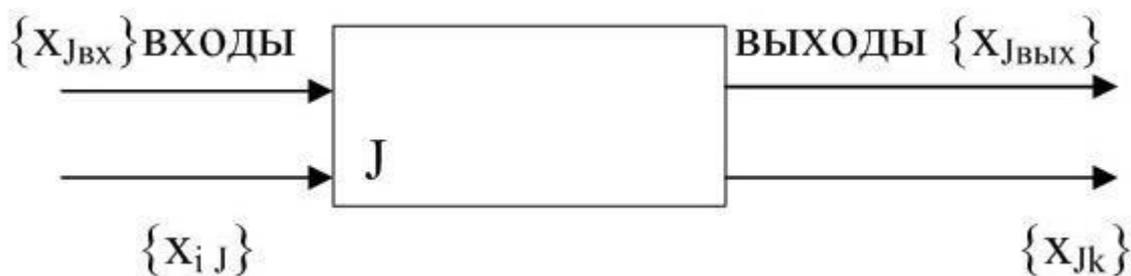


Рисунок 1.6 – Модуль J

$X_{JВХ}$  – внешние (не от системы) воздействия на элементы модуля;

$X_{iJ}$  – связи от других элементов системы;

$X_{JВЫХ}$  – воздействия на выходе системы;

$X_{Jk}$  – воздействия на элементы системы.

В различных разделах науки и техники понятие модуль имеет различные синонимы:

– в технике: агрегат, блок, узел, механизм;

– в программировании: программа, программный модуль, логический блок;

– в организации и управлении – комиссия, подразделение. Входы и выходы можно

интерпретировать как:

- 1) сигнал – отклик;
- 2) воздействие – реакция;
- 3) запрос – ответ;
- 4) аргумент – решение и др.

Перейдем к анализу понятия управления в системах. Процессы управления протекают повсеместно, ежедневно, ежечасно, ежесекундно и охватывают буквально все стороны и моменты человеческой деятельности.

В науке управление стало изучаться и исследоваться сравнительно недавно, и связано это было прежде всего с именем всемирного ученого Норберта Винера (1894-1964). В 1948 году в США и Европе вышла книга «Кибернетика, или Управление и связь в животном и машине», ознаменовавшая своим появлением рождение нового научного направления – кибернетика.

Кибернетические исследования заключаются в изучении общих свойств процессов управления и систем управления в живых и неживых системах.

Направления исследований кибернетики весьма разнообразны:

– теоретические (теория управления, теория информации, информатика);

– механические (исследование и проектирование АИС);

– биологические (нейрокибернетика – обработка информация в нервных тканях человека;

– бионика – искусственные органы;

– гомеостатика – автоматы–роботы;

– экономические;

– социальные.

В современном обществе кибернетика уступила пальму первенства информатике, но значение кибернетики как науки об общих принципах управления в живых и неживых системах, в искусственных системах и в обществе сохраняются и сейчас.

Итак, что такое управление?

Под управлением в самом общем виде понимают совокупность действий, осуществляемых человеком, группой людей или автоматическим устройством.

Эти действия направлены на поддержание или улучшение работы управляемого объекта в соответствии с имеющейся программой (алгоритмом функционирования) или с целью управления.

Управлять – это значит влиять на ход какого-либо процесса или состояния некоторого объекта и его положения в пространстве.

При изучении структуры процессов управления выделяют три основные проблемы. Первая группа задач – это задачи по изучению свойств объектов управления и их характеристик. Вторая группа задач – анализ различных видов органов управления, форм и способов выработки управляющих воздействий, формирование и согласование множества критериев качества (целей) управления, обеспечение устойчивости функционирования управляющих органов. Третья группа задач связана с передачей информации, ее восприятием, оценкой количества, кодированием, защитой от помех, достоверности. Все это относится к той части кибернетики, которую называют теорией информации.

Что же характерно для любого процесса управления?

1 Есть объект, которым управляют (рисунок 1.7). На него через исполнительный орган подается управляющее воздействие. Информация о состоянии объекта по цепи обратной связи поступает на управляющий орган, который выдает управляющие сигналы на исполнительный орган. Очевидно, что самым сложным с точки зрения функционирования, является управляющий орган, так как на него возложено одновременное и оптимальное управление, прием информации от внешней среды, прием информации от объекта.

2 В управлении всегда решается задача с определенной целью (целевой функцией). Решение на выдачу того или иного управляющего воздействия принимается согласно теории принятия решения. Как правило, целевую функцию в моделях описывают математически, задают параметры и ограничения, временные характеристики. Тогда оптимальность управления достигается в обеспечении экстремума выходных характеристик при определенных входных воздействиях.

3 Без информации нет управления.

Информация нужна и внешняя, и внутренняя (о состоянии объекта). Это информация полезная. Есть и информация в виде возмущающих воздействий (неверная, ложная, ненужная, избыточная и т.д.).

Циркуляция информации происходит в информационной системе (ИС), которая является как бы кровеносными сосудами в структуре элементов управления.

Основу информационной системы составляет база данных, в которой хранится большая по объему информация о какой-либо предметной области.

В неавтоматизированном варианте примером информационной системы является книжный каталог библиотеки. В настоящее время все большее развитие приобретают автоматизированные информационные системы. Ее элементы показаны на рисунке 1.7. Пользователь из органа управления выдает запрос в ИС. Этот запрос расшифровывается, формируется поисковое предписание (или поисковый образ), представляющее задание для процедуры поиска в базе данных. Поиск в базе данных осуществляется блоком поиска. Найденная информация выдается потребителю в удобной для него форме. АИС играют большую роль в инфраструктуре общества и являются одним из обязательных звеньев новых информационных технологий (НИТ).

Проанализируем понятие ИНФОРМАЦИЯ.

Отличительной чертой человеческого общества является то, что в течение длительного времени основным предметом труда оставались материальные объекты. Воздействуя на них, человек добывал себе средства к существованию, и на протяжении многих веков решалась задача усиления мускульных возможностей человека с помощью различных инструментов, агрегатов и машин. На это была направлена механизация производства, которая стала интен-

сивно внедряться в наличие двадцатого века. Развитие человеческого общества практически на всех этапах проходило на основе технического прогресса. Это и овладение огнем, и использование паровых машин, и проникновение в тайны атомной энергии, и т.п.

Повышению производительности труда способствовала автоматизация. В процессе формирования трудовых коллективов возникла необходимость обмена знаниями. Первоначально знания передавались устно из поколения в поколение.

Появление письменности позволило по-новому показать накопленные знания: представить их в виде информации.



Рисунок 1.7 – Структурная схема АИС

### Понятие о данных и информации.

В этом разделе рассмотрим и обсудим два основных понятия – это данные и информация, их взаимосвязь. Еще раз подчеркнем, что с философской точки зрения эти понятия однозначно и до конца не определены. Чтобы отличить базу данных от овощной базы необходимо прежде всего дать определение данных.

Данными называют описание в сознании человека предметов, событий и явлений окружающего мира. Существуют три основных формы описания и дальнейшего представления данных:

- символьная;
- текстовая;
- графическая.

*Символьная форма*, основанная на использовании символов – букв, цифр, знаков, является наиболее простой, но она практически применяется только для передачи несложных сигналов и различных событий. (Например – сигналы светофора).

Более сложной является *текстовая форма*, в которой, как и в предыдущей форме, используются символы – буквы, цифры, математические знаки. Однако информация заложена не только в этих символах, но и в их сочетании, порядке следования. Удобство текстовой информации обусловлено взаимосвязью текста и речи человека.

Самой емкой и сложной является *графическая форма* представления информации. К этой форме относятся виды природы, фотографии, чертежи, схемы, рисунки.

Говоря о формах информации важно еще раз подчеркнуть свойство нематериальной информации – для ее существования обязательно должен быть какой-либо материальный объект: свет, воздух, вода, электрический ток, эфир электромагнитных колебаний и т.д.

Итак, носителем информации может быть как непосредственно наблюдаемый физический объект, так и энергетический субстрат. В последнем случае информация представлена в виде сигналов световых, звуковых, электрических и т.д. При отображении на носителе информация кодируется, т.е. ей ставится в соответствие форма, цвет, структура и другие параметры элементов носителя.

Примеры:

1 Почему человек различает цвета воспринимаемого изображения?

Потому, что простой белый цвет имеет частотные составляющие спектры электромагнитных колебаний для цветов: красного, оранжевого, желтого, зеленого, голубого, синего, фиолетового. Сетчатка глаз способна различать частотные спектры цветности и сообщать сведения в мозг человека.

2 Книга – носитель кодированной последовательности букв, цифр, символов, графики. Читая книгу, мы как раз и воспринимаем информацию, записанную на ее страницах, в виде кодовых комбинаций (слов), состоящих из последовательности символов (букв, цифр) принятого алфавита. То же самое можно сказать и относительно информации, сообщаемой в процессе устной речи.

В теории информации особого внимания заслужила наиболее стандартная и единая форма представления информации – двоичная форма. Она заключается в записи любой информации в виде последовательности только двух символов: 1, или «да», или «истина»; 0, или «нет», или «ложь». В ЭВМ эти символы обозначаются наличием либо отсутствием в рассматриваемой точке электрического или магнитного импульса. В этом случае реквизитом информации, т.е. самой малой порцией информации (меньше не может быть) является ответ на любой вопрос в виде «да» или «нет». Эта порция определяет единицу измерения информации, называемую «битом». Последовательность битов может иметь различную разрядность. Запись нулей и единиц производится по правилам кодирования, используемых в ЭВМ.

Поясняя определение данных, мы произвольно стали использовать термин информация. Уясним взаимосвязь этих двух терминов.

Мы с вами уяснили, что информация – это первичное понятие, точного определения которого не существует.

Существует четыре основных направлений толкования термина информация:

1 Информация – это смысл полученного сообщения, его интерпретация. Пример учителя и ученика. Учитель имеет информацию о предмете. С помощью сообщений (рассказ с показом, демонстрация) передает ученику сведения, данные. Ученик получает сообщения и усваивает их. То, что он понял со своей точки зрения и есть информация. Степень адекватности реальному образу проверяется учителем контрольным опросом. В этом отношении компьютер никогда не оценивает смысл информации, ему все равно с какими данными работать. Только человек имеет возможность получить информацию на основе данных компьютера.

2 Информация – это как содержание сообщений, так и само сообщение, данные. В этом смысле примером может служить книга с ценными для потребителя сведениями, газета, кодограмма и т.п.

3 Некоторые ученые и, прежде всего, философы считают, что информация – это третья составляющая основ мироздания (материя, энергия и информация).

Н. Винер в одной из работ написал: «Информация и есть информация, а не материя и энергия».

4 В математической теории информации понятие информация определяется только для случайных событий. В этом отношении информация – это то, что уменьшает неопределенность события.

Приведем пример: Компьютер с помощью генератора случайных чисел выдал число от 1 до 16. Наша задача угадать это число. Мы задаем вопросы компьютеру, а он отвечает «да» (истина, 1) или «нет» (ложь, 0). За какое минимальное количество вопросов можно отгадать число? Сколько нужно информации, чтобы угадать загаданное число? Неопределенность равна 16. Первый вопрос: задуманное число меньше 8? Ответ «да» или «нет» уменьшает неопределенность в два раза и мы получаем информацию, равную одному биту. Если число находится в пределах от 1 до 8, то мы задаем вопрос: число меньше 4? Получаем ответ и еще один бит информации и т.д. Итого, количество информации, необходимое для угадывания числа равно 4 битам.

Подведем итоги и решим проблемный вопрос в однозначном понятии информации:

1 Под словом «информация» (в переводе с латинского «Information») понимается разъяснение, изложение, чего-либо, сообщение о чем-либо.

2 Ответьте на вопрос – материальна или нематериальна информация? Будем понимать так, информация не материальна, но информация является свойством материи и не может существовать без своего материального носителя – средства переноса ее в пространстве и во времени.

Рассмотрим взаимосвязь двух понятий – данных и информации. Если рассматривать процесс передачи данных от источника данных до потребителя (рисунок 1.8), то можно сделать вывод о том, что источник в виде базы данных содержит большое количество различных и неупорядоченных данных, а потребителю информации нужна определенная и необходимая ему информация о конкретной предметной области.



Рисунок 1.8 – Взаимосвязь данных и информации

Исходя из данной взаимосвязи мы можем дать свое определение информации. Информация – это необходимые для получателя данные, переданные по каналу связи от источника данных своевременно и достоверно.

### **Количество и качество информации – как их анализировать?**

В предыдущем разделе мы дали определение информации исходя из взаимосвязи данных и информации, при этом указали два показателя качества информации – достоверность и своевременность. Возникает проблемный вопрос – как анализировать и оценивать качество и количество информации.

В свете идей науки о знаковых системах – семиотики адекватность информации, т.е. соответствие содержания образа отображаемому объекту, может выражаться в трех формах: синтаксический, семантический и прагматический.

Синтаксическая адекватность связана с воспроизведением формально-структурных характеристик отражения независимо от смысловых и потребительских (полезностных) параметров объекта. На синтаксическом уровне учитываются тип носителя и способ представления информации, скорость ее передачи и обработки, размеры кодов представления информации, надежность и точность преобразования этих кодов и т.п. Информацию, рассматриваемую только с синтаксических позиций, обычно называют данными.

Семантическая адекватность выражает аспект соответствия образа, знака и объекта, т.е. отношение информации и ее источника. Проявляется семантическая информация при наличии единства информации (объекта) и пользователя. Семантический аспект имеет в виду учет смыслового содержания информации; на этом уровне анализируются те сведения, которые отражает информация, рассматриваются смысловые связи между кодами представления информации.

Прагматическая адекватность отражает отношение информации и ее потребителя, соответствие информации цели управления, которые на ее основе реализуются. Прагматический аспект связан с ценностью, полезностью использования информации для выработки правильного управленческого решения. С этой точки зрения анализируются потребительские свойства информации.

Три формы адекватности информации соответствуют трем ступеням познания истины, сформулированным философами: «От живого созерцания к абстрактному мышлению и от него к практике – таков диалектический путь познания истины, познания объективной реальности». Первая ступень соответствует восприятию внешних структурных характеристик, т.е. синтаксической стороны информации; вторая – ступень формирования понятий и представлений, выявления смысла, содержания информации и ее обобщения; третья – непосредственно связана с практическим использованием информации в соответствии ее целевой функции деятельности системы. В соответствии с тремя формами адекватности выполняется и измерение информации. Терминологически принято говорить о количестве информации и об объеме данных. Объем данных в сообщении измеряется количеством символов (разрядов) принятого алфавита в этом сообщении. Часто информация кодируется числовыми кодами в той или иной системе счисления. Естественно, что одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта.

Действительно:

$$N = m^n,$$

где: N – число всевозможных отображаемых состояний;

$m$  – основание системы счисления (разнообразие символов, применяемых в алфавите);  
 $n$  – число разрядов (символов) в сообщении.

Поэтому в различных системах счисления один разряд имеет различный вес, и соответственно меняется единица измерения данных. Так, в двоичной системе счисления единицей измерения служит «бит» – двоичный разряд, в десятичной системе счисления – «дит», как десятичный разряд.

Например:

а) сообщение в двоичной системе 10111011 имеет объем данных  $V_d = 8$  бит;

б) сообщение в десятичной системе 275903 имеет объем данных  $V_d = 6$  дит.

В современной ЭВМ наряду с минимальной единицей данных «бит» широко используется укрупненная единица измерения «байт», равная 8 бит.

Определение количества информации на синтаксическом уровне невозможно без рассмотрения понятия неопределенности состояния системы (энтропии системы).

Действительно, получение информации о какой-либо системе всегда связано с изменением степени неосведомленности получателя о состоянии этой системы. До получения информации получатель мог иметь некоторые предварительные (априорные) сведения о системе  $\#$ . Мера неосведомленности о системе  $H(\#)$  и является для него мерой неопределенности состояния системы. После получения некоторого сообщения получатель приобретает некоторую дополнительную информацию  $I_\beta(\#)$ , уменьшающую его априорную неосведомленность так, что апостериорная (после получения сообщения  $\beta$ ) неопределенность состояния системы становится равной  $H_\beta(\#)$ . Тогда количество информации  $I_\beta(\#)$  о системе  $\#$ , полученное в сообщении  $\beta$ , определится как:

$$I_\beta(\#) = H(\#) - H_\beta(\#),$$

т. е. количество информации измеряется изменением (уменьшением) неопределенности состояния системы. Если конечная неопределенность  $H_\beta(\#)$  обратится в нуль, то первоначальное неполное знание заменится полным знанием и количество информации станет равным:

$$I_\beta(\#) = H(\#)$$

Иными словами, энтропия системы  $H(\#)$  может рассматриваться как мера недостающей информации. Энтропия системы  $H(\#)$ , имеющей  $N$  возможных состояний согласно формуле ШЕННОНА, равна:

$$H(\alpha) = - \sum_{i=1}^N P_i \log P_i,$$

где  $P_i$  – вероятность того, что система находится в  $i$ -м состоянии.

Для случая, когда все состояния системы равновероятны, ее энтропия определяется по формуле:

$$H(\alpha) = - \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N} = \log N.$$

Рассмотрим пример. По каналу связи передается  $n$ -разрядное сообщение, использующее  $m$  различных символов.

Так как количество всевозможных кодовых комбинаций определяется по формуле  $N = m^n$ , то при равновероятности появления любой из них количество информации, приобретенной абонентом в результате получения сообщения, будет определяться по формуле ХАРТЛИ:

$$I = \log N = n \log m$$

Если в качестве основания логарифма принять  $m$ , то формула упростится и количество информации станет равным:

$$I = n$$

В данном случае количество информации (при условии полного априорного незнания абонентом содержания сообщения) будет равно объему данных  $I = V_d$ , полученных по каналу связи.

Наиболее часто используются двоичные и десятичные логарифмы. Единицами измерения в этих случаях будут соответственно «бит» и «дит».

Степень информативности сообщения определяется отношением количества информации к объему данных, т.е.

$$Y = I / V_d, \text{ причем } 0 \leq Y \leq 1,$$

где:  $Y$  – характеризует лаконичность сообщения.

С увеличением  $Y$  уменьшаются объемы работы по преобразованию информации (данных) в системе. Поэтому стремятся к повышению информативности, для чего разрабатываются специальные методы оптимального кодирования информации.

#### **Семантическая мера информации.**

Синтаксические меры количества информации в общем случае не могут быть непосредственно использованы для измерения смыслового содержания, ибо имеют дело с обезличенной информацией, не выражающей смыслового отношения к объекту.

Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне наибольшее признание получила тезаурусная мера информации, предложенная Ю.И. ШНЕЙДЕРОМ. Он связывает семантические свойства информации прежде всего со способностью пользователя принимать поступившее сообщение. Используется понятие «тезаурус пользователя». Тезаурус можно трактовать как совокупность сведений, которыми располагает данная система, пользователь.

В зависимости от соотношений между смысловым содержанием информации  $\hat{S}$  и тезаурусом пользователя  $S_n$  изменяется количество семантической информации  $J_c$ , воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус. При  $S_n \approx 0$  пользова-

тель не воспринимает, не понимает поступающую информацию; при  $S_n \rightarrow \infty$  пользователь все знает, и поступающая информация ему не нужна: и в том, и в другом случае  $J_c \approx 0$ . Максимальное значение  $J_c$  приобретает при согласовании  $\hat{S}$  с тезаурусом  $S_n$  ( $S_n - S_{n \text{ opt}}$ ), когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения ( рисунок 1.9).

Следовательно, количество семантической информации в сообщении, количество новых знаний, получаемых пользователем, является величиной относительной.

Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным (семантический шум) для пользователя некомпетентного.

В то же время понятная, но известная компетентному пользователю информация представляет собой для него тоже семантический шум.

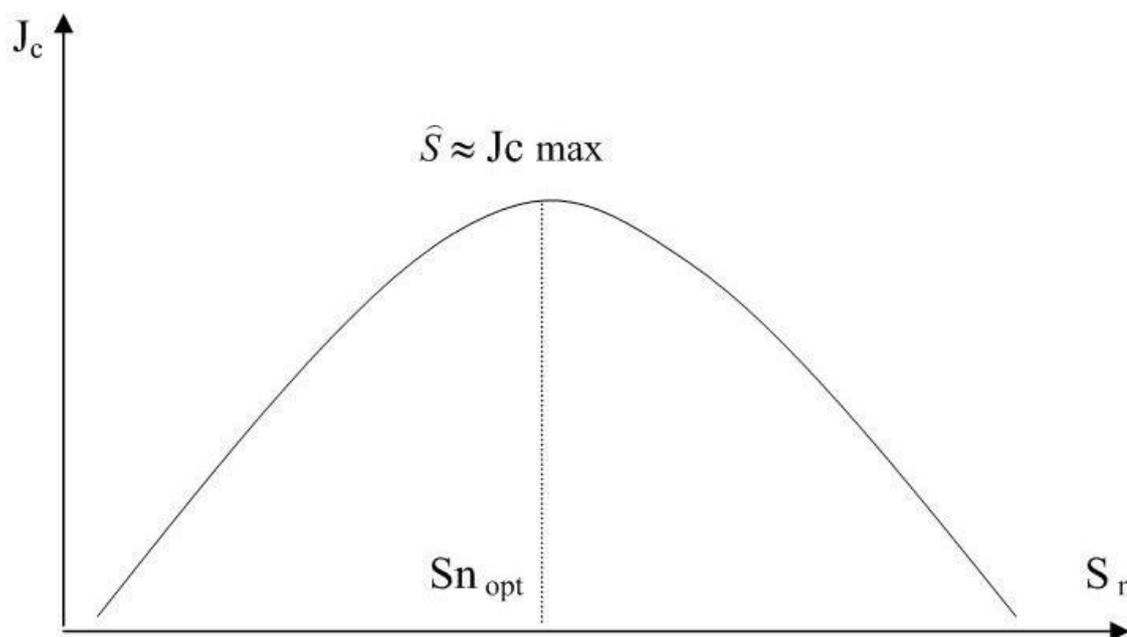


Рисунок 1.9 – Зависимость  $J_c = f(S_n)$

При разработке информационного обеспечения АИС следует стремиться к согласованию величины  $\hat{S}$  и  $S_n$  так, чтобы циркулирующая в системе информация была понятна, доступна для восприятия и обладала наибольшей содержательностью  $S$ , т.е.

$$S = J_c/V_d$$

**Прагматическая мера информации** – это ее полезность, ценность для управления. Эта мера также величина относительная, обусловленная особенностями использования этой информации в той или иной системе.

Ценность информации целесообразно измерять в тех же самых единицах (или близких к ним), в которых измеряется целевая функция управления системой. В автоматизированной системе управления производством, например, ценность информации определяется эффективностью осуществляемого на ее основе экономического управления, или иначе, приростом экономического эффекта функционирования системы управления, обусловленным прагматическими свойствами информации:

$$J_{n\beta}(\gamma) = \Pi(\gamma/\beta) - \Pi(\gamma),$$

где:  $J_{n\beta}(\gamma)$  – ценность информационного сообщения  $\beta$  для системы управления  $\gamma$ ;

$\Pi(\gamma)$  – априорный ожидаемый экономический эффект функционирования системы управления  $\gamma$ ;

$\Pi(\gamma/\beta)$  – ожидаемый эффект функционирования системы  $\gamma$  при условии, что для управления будет использована информация, содержащаяся в сообщении  $\beta$ .

Поскольку экономический эффект функционирования АИС складывается из экономического эффекта решения отдельных функциональных задач, то для вычисления  $J_n$  следует определить:

$Z_\beta$  – множество задач, для решения которых используется информация  $\beta$ ;

$F$  – частоту решения каждой задачи за период времени, для которого оценивается экономический эффект;

$R_\beta$  – степень влияния информационного сообщения  $\beta$  на точность решения задачи,  $0 \leq R \leq 1$ .

Тогда:

$$J_{n\beta}(\gamma) = \Pi(\gamma/\beta) - \Pi(\gamma) = \sum_{j=1}^{Z_\beta} F_i R_{\beta j} \Pi_j$$

где  $\Pi_j$  – экономический эффект от решения  $j$ -й задачи в системе.

В такой постановке единицей измерения ценности информации АИС является обычно рубль.

#### **Качество информации.**

Информация в АИС является и предметом труда и продуктом труда, поэтому от ее качества существенно зависят эффективность и качество функционирования системы.

*Качество информации* можно определить как совокупность свойств, обуславливающих возможность ее использования для удовлетворения определенных в соответствии с ее назначением потребностей.

## **Конец ознакомительного фрагмента.**

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.